ThermoNet: Deep Neural Network Thermogram Analysis of Human Calves during Physical Exercise

Daniel Andrés López

June 24, 2024

DOI: 10.25358/openscience-10770

Institute of Computer Science

JG|U

ThermoNet: Deep Neural Network Thermogram Analysis of Human Calves during Physical Exercise

Dissertation

submitted for the award of the title

“Doctor of Natural Sciences”

to the Faculty of Physics, Mathematics and Computer Science
of Johannes Gutenberg University Mainz
in Mainz

Daniel Andrés López

Born in Bad Soden am Taunus

Mainz, June 24, 2024

Daniel Andrés López
ThermoNet:
Deep Neural Network Thermogram Analysis of Human Calves during Physical Exercise
Dissertation, June 24, 2024
Reviewers: [name removed] and [name removed]

Date of the oral examination: October 1, 2024

Johannes Gutenberg University Mainz
Computational Geometry
Institute of Computer Science
FB08
Staudingerweg 9
55128 Mainz

Abstract

Applied infrared thermography allows practitioners and researchers to evaluate human thermoregulation and gain physiological insights based on pattern recognition of non-invasive acquired thermograms. Current research in sports science and medicine already utilizes thermography in several areas, including injury detection and prevention, disease detection and monitoring, as well as understanding the metabolism and physiology of individuals under external physical load or stress. Studies are limited by manual image selection and analysis or the application of specialized hand-crafted algorithms to detect regions of interest. Thermal features are extracted and analyzed from a few thermogram samples. This dissertation proposes the end-to-end acquisition and segmentation pipeline “ThermoNet” to acquire radiometrically calibrated thermograms, segment regions of interest, automatically extract thermal features, and fuse them with additional external sensor data such as heart rate or breath analysis. An entire experiment, measured with a high-speed, high-resolution thermographic camera, is now fully analyzable, instead of being examined only on cherry-picked samples. Contrary to common practice, radiometric calibration is performed in each thermogram using a custom two-point calibration device. Regions of interest include body part extraction, i.e. left and right calf, and vascular-related patterns: superficial vein and perforator patterns. The patterns are additionally analyzed among their individual instances, allowing for further differentiation in explaining thermoregulatory processes. Two specialized deep neural networks semantically segment the thermograms. Therefore, this thesis explores the development of these networks, including the construction of appropriate manually annotated datasets. The work focuses on the backside of runners on a treadmill to evaluate their calves. Other regions of interest are not yet included. To mitigate the lack of initial datasets for these regions, a method for bootstrapping an annotated dataset based on a stereo system with a thermal camera and a visual + depth camera is presented. Application of the system results in automatically annotated datasets that provide a starting point for new segmentation models and reduce the need for large manually annotated datasets. The processing pipeline “ThermoNet” allows analysts to apply further investigation to the time series of an entire experiment. Several studies revealed relationships between skin temperature radiation and other physiological attributes. Thus, the work integrates into several areas of sports science and medicine.

Zusammenfassung

Die Infrarot-Thermografie ermöglicht es Forschern, die Thermoregulation des Menschen zu untersuchen und physiologische Erkenntnisse durch Mustererkennung in nicht-invasiv aufgenommenen Thermogrammen zu gewinnen. In der aktuellen Forschung in der Sportwissenschaft und der Medizin wird die Thermografie bereits in einigen Bereichen eingesetzt, dies schließt die Erkennung und Prävention von Verletzungen und Krankheiten, das Verständnis des Stoffwechsels und der Physiologie von Menschen unter äußerer körperlicher Belastung oder Stress ein. Die manuelle Auswahl und Analyse von Thermogrammen oder die Verwendung spezialisierter Algorithmen zur Erkennung interessanter Regionen schränken die derzeitigen Studien ein. Nur aus wenigen Thermogrammen werden thermische Kennzahlen extrahiert und analysiert. In dieser Dissertation wird zunächst die durchgängige Erfassungs- und Segmentierungspipeline „ThermoNet“ vorgestellt, um radiometrisch kalibrierte Thermogramme zu erfassen, interessante Regionen zu segmentieren, thermische Kennzahlen zu extrahieren und diese mit zusätzlichen externen Sensorendaten wie der Herzfrequenz oder einer Atemanalyse zu kombinieren. Ein komplettes Experiment, das mit einer hochauflösenden Thermografiekamera aufgenommen wurde, kann vollständig analysiert werden, anstatt nur Stichproben zu nehmen. Im Gegensatz zur üblichen Praxis erfolgt die Kalibrierung in jedem Thermogramm durch eine Zwei-Punkt-Kalibrierung. Der Fokus liegt auf der Extraktion der linken und rechten Wade und der automatischen Erkennung von Venen- und Perforationsmustern der Blutgefäßstrukturen, die zur differenzierten Erklärung thermoregulatorischer Prozesse zusätzlich instanzbasiert analysiert werden. Zwei tiefe neuronale Netze segmentieren die Thermogramme. Daher wird in dieser Arbeit die Entwicklung dieser Netze untersucht, einschließlich der Erstellung geeigneter manuell annotierter Datensätze. Die Arbeit konzentriert sich auf die Rückseiten der Waden von Läufern auf einem Laufband. Es gibt keine annotierten Thermogramme für weitere Körperteile. Um diesen Mangel zu beheben, wird eine Methode zum Bootstrapping eines annotierten Datensatzes vorgestellt, die auf einem Stereosystem mit einer Wärmebildkamera, einer visuellen Kamera und Tiefenkamera basiert. Die Anwendung des Systems liefert automatisch annotierte Datensätze für das Training neuer Segmentierungsmodelle und reduziert den manuellen Annotationsaufwand. „ThermoNet“ ermöglicht es Analysten, weitere Untersuchungen auf die Zeitreihen eines gesamten Experiments anzuwenden. In mehreren Studien wurden Zusammenhänge zwischen der Hauttemperaturstrahlung und anderen physiologischen Merkmalen gefunden. Die Arbeiten können daher in verschiedene Bereiche der Sportwissenschaft und Medizin integriert werden.

Contents

Part I

Foundations

1 Introduction

Human perception of the world is largely based on the visual system. Many applications are performed subconsciously by the brain's visual processing and combined with other senses. These include applications such as object detection and instance segmentation, tracking over time, motion capture, 3D world building, and feature recognition in multiple scenarios and views. Research and industry are trying to mimic these applications with camera-based systems and enhance them with automatic analysis methods. Examples include tracking athletes in soccer games or assisting surgeons in medicine. Multiple cameras combined with sensors from many domains can access information and provide them in a quantified and persistent way. Because visual impressions are representations of light in the human visible spectrum, they are limited to a small range of light wavelengths. Technology has developed many systems to detect photons in other ranges, revealing hidden features and enabling new applications. An example is X-ray technology, which provides an internal view of objects based on their permeability to the particular radiation, for example, medical diagnosis of broken bones or security screening of luggage at airports. In addition to visible light and X-rays, there are many other spectra to discover for further insight.

One promising spectrum is infrared, which contains the thermal radiation spectrum of objects. Structural insight can be gained just by inspecting the thermal radiation properties of objects. The industry has developed many applications ranging from production process support to quality control, defect detection, heat flow analysis, and many more. Stationary and mobile systems are available for all types of applications. As sensor technology continues to improve, systems are achieving higher accuracy, speed, and resolution. These are important factors when applied to small and moving objects/structures. Therefore, medicine has discovered this field for supporting diagnosis of several diseases such as breast cancer. In recent years, the information has also been evaluated by sports scientists to analyze an athlete's performance or body status, e.g., to detect muscle anomalies or to prevent injuries.

1.1 Motivation and Problem Statement

Measuring the thermal activity of a body provides several physiological insights into the thermoregulatory system [176, 134, 61, 135]. Currently, the diagnosis and monitoring of human internal states in medicine and sports science often requires sensors attached to the body. Whether to diagnose pathophysiological problems or to analyze a person's physiological constitution, sensors are attached to specific parts of the body for continuous measurements. Attachable sensors can measure many characteristics such as heart rate, blood flow, breath components and respiratory rate, skin temperature, core body temperature, sweat rate, cardiac response. Additional unique events can be recorded, including blood samples for lactate concentration. There are other devices that allow non-contact analysis, e.g., cameras, X-rays or MRI. The former can capture motion and display it on a screen for visualization and analysis. X-rays, MRIs, and similar methods introduce a new perception because their imaging system is based on different wavelengths of radiation than a common visible light camera. This allows images to be taken from inside the body. However, these methods require a lot of equipment and can also be invasive (X-rays). In addition, it is not possible to take images of a moving person. Over the past decade, infrared thermography has become increasingly popular in medical and sports science. The technology enables the detection of the thermal radiation emitted by any material and relate it to a temperature at the object's surface. The new knowledge allows medical and sports scientists to observe the human body non-invasively. Many studies have already identified applications in breast cancer detection, early inflammation or rehabilitation monitoring. In addition, insights can be derived for animals, especially in equestrian sports to measure the health and performance of horses. In the field of sports science, ThermoHuman [@21] has pioneered thermographic applications in professional sports teams (mostly soccer). Many studies lack the ability to measure data in motion [130]. Analysis strategies typically use a fixed, normalized posture to extract regions of interest (ROIs) automatically or manually. However, this approach requires a high degree of standardization in image acquisition and experimental design. Dynamic information gain is not possible because thermograms are often not acquired during activities. Some studies have addressed this problem and started to perform motion analysis. In [163] infrared thermography (IRT) is applied while running, but not continuously. Only every 5 minutes an image is taken. Racinais et al. [140] provide live analysis during running

races, but only with two time points (first and last lap). A fully automated processing pipeline is still missing. Due to the manual selection of ROIs, the analyzed data depend on the investigator. The reliability between different individuals, work groups and laboratory environments is discussed in [104, 134]. Standardization also limits the application of IRT because it requires highly trained personnel and is prone to errors for small deviations from the specified protocol. Most research focuses on regional analysis of surface radiation temperature (T_{sr}). However, thermograms also contain information about the cutaneous vasculature, which influences the heat distribution of the skin. These patterns are already known, but due to limited image resolution, the patterns have not been well classified. Newer technologies allow more detailed analysis and are gaining interest among research groups [61]. The authors refer to these methods as “exercise radiomics”. A general detection method for the vascular patterns has not yet been developed.

To overcome the limitations of current methods in human IRT analysis, we aim to automate the processing pipeline from acquisition, ROI selection and feature extraction. In addition to body regions, vascular-related ROI patterns should be found and analyzed. The automation should also allow the analysis of non-stationary postures, therefore a comprehensive processing of all accessible moving body parts is encouraged in this work. The topic of image acquisition is mostly associated with standardized laboratory environments. In the low temperature range, there are only proprietary calibration methods for radiometric images. It is common practice to place a blackbody with known properties within the thermogram and apply a shift correction [134, 176]. However, this method is still error-prone and requires manual intervention. For ROI detection, many studies present problem-specific solutions for semi-automatic feature extraction, but these are difficult to reproduce due to the lack of common datasets. Data-driven algorithms have a high demand for large numbers of examples, underscoring the need for an independent and extensible approach to generating segmentation datasets. Thermography applied during exercise without interfering with the participant’s movements is not yet common. Therefore, new methods need to be applicable during exercise and also need to extract meaningful features, such as valid straight leg postures. The detection of vascular patterns defines a new understanding of physiological processes when integrated into the pipeline. These vein and perforator patterns must be reliably distinguished. Finally, IRT data is often applied in conjunction with other sensors. Data fusion, which is crucial for reliable correlations and other comprehensive statistics, is not clearly men-

tioned in the literature. Overall, many studies provide non-reproducible work, which also lacks current state-of-the-art image processing, and therefore the full potential of infrared thermography in medicine and sports science is not yet discovered.

1.2 Thesis Structure

The thesis deals with an overall system for the development of an integrated automatic IRT analysis pipeline. Part I provides the foundation for this work. Background information on human physiology, including the necessary biological and sports science knowledge will be given in chapter 2 (p. 9). It continues with image processing and camera basics in chapter 3 (p. 21). IRT and time-of-flight (ToF) cameras are further explained.

Part II explains the methods, data and hardware involved in this thesis. The entire image processing pipeline consists of five main steps (figure 1.1). First, the acquisition step is described, including the hardware setup and the custom radiometric calibration method (chapter 4, p. 39). Also included is the integration of other cameras to create a stereo setup, the associated stereo calibration and other sensors. Due to the lack of publicly available datasets on this topic, several studies have been conducted to create a custom dataset and investigate the thermographic properties. The studies were led by [name removed] of the Department of Sports Medicine, Prevention and Rehabilitation, Institute of Sports Science, Johannes Gutenberg University Mainz, Germany. These are described with their characteristics in chapter 5 (p. 53).

The automatic thermogram processing (steps 2 and 3) is explained in the following chapter 6 (p. 59). The model for step 2 is called body part network (BPN) and for step 3 vessel network (VN). The chapter includes details on the design choices for the deep neural network architecture, training procedure, and data annotation. The last part of the chapter deals with step 4, the extraction of thermal features from the detected regions of interest. In the following chapter 7 (p. 91), an automatic approach for data annotation based on a stereo pair with a thermal and a color+depth (RGBD) camera is proposed. With automatic annotation, a study is conducted to create a custom dataset for deep learning with the steps listed in figure 1.2. Finally, step 5 of the processing pipeline is the sensor fusion in chapter 8 (p. 109) with the description of multi-sensor integration and post-processing methods for time series analysis.

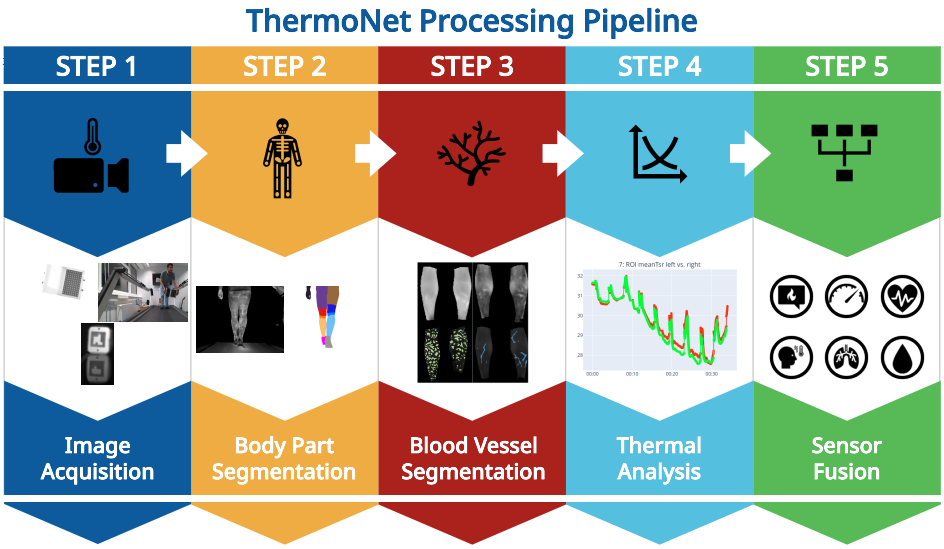


Fig. 1.1.: There are five steps in the ThermoNet processing pipeline: What is needed to acquire thermal images reliably and as automatically as possible (1), how to segment the acquired thermograms into object-related parts such as human body parts (2), how to find internal structures such as blood vessel patterns (3), how to extract features from the ROIs for each image from an experiment (4), and finally how to merge the data with external sources and other sensors (5).

Part III presents the main results (chapter 9, p. 125) and a discussion of the advantages and disadvantages of the developed system (chapter 10, p. 155). Finally, the thesis concludes with chapter 11 (p. 197).

1.3 Publications

This dissertation extends the methodological description of our published work on the thermogram processing pipeline in [60, 59, 58, 6, 7]. In [60] we introduced a deep neural network to semantically segment thermograms of humans in motion to automatically find the ROI. To prove the validity of our work, we compare the results with the common approach of a manual analysis strategy and find a high correlation between both methods. The automated approach is superior because it processes all of the captured thermograms, rather than limiting the analysis to a small fraction for manual investigation. Based on this work, we extended our processing method with more detailed label masks and more statistical features in [59]. As a final development, we presented the complete processing pipeline in the work [7], including a

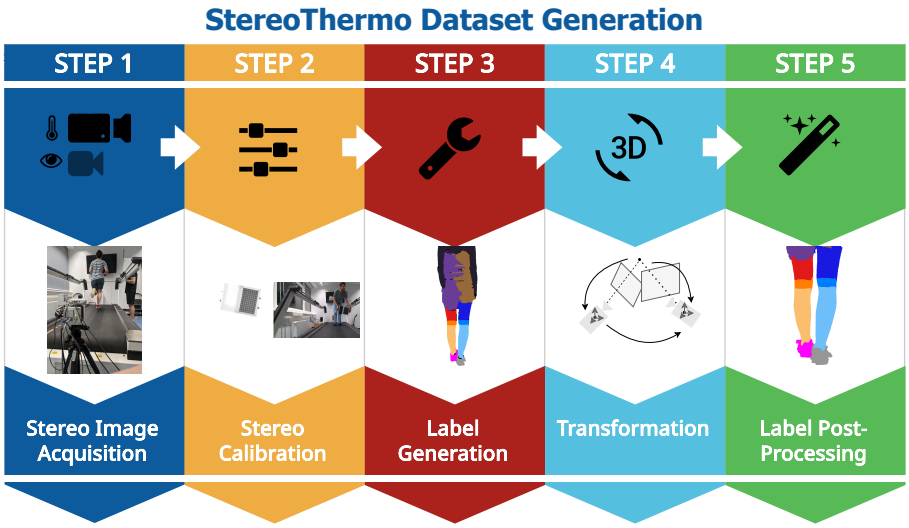


Fig. 1.2.: The five steps to automatically create a custom thermal dataset with image transformation from the RGBD image domain to the IRT image domain: acquire the images synchronously (1), calibrate the cameras and register them in a stereo system (2), generate labels in the common (RGBD) domain (3), transform the label data to the IRT domain (4), and post-process the transformed label to get a final result (5).

detailed description of the acquisition, deep neural network application, and statistical feature analysis. In [6] we present a method to achieve automatic segmentation for new ROIs faster. The main idea incorporates a stereo system consisting of a visual camera, a depth camera, and a thermal camera. While the segmentation in the visual domain is available through already developed algorithms, it can be transformed to the thermal spectrum by the stereo system and employed as training data. The automatic labeling system allows rapid prototyping of new applications in human thermography and is successfully applied to the posterior legs, providing a new dataset [5]. In addition, we are working on another multi-camera scenario. Three thermal cameras capture three different ROIs (calves, forearms and the face) of a single human during exercise on a cycle ergometer and compare the thermal body response during the experiment [58]. Additional sensors helped us to correlate the thermal time series of each ROI with body functions and external load. Although this work does not include a fully automated analysis, it helps to define an analysis strategy to new ROIs and points to possible applications.

2 Human Physiology

This chapter covers the basic concepts of human physiology as far as they relate to this work. First, the concepts of the vascular system and its ability to change its capacity are explained. The next section covers thermoregulation to provide insight into the human regulatory systems that maintain a stable core temperature. In the context of this thesis, the basic principles of external load applied to a body and its reactions are also shown. Finally, the current state of the art for measuring the physical fitness of a body with exercise testing is presented.

2.1 Vascular System

The human body's vascular system carries blood to deliver nutrients, oxygen, and other components to cells and to dispose of unwanted molecules in specific organs. Blood vessels can be divided into arteries and veins. Arteries carry oxygen-rich blood to target organs, while veins carry oxygen-poor blood away from target organs. Further subdivisions are shown in the basic structure diagram figure 2.1 by Pugsley and Tabrizchi [137]. Arteries carry blood from the heart to the tissues of the body. However, they are quite large and cannot interact with the tissues as easily as they should. Therefore, they perforate into smaller arterioles that cover many parts of the tissues, such as organs, muscles, or skin, which is also called perforasome [149]. The interaction with these tissues takes place in the small capillaries. After use, the capillaries join together to form larger venules and later veins to carry the blood back to the heart. Veins also have valves inside them that force the blood not to flow backwards when they are constricted by muscle activity [166]. When a muscle contracts, the valves close and the vein expands a bit to hold the blood in place if it is not pushed forward. Alberts et al. [3] describe in their book how blood vessels are formed and how they can dynamically reshape and adapt to the current needs of the body part, where they are located. A blood vessel is a tubular structure that has a wall consisting of several types

of cells and is loosely connected to the surrounding tissue. It is surrounded by a small smooth muscle layer and the endothelial layer (shown in figure 2.1), which interacts with the autonomic nervous system (ANS) and has other roles in providing access to the blood transport system.

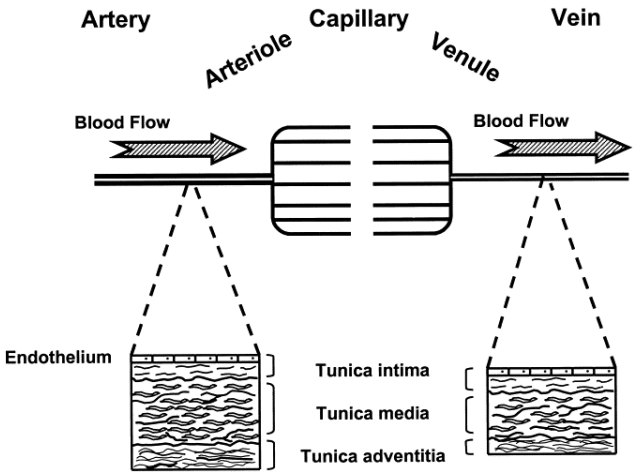


Fig. 2.1.: The figure provides an overview of blood flow with different vessel types and the basic vessel formation [137]. The endothelial layer (endothelium) plays a role in communicating with the autonomic nervous system for vasoconstriction and dilation.

In [153] the authors describe how the body's ANS interacts with the vascular system, which is briefly summarized in the following paragraph. It is responsible for regulating the body's energy supply. The ANS consists of two subsystems: the sympathetic and parasympathetic nervous systems (SNS and PSNS). The former is associated with the activation and alarming of body functions and organs for increased reactivity and performance of body functions necessary in fight or flight situations. While the PSNS correlates with states of recovery or digestion. Both systems work together to regulate body functions. The neurotransmitters released by this system have multiple effects. In the case of the vascular system, the vessel may constrict or dilate, which also affects blood pressure and the interaction between blood components and energy supply. The most dominant neurotransmitter of the SNS for vascular function is noradrenaline/norepinephrine (NE), which causes vasoconstriction. In addition, the energy source adenosine triphosphate (ATP) and Neuropeptide Y (NPY) also have a small vasoconstrictive effect. The PSNS primarily releases two other neurotransmitters that have vasodilatory effects. It also influences other major organs involved in physiological processes such as breathing and heart rate regulation. The PSNS is part of the vagus nerve, the largest nerve in the body. Acetylcholine (Ach) and calcitonin gene-related

peptide (CGRP) are both associated with vasodilation. Blood vessels that need to constrict or dilate also release transmitters through the endothelial cells. They release nitric oxide (NO) during the relaxation phase and endothelin during the constriction phase. During sudden events such as inflammation, disease, or activity, the balance between the two vasoactive inducers is disturbed, resulting in abnormal behavior. The figure 2.2 from [153] illustrates the interaction between the ANS and blood vessels. The ANS does not only facilitate the transport of neurotransmitters to muscle cells for muscle activity, as previously stated; it also interacts with the blood vessel. In contrast, the endothelium cells of the blood vessels releases transmitter particles to interact with the ANS. Vascular disease may also be related to the ANS, such as diabetes mellitus. Vasoconstriction and dilation may not work together as they do in healthy systems.

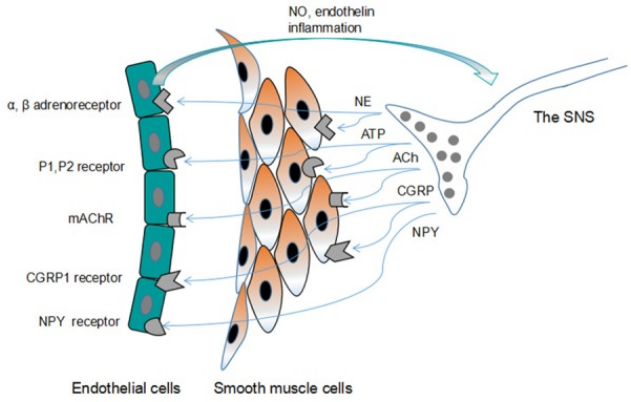


Fig. 2.2.: Interaction of the ANS with blood vessels and muscle cells. Neurotransmitters from the ANS cause the blood vessel to vasoconstrict or dilate, while the endothelial cells send particles to the ANS to relax or constrict. [153]

2.2 Human Thermoregulation

The human organism tries to maintain a stable core temperature, the mechanisms are called thermoregulation [9, 79, 134, 4, 129]. The processes involved are affected by heat production within the body and heat loss to or consumption from the environment. The thermoregulatory system must adjust the body's function when there is a change in physical activity that increases heat production or when environmental conditions change. According to [9, 79] the body has several ways to control heat flow, and some methods of heat transfer are not internally controllable. External factors in heat transfer are

environmental characteristics such as ambient temperature, wind, or solar radiation. These characteristics can be controlled by changing the position of the body itself. Internally, heat is produced in the muscles and transported to other areas, such as blood vessels, by conduction[1] and by blood convection[2], which transports heat to other areas. To maintain a stable temperature, the body must release heat to its environment. Therefore, the three main thermodynamic heat transfer processes can be involved: radiation, conduction and convection. A human body also emits radiation in the thermal infrared wavelength [134]. The amount of radiation cannot be controlled by the body. Hymczak et al. [69] state that the major heat loss is done through the body skin with about 90%. The most relevant heat transfer is thermal radiation from the skin, which is about 65% according to [69]. However, underlying tissues and structures can affect the local emission intensity of the skin. Conduction heat loss can be counted as a macro solution because it would be controllable by the actions of the person when he touches other materials with his body to initiate conductive heat transfer. Conductive heat loss also depends on the material, which has different heat transfer rates. Convection is either related to the environment, such as wind, or to the internal heat transfer of the vascular system, and is therefore also controllable. The convective effect adds up to 10% to 15% of the heat loss. Breathing has a small effect, which heats the air in the body, pushing it out and drawing in new, colder air. However, (sweat) evaporation is more important and localized, with heat loss ranging from 20% to 85% depending on physical activity [69]. The liquid sweat exchanges its heat with the wind or ambient air, which cools the body. Figure 2.3 by [129] provides an overview of the different ways in which heat is transferred. It also mentions factors that affect the transfer process, such as ambient temperature, humidity or solar radiation, as well as individual factors such as clothing and metabolic heat production or physical activity. Personal parameters such as sex, age and individual sweat rate also influence how the body manages the heat transfer process. In addition, the current state of the body plays a role in optimizing the process. An important factor is the body's fluid balance and hydration. With less hydration, the body cannot optimize internal and external heat management as it can with more hydration. However, excess hydration also impairs the ability to optimize heat management.

[1] Conductive heat transfer occurs between two solid materials.
[2] Convection is the heat transfer between solid and liquid or gaseous materials.

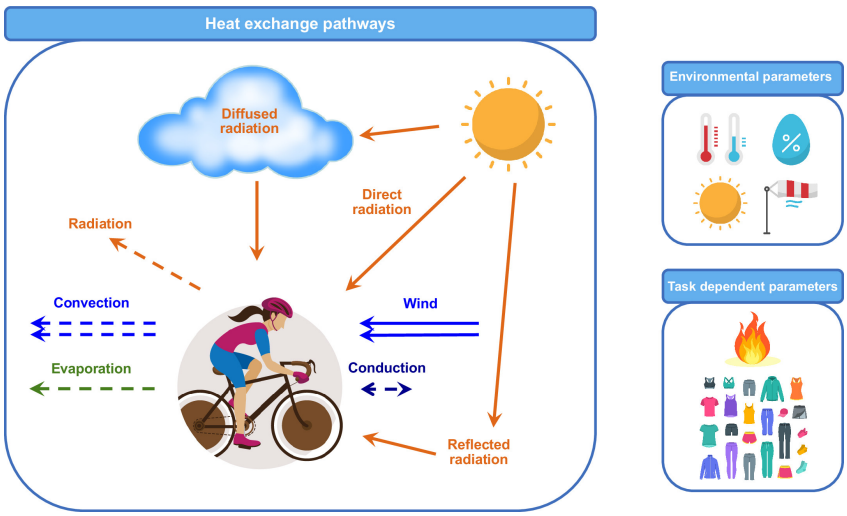


Fig. 2.3.: Various heat transfer processes and factors for human heat transfer, including thermal radiation, convection, and conduction, as well as environmental factors such as solar intensity, humidity, or wind. Specific conditions such as clothing are also important for human thermoregulation. [129]

Taking a closer look at the body's ability to control heat dissipation, we come to the ANS and the vascular system [33]. Internal convection can be increased by increasing blood flow. As explained in the previous section, blood flow can be controlled by constricting and dilating the vessels through signals from the ANS. When heat dissipation is needed, the vessels vasodilate. In addition, local sweat production increases to move heat from the inside to the outside for better convective heat dissipation. Sweat production is also controlled by the ANS. The internal body conducts heat through the tissues and skin to the surface and emits thermal radiation, which is measured as surface radiation temperature (T_{sr}). However, it is not controlled by the body, but is influenced by internal structures such as tissue type, skin function, any external creams applied, and local hair density. Sweat also alters thermal radiation behavior. In addition to forming a small liquid surface, sweat can flow down the skin, increasing convective heat transfer. Surface radiation on the skin accumulates from the lower parts, but is diffused. However, with the infrared thermography (IRT) imaging, the blood vessels in the skin are visible and can be related to the deep veins and arteries that perforate into small superficial vessels, which in turn pass through the underlying tissue and muscle, as explained by Hillen et al. [61]. Muscle activity can also be visualized.

Thermoregulation works in both directions: when the body temperature becomes too high for various reasons, such as external stress. Heat must be

removed from the body. The other direction is to keep heat in the body as much as possible, as is the case in cold environments. Two main controllable methods have been developed: vasoconstriction reduces blood flow especially to the surface areas to reduce convection heat dissipation and as a second method thermogenesis by increased activity, most likely shivering of the muscles. However, the scenario of cold environment and heat retention is not further analyzed in this work, since we work with external load for humans and have controlled room temperature.

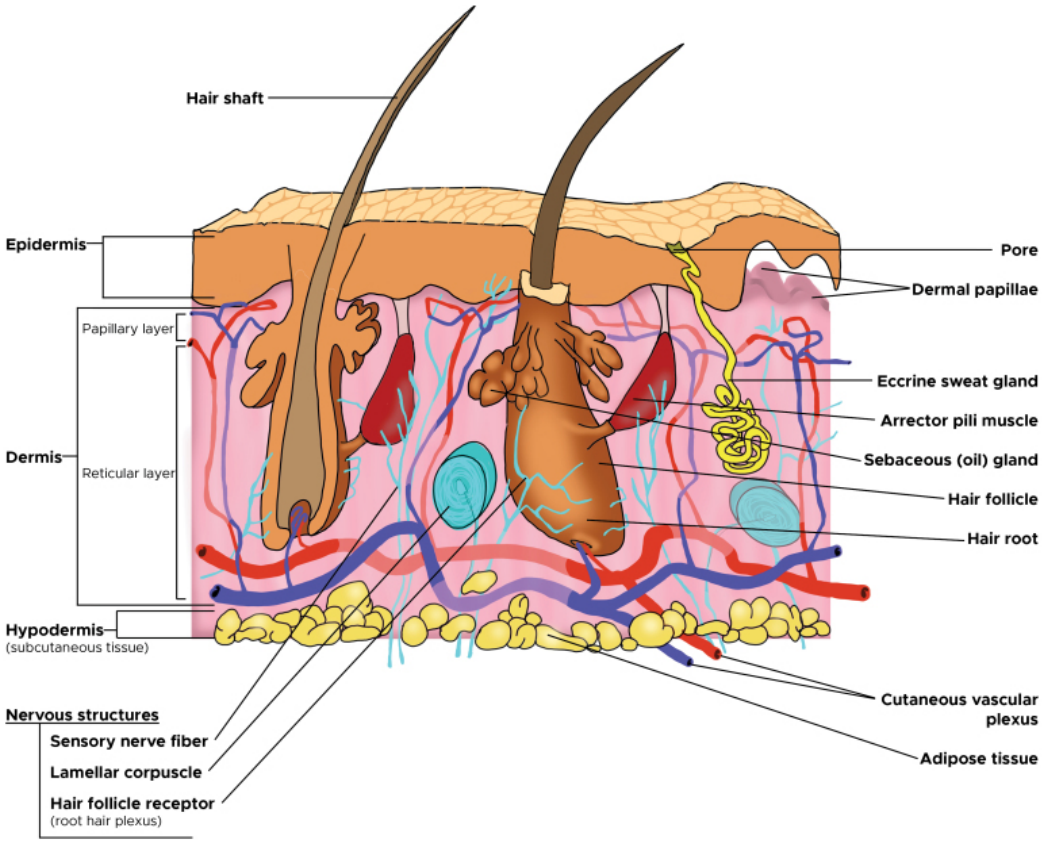


Fig. 2.4.: Human skin is made up of several layers. The dermis layer contains superficial blood vessels, sweat glands and other parts. [181]

Understanding thermoregulation requires an overview of skin formation. Igarashi et al. [71] and Yousef et al. [181] provide insight into the different layers and types of skin and how different functions are achieved by different cells. Heat transport is based on conduction within the skin and appears as thermal radiation from the surface. Figure 2.4 from Yousef et al. [181] show a schematic of the skin. The blood vessels and their perfusion are visible, as are the sweat glands. This demonstrates that the heat from these parts must pass through the dermis and epidermis layers and may originate from

deeper layers. Therefore, thermal surface radiation of the skin is derived from multiple deeper structures.

2.3 Thermoregulation with External Exercise Load

In this thesis we further focus on the effects of thermoregulation during exercise. In a standardized manner with an external load of running or cycling a person has controlled physical activity and is observed for the effects of its body thermoregulation process to gain insights into the physiological processes and examine pathophysiological properties. Hillen et al. [61] provide a review of how the body absorbs heat stress due to physical activity, examining various observations while observing the participants skin involving thermal imaging and providing an overview of current knowledge in the field. The authors identified three main patterns that occur during different types of exercise or after a certain amount of time compared to the natural thermal distribution, which is more or less similar. The first observed pattern correlates with perforasomes. As shown in figure 2.5 by [61], it reflects the small capillaries that come from deeper arteries and divide into tree-like shapes up to the skin, including through the muscles. The temperature is based on the core temperature of the body and is transported to the skin by conduction and convection and dissipates in small structures as thermal radiation. The second observation also applies to blood vessel structures. These vessels are within the muscle tissue and are not deep vessels. Temperature refers to the internal temperature of the active muscle. The associated thermal radiation pattern is more tubular and not as divergent as the snowflake structure. As a final observation, in areas of the body with smaller and near-surface muscles, heat is dissipated through a diffuse, homogeneous area. The third pattern type is not addressed in this work. According to [61, 72] and others, skin temperature decreases during exercise due to the effects of thermoregulation. Sweating plays a large role in the active thermoregulation of the body, while air exchange through breathing has a small effect.

2.4 Measuring Physical Capacity

During physical activity, many internal adaptations take place to ensure that the muscles are functioning for the current activity. In sports science, there

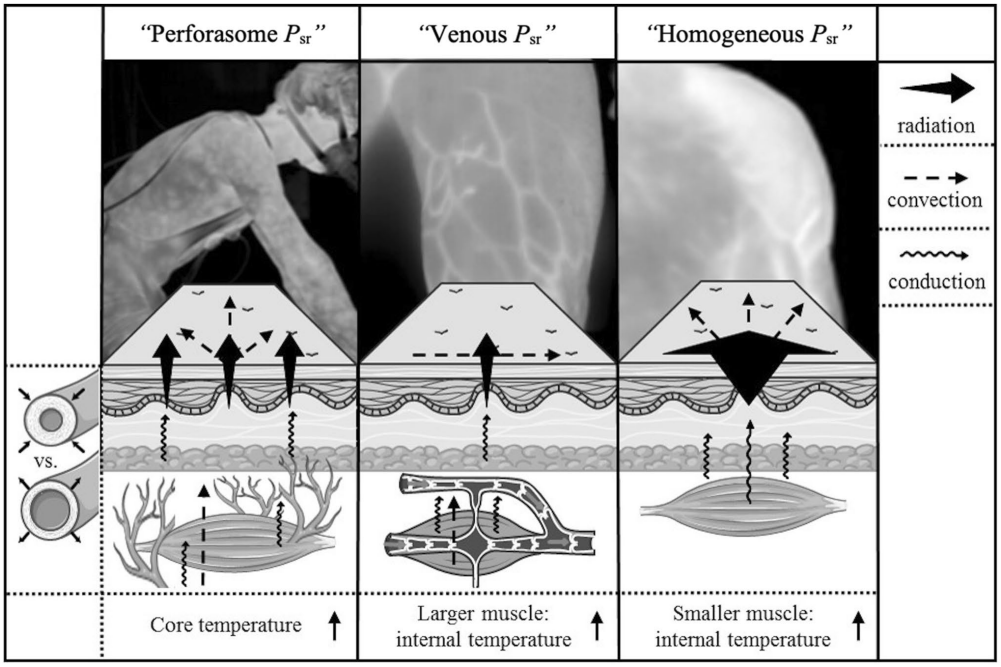


Fig. 2.5.: Three distinct patterns of human surface thermal radiation are associated with superficial perforators, veins, or smaller muscles during physical activity. [61]

are several ways to measure the physical capacity of the body, which provides insight into the processes of the body's ability to sustain the required external loads. The following measurements are state of the art and widely adopted in practice. The medical and technical background is not fully covered in these chapters, as this is not the main focus of this work. The data will be retained for the development of our new method and for comparisons.

RPE A standard measure is the rate of perceived exertion (RPE). It is a personal assessment of current condition in relation to current physical activity. It does not require a measuring device. Borg [24, 23] proposed a method to assess RPE with a standardized scale of 6–20, where 6 means "no exertion at all" and 20 "maximum exertion". For safe experimental evaluation, the RPE can be checked periodically to ensure that the person is able to maintain the current load.

Heart Rate Blood is carried by the vascular system described above. Heart rate determines how many times (beats) the heart pumps blood per minute [bpm]. During exercise, oxygen and energy consumption increases and blood

must flow faster, resulting in a higher heart rate. Heart rate changes constantly. It ranges from about 40 to about 200 bpm. In sports science, heart rate ranges can be used to easily identify a specific body state during exercise. Sammito and Böckelmann [151] provide a systematic review of the measurement of contact and non-invasive methods such as chest belt sensors. In addition to the established contact sensors, new methods have been developed to estimate heart rate in facial image streams using remote photoplethysmography, as reviewed in [93], which are not used in this work.

Lactate The lactate metabolism describes the energy supply and it's end product *lactate* in the body. The metabolism is influenced by various sources such as inflammation, disease or physical activity as Li et al. [96] reported in a review. Cells require oxygen, energy and other nutrients to function. During physical activity, muscles consume more than usual. Then, internal chemical processes build up molecules that must be removed from the cells through the blood. Oxidation processes produce carbon dioxide (CO_2), which is expelled from the body through the pulmonary system. However, when the oxygen supply through the blood system is not fast enough during more intense activity, the body uses other mechanisms to provide energy to the cells. These reactions cannot be sustained for long time and unwanted molecules are produced. One commonly measured product is lactate. This molecule builds up in the muscles and prevents them from working properly. The process of producing lactate through non-aerobic reactions increases with the intensity of the external load. The amount of lactate can be measured in blood samples from many areas of the body because the blood system is completely interconnected and the lactate concentration is assumed to be rapidly distributed. Blood samples are taken from the earlobe for ease of assessment.

Faude et al. [48] reviewed different applications of continuous lactate measurement during physical activity and compared the current state of the art. Many researchers search for a threshold within the lactate concentration curve that describes a breakpoint in lactate metabolism and therefore in the body's ability to sustain the current load. A common strategy is to define a maximum lactate steady state (MLSS) where lactate concentration does not increase under the same external conditions. The same is true for other indicators such as heart rate, oxygen uptake and CO_2 output. The MLSS should refer to the maximum capacity of the oxygen metabolism, above this threshold the anaerobic metabolism takes over without the necessary oxygen supply.

Contrary to its name, it is well known in the literature that the MLSS is not a static threshold, but rather a range and can be influenced by many factors. In addition, many models have been proposed to determine the threshold, also known as the individual anaerobic threshold (IAT). As reported in the review, many researchers have come up with their own definitions, in this thesis the definition of Dickhuth et al. [44, 45] is used. Each IAT calculation has advantages and disadvantages, but the Dickhuth threshold is a commonly applied. The time t_{IAT} is defined when the blood lactate concentration $bLa(\cdot)$ reaches $bLa(t_0) + 1.5$ mmol/l where $bLa(t_0)$ is the lactate concentration at the slowest walking stage of the protocol. If the lactate curve shows a more pronounced increase (shifted to the left) a person is less fit, while a shift to the right (longer low level increase in lactate) indicates a more fit person.

Cardiopulmonary Tests Hollmann and Prinz [65] provide insight into the history and common practice of spiroergometry testing, also known as cardiopulmonary exercise testing (CPET). It consists of two components, the exercise as work on an ergometer (bicycle, treadmill, rowing, etc.) and the spirometry, which examines a gas analysis of the breath. The method is well established in many fields such as sports medicine, cardiology or pneumology for therapy, rehabilitation, research and training control.

While performing the given training on the ergometer, where the load can be precisely controlled in terms of resistance (watts) or speed (km/h), the participant wears a mask that captures the breath and provides fresh air with known characteristics. The air is analyzed for its components and volume. From the combination of the air, conclusions can be drawn about the metabolism. Breath frequency is estimated. Important components such as the volume of inhaled oxygen (VO_2) and the volume of exhaled carbon dioxide (CO_2) are estimated for each breath. The respiratory ratio $RER = VO_2/VCO_2$ gives information about the efficiency of a breath.

As in lactate analysis, individual thresholds can be defined on the VO_2 and VCO_2 curves in spiroergometry [112]. Usually two ventilatory thresholds (VTs) are used. In addition, the maximum VO_2 that a person can achieve at their maximum level of physical activity is described as $VO_{2,peak}$, which is often normalized to kg for better comparability. The higher the maximum oxygen uptake, the fitter a person is. The authors Mazaheri et al. [112] define VTs as the local minimum in the VO_2 and VCO_2 curves from which the slope

trend increases. Metabolic efficiency is directly related to the supply of energy from aerobic (oxygen) or anaerobic sources. VTs help quantify aerobic metabolism.

Training Zones Based on the MLSS/IAT sports scientists define training zones. There are many different definitions. We use a division into 4 zones as shown in table 2.1 inspired by Cleveland Clinic [@4]. Depending on the training goal, a person can train in the targeted zone. Metabolism changes with intensity and primarily uses different fuel sources. The IAT is typically estimated using a treadmill or bicycle exercise test. With the relative speed/power definition of the zones, corresponding heart rate ranges are calculated for training control.

Zone	Name	Abbr.	Relative speed to IAT	Intensity
1	Regeneration	REG	50–70%	low
2	Basic endurance 1	GA1	70–85%	moderate
3	Basic endurance 2	GA2	85–100%	moderate-high
4	Developmental zone	EB	100–110%	high

Tab. 2.1.: Lactate threshold training zones. The (virtual) speed at the IAT defines the reference (100% speed) for the relative running speed definitions of the zones. Modified table from [@4].

Core Temperature In medical practice, many methods have been developed to measure the internal core temperature of the body. However, the scientific community has not fully clarified which is the most representative location in the body to measure the temperature. Hymczak et al. [69] provide an overview of the most common approaches with their advantages and disadvantages. Easily available methods often lack accuracy and have a high risk of misplacing the thermometer. This is the case, for example, with axillary, oral, or body surface measurements. Other measurements, such as the tympanic membrane in the ear, are easily accessible but have high measurement errors due to handling errors. Probes that must be inserted into the body, such as rectal, oral, or bladder and esophageal catheter methods, are more reliable but also have a higher latency and are not suitable for measurements during exercise. The esophagus has been established as the gold standard measurement in medical settings. Pulmonary artery catheters are considered the most accurate, but are invasive and therefore only available in intensive care units.

Sweat Sweating is an important process for the human body to control local and systemic heat loss and maintain core temperature. Baker [12] summarize the role of sweat in thermoregulation and its formation. The body has two main types of sweat glands: eccrine and apocrine. Only the former are used for thermoregulation and are located in the skin all over the body. Sweat is a fluid with two main components, water and NaCl, as well as other molecules. Heat is lost to the environment by evaporation and convection. Measuring sweat is difficult and depends on the application and whether the sweat rate as a whole or the components are being studied. Whole body sweat rate (WBSR) can be measured by the difference in weight between before and after the experiment. However, as Cheuvront and Kenefick [37] discussed, this simple method is also subject to unclear errors. In addition, a local sweat rate can be determined by applying a pad of predefined weight to the skin and measuring it after a strict time frame. Recently, electronic sensors have been improved to measure local sweat rate continuously, but still have limitations [182, 70].

Body Skin As mentioned above, the skin is responsible for about 90% of the body's heat loss. In the review [106], the authors analyzed studies going back to 1960 for contact skin temperature T_{sk} measurements. They identified several problems with the repeatability and accuracy of the T_{sk} measurement. Typically, a sensor is placed directly on a single point on the skin and reports its readings periodically. The sensor has to deal with heat accumulation under its own material. In addition, it can be influenced by various sources such as sweat or ambient air.

3 Imaging Basics

This thesis relies on several imaging concepts and technologies. In this chapter the basics for understanding the following chapters are given. First, a generic, ideal camera model is introduced. It is the common model for imaging visible light, but can also be applied to other electromagnetic waves such as infrared. Infrared imaging technology (thermography) and how to calculate temperatures from measured radiation are also explained. Another infrared imaging technology will also be discussed, a time-of-flight system with an active near-infrared system to capture depth information of a scene. Thermography and depth cameras can be combined to obtain a depth map along with temperature radiation information. Therefore, epipolar geometry is the key to registering the two systems together to create a stereo system.

3.1 Camera Model

To analyze the real world and *see* computationally, information from the real world must be transformed into computer-readable information. Images represent a particular view of the world with discrete and finite information. The image formation process of cameras is modeled by the pinhole camera model. Additionally, a lens distortion model is integrated. All cameras employed in this thesis are based on these principles. The pinhole camera works on the principle that light rays reflected from objects pass through a small pinhole and fall onto the sensor plane. The sensor captures the light rays and provides a discrete accumulated intensity value when read out. Each pixel requires a separate sensor element (e.g. photodiode). A lens is placed in front of the camera to focus the light rays onto the sensor array. In this way, the 3D points of the real world are mapped onto 2D points of the sensor plane (also called the image plane). Figure 3.1 by [31] shows the transformations to project from the 3D world coordinate system to the 2D image coordinate system, which is represented by a discrete grid of pixels.

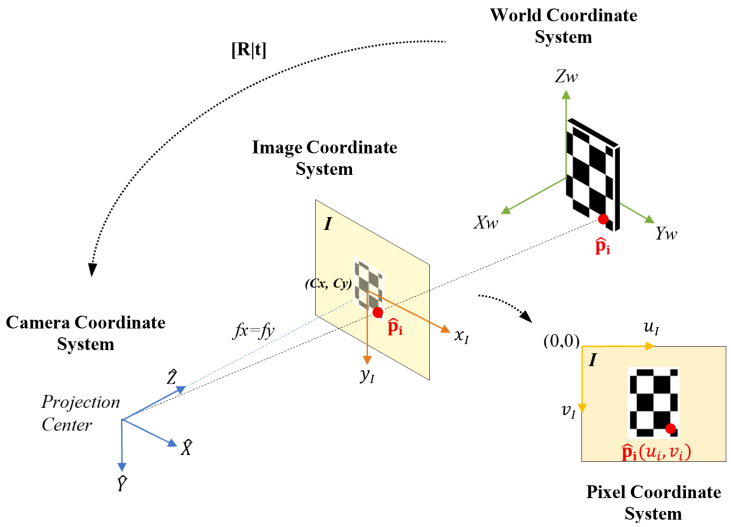


Fig. 3.1.: Pinhole camera model components: The object in the world coordinate system is projected through the projection center of the camera onto the (virtual) image plane (image coordinate system). The image plane is built with a sensor array (pixel coordinate system). [31]

3.1.1 Points, Transformations, and Projections

When analyzing cameras and objects in 3D and projecting them to images in 2D, different spaces and coordinate systems are involved. 3D points are represented by $\boldsymbol{X} = \begin{bmatrix} x & y & z \end{bmatrix}^\top \in \mathbb{R}^3$. However, for more convenient description and modification, we choose the equivalent representation with homogeneous coordinates in projective space $\overline{\boldsymbol{X}} = \begin{bmatrix} \overline{x} & \overline{y} & \overline{z} & w \end{bmatrix}^\top \in \mathbb{P}^3$ where the last coordinate w is the homogeneous coordinate. Points in 2D are written as $\boldsymbol{x} = \begin{bmatrix} u & v \end{bmatrix}^\top \in \mathbb{R}^2$ with the homogeneous extension in projective space $\overline{\boldsymbol{x}} = \begin{bmatrix} \overline{u} & \overline{v} & w \end{bmatrix}^\top \in \mathbb{P}^2$. When working with multiple cameras, multiple views of the scene and its points are available. Each view is represented by its own coordinate system definition. The index B (right subscript) in $\boldsymbol{x}_\mathrm{B}$ and in $\boldsymbol{X}_\mathrm{B}$ indicates the name of the coordinate system. A transformation A is a mapping function that changes the coordinate system of a given point from B to C ($A : \mathrm{B} \mapsto \mathrm{C}$). Transformations in the projective space $\mathbb{P}^3$ can be expressed as a quadratic homogeneous matrix, with the homogeneous coordinate in row and column 4. The coordinate system change is indicated in the subscripts of the matrix with the target and source coordinate systems: $A_{\mathrm{C,B}}$. A transformation can be concatenated with another transformation by matrix multiplication to combine coordinate system changes (3.1). Applying a transformation $A_{\mathrm{C,B}}$ to a point $\overline{\boldsymbol{X}}_\mathrm{B}$ changes the coordinates from B to C by a matrix-vector multipli-

cation (3.2). Additionally, transformations are invertible $A_{\mathrm{B,C}}^{-1} = A_{\mathrm{C,B}}$, and it holds that $A \cdot A^{-1} = A^{-1} \cdot A = I$, where $I = diag(1)$ is the neutral element, also called identity.

$$A_{\mathrm{C,D}} = A_{\mathrm{C,B}} \cdot A_{\mathrm{B,D}} \tag{3.1}$$

$$\overline{\boldsymbol{X}}_{\mathrm{C}} = A_{\mathrm{C,B}} \cdot \overline{\boldsymbol{X}}_{\mathrm{B}} \tag{3.2}$$

In this work, the relevant transformations are translations, rotations, and perspective transformations. Translations have the effect of moving a point by a certain amount with the vector $t = \begin{bmatrix} t_1 & t_2 & t_3 \end{bmatrix}^\top \in \mathbb{R}^3$, creating a transformation matrix $A = \begin{bmatrix} \mathbf{0} & t \\ \mathbf{0}^\top & 1 \end{bmatrix}$. The group $SO(3)$ (Special Orthogonal Group) contains in Euclidean space (dimension 3) all rotations $R \in SO(3)$. The rotation is in the top left part of a transformation: $\begin{bmatrix} R & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{bmatrix}$. In $SO(3)$, the inverse of an element is the same as the transposed matrix: $R^{-1} = R^\top$. The $SE(3)$ (Special Euclidean Group, also called rigid body motions) group denotes the transformations in Euclidean space (dimension 3) consisting of rotational and translational parts. However, the representation of these transformations in a single transformation matrix is only possible in the projective space $\mathbb{P}^3$: $\begin{bmatrix} R & t \\ \mathbf{0}^\top & 1 \end{bmatrix}$. Together with the scaling and shearing transformations, all of the above form the group of affine transformations.

Perspective projections are special transformations that do not change the coordinate system. They change the space: $K : \mathbb{R}^3 \mapsto \mathbb{P}^2$. In our case from Euclidean 3D space to projective 2D space. The perspective projection is $K \cdot \boldsymbol{X} = \begin{bmatrix} \boldsymbol{x} \\ z \end{bmatrix}$ with the camera calibration matrix K (3.3). The z coordinate of a 3D point gets a homogeneous coordinate in 2D.

$$K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{3.3}$$

3.1.2 Camera Projection

Points $\boldsymbol{X}$ in the world coordinate system are named $\boldsymbol{X}_{\mathrm{W}}$. Each camera has a corresponding (virtual) image plane on which $\boldsymbol{X}$ is projected with the camera

calibration matrix K (3.3). The points are projected onto a 2D plane in the projective space $\mathbb{P}^2$ according to the projection equation:

$$\begin{bmatrix} \boldsymbol{x} \\ 1 \end{bmatrix} = \frac{1}{z} \cdot K \cdot \boldsymbol{X} \tag{3.4}$$

z is the distance, but as a homogeneous coordinate it is normalized to 1. The depth information z is lost in the projection process. When depth is estimated together with the camera acquisition process, these depth values are called $\mathrm{d} \in \mathbb{R}$.

In addition to 3D to 2D projections and point definitions, a rigid body transformation change the coordinate systems. These parameters are called extrinsic parameters and have the shorthand notation $[R|t]$. Together with a camera calibration matrix (intrinsic calibration) they form the camera projection matrix P, which transforms points from the coordinate system B to the camera C and then projects them onto the image plane with K_{C}.

$$P_{\mathrm{C,B}} = K_{\mathrm{C}} \cdot [R|t]_{\mathrm{C,B}} \tag{3.5}$$

3.1.3 Optics and Lenses

Real cameras have lenses attached to the camera body. They introduce distortions (coefficients D). According to Zhang [185], the most common distortions are radial distortions, but tangential distortions can also be considered. For our purposes, we consider a simplified model with 3 coefficients for radial distortion and 2 for tangential distortion as described in the OpenCV library [25]. Radial distortions k_1, k_2, k_3 affect the pixels in the image plane according to (3.6) with (3.7). Tangential influence occurs when the lens is tilted or shifted relative to the image plane. It is modeled with p_1, p_2 in (3.8).

$$\begin{bmatrix} u \\ v \end{bmatrix}' = (1 + k_1 \cdot r^2 + k_2 \cdot r^4 + k_3 \cdot r^6) \cdot \begin{bmatrix} u \\ v \end{bmatrix} \tag{3.6}$$

$$r^2 = u^2 + v^2 \tag{3.7}$$

$$\begin{bmatrix} u \\ v \end{bmatrix}'' = \begin{bmatrix} u + (2p_1uv + p_2(r^2 + 2u^2)) \\ v + (p_1(r^2 + 2y^2) + 2p_2xy) \end{bmatrix} \tag{3.8}$$

The distortions place the image pixels at different positions on the sensor, as it would be expected from the camera model. The effect of radial distortions is

illustrated in figure 3.2 by [25]. Removing distortion from an image is called rectification.

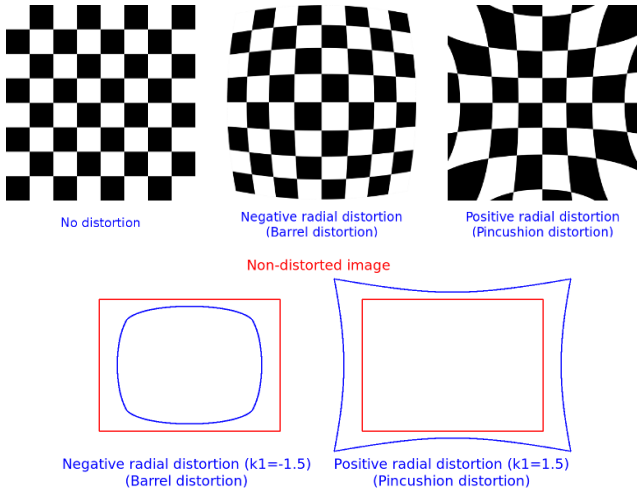


Fig. 3.2.: Example of the effect of radial distortion on a checkerboard (first row) and, in the second row, how an image (red) will be undistorted (blue) in different cases. [25, @18]

3.1.4 Camera Calibration and Epipolar Geometry

Intrinsic and extrinsic parameters are crucial for computer vision tasks in this thesis. Standard algorithms for obtaining them work with all imaging modalities that have the same pinhole camera and lens model. However, finding the necessary corresponding points is more challenging and will be discussed further in section 7.1. This section describes the basic parts according to Hartley and Zisserman [53].

Intrinsic calibration estimates the camera calibration matrix K and the distortion coefficients D. Tsai [170] refined the camera calibration by using a direct linear transformation and solving it with least squares to estimate intrinsic and extrinsic parameters simultaneously from a single view. The approach includes first estimation of camera parameters and second correction of lens distortions to achieve high accuracy results. An optimization of Tsai's method is provided by Zhang [185]. The point correspondences of 3D-2D points are based on views of planar patterns with known properties, such as squares of fixed size and known length.

The extrinsic parameters of the camera calibration are the relative position t and orientation R of the camera coordinate system C to the world coordinate

system W: $A_{\mathrm{C,W}} = [R|t]_{\mathrm{C,W}}$. With a single camera, the need for extrinsics depends on the application. In stereo vision, however, the extrinsics are crucial, since they define the relative pose between the two stereo cameras. Beschi et al. [18] summarize several ways to obtain the calibration parameters, including R and t.

Stereo vision describes the combination of two cameras and their relationship [53]. The cameras are relatively fixed in their position and orientation, as well as their lens settings such as focus and zoom. Without loss of generality, we assume that one camera coordinate system L (left) is the world coordinate system W. The other camera R (right) is located relative to L with $[R|t]_{\mathrm{R,L}}$, where R is the rotation and t is the translation from the left to the right coordinate system. Stereo allows to determine depth information from two corresponding views (captured at the same time) of the same scene, or at least overlapping elements, but from different poses in the world. In order to reconstruct the original 3D point, the ray between the optical center and the image point is estimated for each camera. Since depth is lost in projection, the intersection of the two camera rays determines the exact location of the 3D point. However, the theoretical recovery fails due to several errors such as uncertainties and discretization in the image acquisition or less accurate relative extrinsic estimation. These errors result in reconstructed rays that do not intersect and therefore no depth information can be computed. An important step in preparing for triangulation is to find corresponding image points in both views. There is a relationship that connects a point in one image to a corresponding line in the other image. The corresponding point in the first image must be on that line. The constraint is called the epipolar constraint, and the line epipolar line l. All epipolar lines of an image intersect at a single point, the epipole. The formalization of the epipolar constraint includes the fundamental matrix $F \in \mathbb{R}^{3\times3}$ with $det(F) = 0$ and f_{33} as a homogeneous coordinate. The fundamental matrix is the basic information that defines the relationship between the corresponding points in the stereo system. For the point $\overline{x}_{\mathrm{L}}$ in the left image and the corresponding epipolar line in the right image l_{R} (and vice versa) the following equations hold:

$$F \cdot \overline{\mathrm{x}}_{\mathrm{L}} = l_{\mathrm{R}} \tag{3.9}$$

$$F^{\top} \cdot \overline{\mathrm{x}}_{\mathrm{R}} = l_{\mathrm{L}} \tag{3.10}$$

$$\overline{\mathrm{x}}_{\mathrm{R}}^{\top} \cdot F \cdot \overline{\mathrm{x}}_{\mathrm{L}} = 0 \tag{3.11}$$

By calculating epipolar lines in associated images, corresponding points can be found more easily. Any point on the epipolar line can be used for triangulation, since its ray would intersect the ray from the source point. However, the fact that the epipolar line is not aligned with the pixel grid of the images still hinders efficient correspondence matching. To further improve the process, both images are rectified. In rectification, both images are brought into a common image plane by rotating, shifting, and scaling them so that there is only a horizontal or vertical offset between the images, depending on the horizontal or vertical physical hardware setup. Now the epipole is at infinity (the homogeneous coordinate is 0), which means that all epilines are parallel, also to the baseline (connection between the two principal points). Thus, a corresponding epiline is directly aligned to the image grid and can be processed more efficiently. The rectification can be done in many ways [53]. Virtually it can also be represented as a new camera and is denoted by the prefix r like $P_{\mathrm{rR,W}}$, $P_{\mathrm{rL,W}}$. New transformations for this coordinate system can be applied as intermediate steps like other transformations.

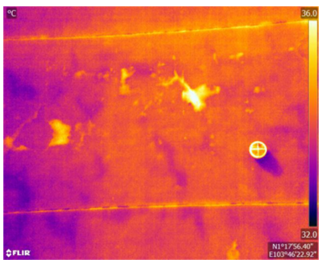
3.2 Infrared Thermography

Infrared thermography (IRT), infrared thermal imaging, or simply thermography, is a rapidly growing technology with many applications in surveillance, construction inspection, assembly quality assurance, medicine, sports science, and many other fields [172, 16, 176]. The examples in figure 3.3 give an insight into the visual impression of the infrared modality, which is completely different from images of the visible light spectrum. The main principle is based on the physical property of thermal radiation of objects themselves.

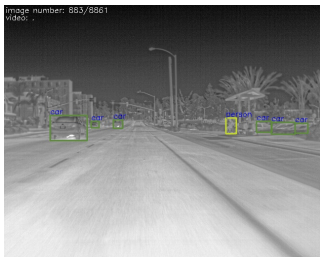
3.2.1 Infrared Radiation

Infrared radiation is a part of the electromagnetic wave spectrum. Other parts are visible light, X-rays, microwaves, and so on. In the context of this thesis, the wave representation from the particle-wave dualism is sufficient. The thermal infrared spectrum is part of the infrared spectrum and ranges from ~1.4 to ~15 µm. According to figure 3.4 by [176], it can be divided into three subdivisions: short, medium, and long infrared with wavelengths of 0.9 to 1.7 µm, 3 to 5 µm, and 8 to 14 µm. The origin of the thermal radiation and

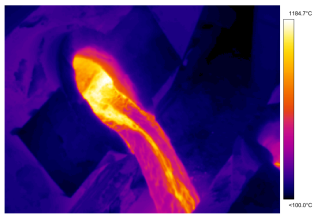
(a) Visual image of a building wall. [169]



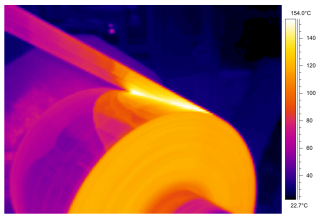
(b) Thermal image of (a) with material delamination visible in the thermogram but not in the visual image. [169]



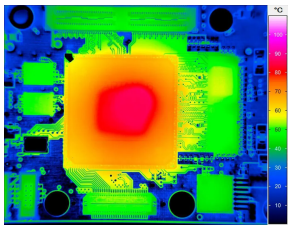
(c) Traffic scene captured with a thermal camera. Labels are available for people, cars, and bicycles. [@27]



(d) Liquid pig iron is analyzed for its constituents in the stream. [172]



(e) Quality assurance measurements for steel strip rolling. [172]



(f) Fault detection of circuits and electronic devices. [@13]

Fig. 3.3.: Examples of IRT imaging in buildings, traffic scenarios, and industrial materials inspection.

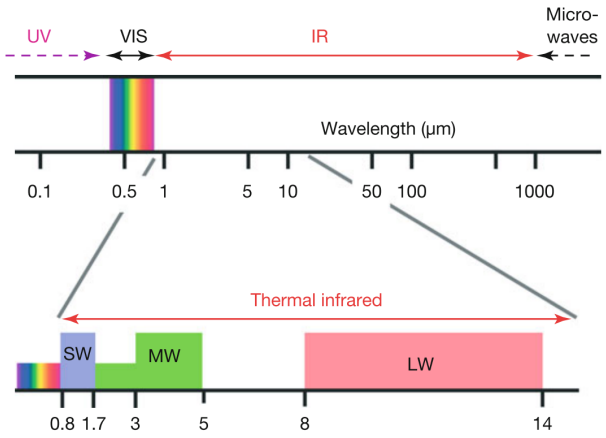


Fig. 3.4.: Wave spectrum of thermal infrared. SW marks short wavelengths, MW mid wavelengths and LW long wavelengths. [176]

its specific wavelength depends on the material itself. Any material with a temperature above the absolute minimum of -273.15° C (= 0 K) emits radiation. The intensity and wavelength depend on the material.

The intrinsic influence of the material on the amount of radiation emitted is called the emissivity $\varepsilon \in [0, 1]$. It is a fraction of the maximum possible thermal radiation from an object at a given temperature and wavelength. A blackbody is theoretically a perfect emitter. It radiates at the maximum for a given temperature and wavelength, absorbs all other radiation from any direction and wavelength, and acts as a Lambertian radiator where the radiation is equal in all directions. Blackbody radiators are often employed as calibration devices. Real world radiators cannot be built to be perfect blackbodies and therefore have $\varepsilon < 1$. The emitted radiation does not have a single wavelength. Radiation is given by a spectrum with a specific peak at a certain temperature. Blackbody radiation reflects the theoretical possible radiation spectrum, while other materials fall below it. Figure 3.5 by [172] shows the radiance $E(\mathrm{T})$ of blackbodies at different temperatures. The maximum radiance for room temperature (300 K) has a wavelength of about ~9.7 µm.

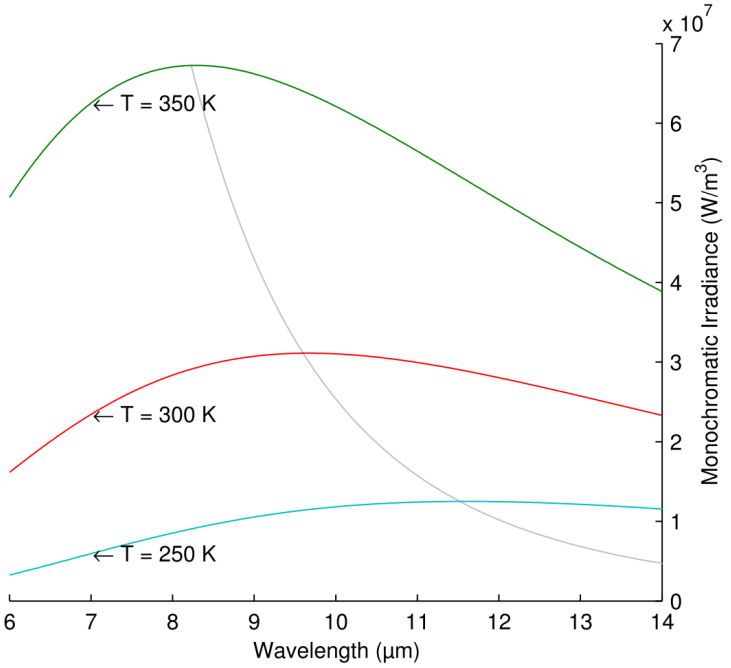


Fig. 3.5.: Radiance power of a blackbody radiator at different temperatures. For each given temperature, the radiance power at different wavelengths is given. The gray curve connects all maxima of the individual temperature curves. [172]

The emission of objects is influenced by various properties. The emissivity is different for each material, angle, temperature and wavelength. Objects also reflect, absorb, or transmit (transmittance factor τ) other radiation. The total amount of energy received by an object through radiation must be conserved

and consists of all three parts. These effects will interfere with internally emitted radiation. Figure 3.6 by [172] illustrates the three sources of radiation captured by an infrared camera: Emission from the object E_{obj}, reflected emission E_{refl}, and emission from the atmosphere E_{atm}. The total radiation detected by the camera is the sum $W_{\text{cam}} = E_{\text{obj}} + E_{\text{refl}} + E_{\text{atm}}$. The transmittance for a path is estimated by $\tau_{\text{atm}} = e^{-a \cdot d}$ with the absorption constant a [km^{-1}] and the distance d [m] the radiation must pass through. Blackbodies have zero transmittance and reflectivity and maximum absorption.

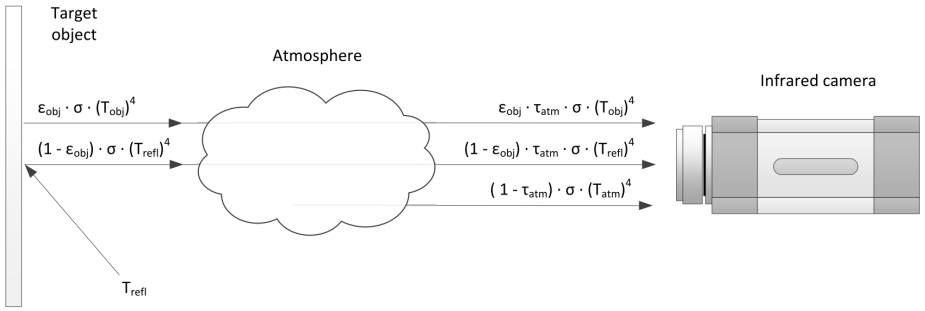


Fig. 3.6.: The infrared camera receives radiation from three sources: emitted radiation, reflected radiation, and atomospheric radiation. [172]

Applying the Stefan-Boltzmann formula ($E(A, \mathrm{T}) = \sigma \cdot A \cdot \mathrm{T}^4$) to the maximum radiation of a blackbody at a given temperature T, a surface area A and the integration constant σ, W_{cam} can be rewritten as (3.12) based on the temperatures with respect to their source (object temperature T_{obj}, reflected temperature T_{refl}, and atmospheric temperature T_{atm}). The effect of absorption is not reflected in Figure 3.6, it will reduce the amount of reflected radiation.

$$W_{\text{cam}} = \tau_{\text{atm}} \cdot [\varepsilon \cdot E(\mathrm{T}_{\text{obj}}) + (1 - \varepsilon) \cdot E(\mathrm{T}_{\text{refl}})] + (1 - \tau_{\text{atm}}) \cdot E(\mathrm{T}_{\text{atm}}) \tag{3.12}$$

Emissivity depends on the angle to the surface normal. There is a big difference in the behavior of non-conducting and conducting materials (mostly metals). From 0 to 45° both show very small changes in ε which are often neglected. However, the non-conducting materials show a fast decrease of ε afterward, while the conducting ones counterintuitively increase their emissivity (see figure 3.7 by Vollmer and Möllmann [176]). In the case of this work, mostly non-conductors are measured, except for the aluminum calibration pattern, which has a non-conductive coating. The temperature dependence of the emissivity, which is also material dependent, must also be considered.

This is especially true for changing aggregates of a material or for large temperature changes.

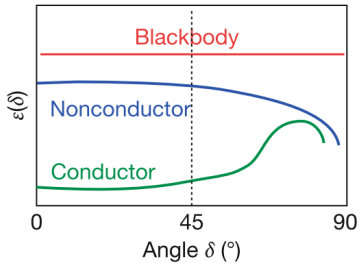


Fig. 3.7.: Schematic change of emissivity ε vs. viewing angle in relation to the surface normal for non-conductive and conductive materials. [176]

3.2.2 Thermal Imaging

Thermal imaging is the process of capturing heat radiation without direct contact by various methods. Thermal imagers use the same principles as conventional cameras (pinhole camera model) [176]. Radiation within the visible light spectrum can be detected with a standard camera. However, the analysis of thermal images requires a different approach due to its unique spectrum. For the purposes of this thesis, the thermal band will be long wave, as this range covers the thermal radiation emitted by human subjects at room temperature and can also be detected by certain cameras and optics. However, not all of the radiation of a single band is detected, but the majority, including the peak radiance point.

Thermal radiation has different characteristics than visible light. A major problem is the inability to see through glass, as this material has a very low transmittance in the target band. Standard optical glass lenses have low transmittance and block all thermal radiation in our target wavelength range. Therefore, the lenses must be built from other materials, typically germanium-based. In addition to the optics, the camera's photosensor itself must be sensitive to the specified range. A visible light sensor usually detects radiation from 120 nm to 850 nm (visible light and near infrared radiation), the thermal camera in this work: 7.5–14 µm (long wave infrared radiation).

A common method for detecting thermal radiation is the bolometer technology. Besides bolometer technology, there are other infrared detectors such as thermocouples or photodetectors. Bolometers were first developed in 1880, and

the first uncooled focal plane arrays (FPAs) appeared in the late 1980s [118]. The principle of thermal radiation measurement is that each bolometer cell is heated by the radiation received from the receptor band. The accumulated heat is converted into an electrical signal that is amplified for further processing. A schematic bolometer cell is given in figure 3.8 by Thomas et al. [167]. The process takes about 8–10 ms to get a stable result. FPAs are uncooled devices that must compensate for internal thermal differences with their readout circuit. There are several sources of noise in this process. The noise equivalent temperature difference (NETD) is defined by Niklaus et al. [118] as the difference between two identical blackbodies captured by a camera. In other words, the lower the NETD, the more sensitive the camera is to small temperature radiation differences that can be detected in conjunction with a noisy signal. It depends on the $1/f$ noise of the signal source itself. Johnson noise (thermal noise) also accumulates in the total noise. This type of noise is introduced by the electrons of the circuit interacting with the material and affecting the thermal state. Another type of noise is introduced by interaction from the environment, such as heat transfer from the camera body. Finally, the readout circuit introduces noise through its electronic components. Typical FPAs achieve a NETD of about 40 mK [183], while low-cost devices perform worse with higher NETD and lower frame rate (~70 mK NETD and ~9 Hz frame rate) [174].

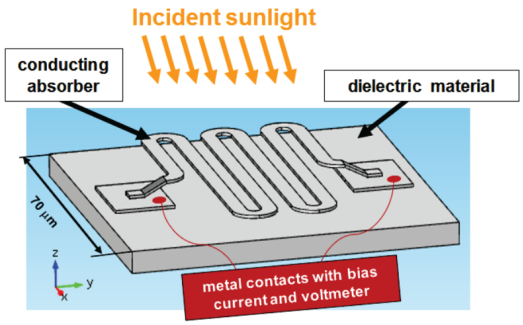


Fig. 3.8.: Principle of Bolometer Technology. Radiation is received and heats the conductive absorber material. The heated absorber generates a current between the two connections to the dielectric element. [167]

Bolometer technology is limited by the rolling shutter readout of the entire bolometer FPA. Figure 3.9 by [@2] shows the difference between global and rolling shutter readout of image sensors. In the global case, all sensor pixels are illuminated and evaluated at once. In the rolling case, a time scheme is used, each row of the sensor pixel array starts its integration time, and therefore its readout is slightly time shifted, so the total time for a full frame

is higher than the integration time. Rolling shutter has some disadvantages, especially for moving objects. If the object is faster than the integration time, then the world points will be detected by multiple pixel points, resulting in visual effects such as misplaced, enlarged, shrunken, or deformed object projections.

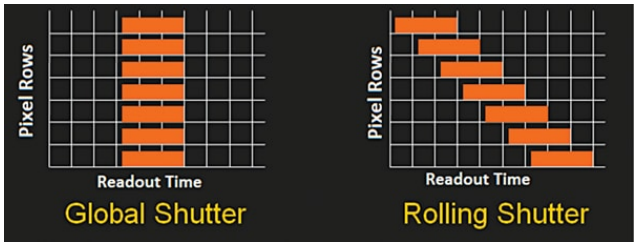


Fig. 3.9.: Global shutter and rolling shutter readout scheme. [@2]

With the new spectrum, the images are not recognizable as ordinary images and have a different look and feel than images that capture visible light. Nevertheless, the imaging principles remain the same. The objects themselves emit the main source of thermal radiation, rather than reflecting light from other sources. Features can disappear completely or appear in homogeneous regions as the heat distribution within the object changes. Commonly detected features such as object edges also behave differently. Edges are only visible if the foreground and background have different radiation. If they are similar, e.g. the measured object has the same temperature as its surroundings, no edges can be distinguished. In [172, 176] the authors explain in detail how images are formed in thermal imaging.

Unusual feature behavior in computer vision applications requires rethinking and adapting existing approaches. Thermography is widely applied in non-destructive assembly inspection, surveillance, building and infrastructure control, and many other industrial applications [176]. However, the integration of complex detection systems is currently insufficient. In current medical applications, such as breast cancer detection, precise identification of at-risk locations is critical. While image analysis is generally performed manually, as explained in [61, 134, Chapter 12], there are significant efforts by research teams to automate the analysis of medical thermographic data, as shown in a recent meta-review by Magalhaes et al. [107]. Breast cancer detection and localization is the most commonly studied application. The reviewed publications typically present a hand-crafted algorithm with a narrow focus and lack of generality. In addition, the acquisition protocol relies only on high-quality

still images without any motion. The standardized recording process results in high-quality input images that are easier to analyze due to the consistently positioned persons, with thermally neutral background. Therefore, in sports science studies, persons are often analyzed in predetermined poses without movement, as seen in [21, 162]. But in sports, movement is essential. This thesis develops a new approach to the analysis of thermograms of moving people.

3.3 Time of Flight Camera

In computer vision, there are several approaches to recover the 3D information of a scene and its object from camera images. During the projection process of the camera image acquisition described above, the 3D information is mapped onto a 2D image plane using projective geometry. However, many applications require knowledge of the position, orientation, and shape of objects within a scene in 3D and over time. In general, there are two groups of methods based on the main characteristics of the approach: active and passive. In active methods, the measuring device emits electromagnetic waves that are reflected by objects and scene backgrounds. A camera captures the process and computes a depth map to create 3D objects. Various approaches implement this principle, such as structured light projected on the target object to reconstruct the 3D image from the distorted projected points on the object surface [49]. Another approach is to measure the time it takes light to travel from an emitter to a light receiver (camera) [66]. Knowing the speed of light and the phase of the emitted light, the distance can be calculated. This technique is called time-of-flight (ToF). Passive approaches include the approach structure of motion [178]. It takes multiple images of a moving object or a moving camera and matches object points based on their features to estimate the shape. Stereo vision is also a way to estimate 3D information. Based on epipolar geometry, depth information can be triangulated with a fixed pair of cameras observing the same object from different poses but with overlapping fields of view [53].

According to Horaud et al. [66], there are two different principles for ToF cameras: pulsed light and continuous wave. ToF cameras are less accurate than laser-based cameras and often come with 2D detector arrays like other cameras. Both ToF methods consist of light emitters and receivers. The former emits light pulses that, when reflected from an object, are detected by the

receiver's photodiodes, which are also coupled to high-precision time sensors. Time is measured from a common start time, when the light was emitted, and an individual pixel receive time. Distance is recovered from half the time and the speed of light. The other approach demodulates a phase shift to estimate distance. A modulated light pulse signal is emitted and the detectors measure the reflected light four times to recover the phase. Depending on the distance of the objects, a phase shift occurs that is used to calculate the distance. The method can also detect unmodulated background waves and remove this offset, reducing errors and noise. ToF cameras are typically modeled as pinhole cameras and can be used in conjunction with other modalities.

PART II

METHODS

4 Experimental Hardware Setup

This chapter describes the experimental hardware setup in which the measurements of this thesis are performed. Data acquisition is performed by infrared thermography (IRT) and visible light (VIS)+depth cameras and other hardware components. They are part of the first step of the ThermoNet pipeline (figure 4.1). Radiometric calibration is required to convert IRT camera pixel intensities to temperature values. A two-point calibration body and a single frame calibration routine are presented. In addition, the sensors for gold standard measurements are introduced.

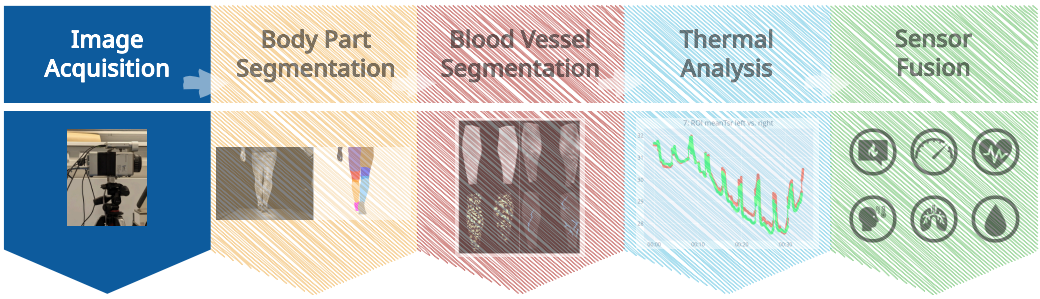
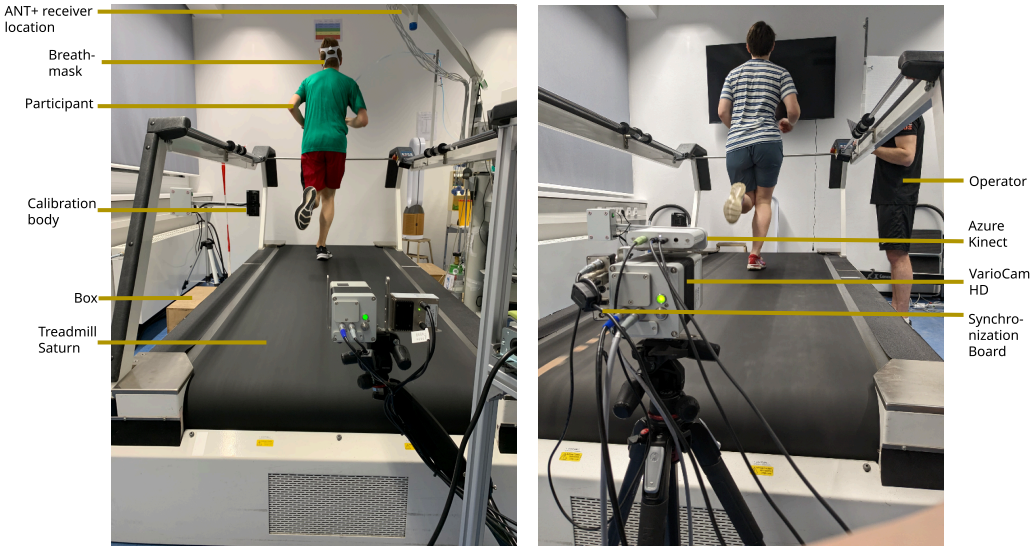


Fig. 4.1.: The hardware description is part of step 1, the acquisition step, of the ThermoNet pipeline.

Most studies are conducted on a treadmill to perform walking or running protocols. The treadmill model is a Saturn, h/p/cosmos sports & medical GmbH, Germany. Its speed can be manually controlled in 0.2 km/h steps up to 45 km/h. It is also possible to change the elevation up to +25%. The treadmill does not have an interface to report its current speed to a computer interface. Instead, speed of the treadmill belt is measured by a radar sensor (see section 4.2). The speed sensor is placed over the end of the treadmill. On the right side of the treadmill there is a place for the experimenter who is in charge of the trial. In addition, the spiroergometer control system is placed on the right side along with the control and measurement computer for the system. The medical staff taking blood samples for lactate measurement stand on the left

side on a small box. An ANT+[1] receiver platform is also placed on the top arm to pick up multiple ANT+ devices simultaneously and record their data to a custom management computer. The receiver platform is custom-built by OptoPrecision GmbH. Figure 4.2 shows two people running on the treadmill. (a) shows a participant wearing a mask, the calibration device on the left side, the box in front of the calibration device and the thermal camera in front of the image. (b) features the stereo system described in section 4.1.4.



(a) IRT system with additional sensors.

(b) Stereo system with VIS+depth and IRT camera.

Fig. 4.2.: Pictures of people running on the treadmill, including the hardware setup.

4.1 Camera Setup

Several cameras are rigidly mounted on a rack and placed on a tripod for imaging. Two different models of thermal cameras are utilised.

4.1.1 VarioCam hr

The first IRT camera is a Jenoptik VarioCam hr head with an uncooled focal plane array (FPA) of microbolometer sensors. 25 fps (frames per second) are

[1] ANT+ is a proprietary wireless communication protocol for exchanging data, such as readings from sports wearables, over a low-power 2.4 GHz network [@1].

captured with a fixed exposure time of about 8 ms with a rolling shutter. This camera connects to a control computer via a FireWire 400 (IEEE-1394) port. The manufacturer's software runs on Windows XP systems. Unfortunately, the radiometric images are saved as normal JPEG files. The original information could be affected by JPEG compression, and the depth is limited to 8 bits, even though the camera provides 16-bit A/D conversion of its sensor data. In addition, the saved thermogram contains overlaid information: temperature scale, current time and a logo. The acquisition can be controlled by the included software by setting emissivity, temperature scale and other properties. Images are stored as individual files with ascending index numbers. Since the image generation is performed with a fixed setting, the thermogram scale or other radiometric properties cannot be changed afterward. Radiometric calibration was not available when this camera was in operation. The camera periodically performs a nonuniform calibration (NUC) for about 250 ms during which no images are captured (see section 4.1.5). NUC time points are not stored in the data. The spatial resolution is 640×480 pixels. The thermal resolution is defined as 20 mK with a systematic offset error of ± 2 K at 20° C. The attached lens has a field of view of 32×20° C and an autofocus range from 0.3 m to infinity. The camera does not support a current API and cannot be integrated into our hard- and software system. However, previously recorded data can be analyzed with the methods presented in this thesis, but not synchronized with other sensory data.

4.1.2 VarioCam HD

The second thermal camera model has superior characteristics. It is from the same manufacturer, Jenoptik AG, Germany, and is distributed by InfraTec GmbH, Germany. The sensor is also a FPA microbolometer with rolling shutter and 8 ms exposure time per row. The pixel row readout starts at the bottom row. The total image acquisition time is about 30 ms. Therefore, the frame rate is 30 fps. The resolution of the sensor is 1024×768 pixels. Connectivity is provided by a GigE Vision interface (ethernet connector) and a combined control&power cable, both with LEMO connectors [@15] on the camera housing. The thermal sensitivity is 20 mK and the offset error is ± 1 K at 20° C. The lens (focal length 30 mm) has a field of view (FOV) of 32×24°, autofocus from 0.3 m to infinity and an aperture number of $f/1.0$. The resolution for each pixel is given by the instantaneous field of view of 0.57 mrad. The camera also needs to perform a NUC periodically. The camera is controlled by an API,

which is packaged by our cooperation partner OptoPrecision[2]. Images are stored as 16-bit raw data (pixel intensities). Image data does not contain radiometric information, additional information such as focus setting is stored with the raw data. This thermal camera will be integrated with other systems by an acquisition software from OptoPrecision to control multiple sensors and cameras simultaneously on a single computer.

4.1.3 Azure Kinect

To capture images in the visual domain along with a depth domain, we chose a Microsoft Azure Kinect DK. It has a color sensor for the VIS spectrum and a depth sensor from a time-of-flight (ToF) camera system in an integrated housing. Both have fixed lenses with a FOV of 90×59° for VIS and 75×65° for ToF. There is also an integrated wide angle lens for the VIS camera, but it is not considered in this work because the FOVs are much more different than for the other camera. The image resolution is set to 1920×1280 for the VIS camera, as this is the best compromise between memory persistence speed and image resolution. The camera would also be capable of higher resolutions. The native ToF resolution is 640×576 pixels. However, the camera is configured to align a corresponding depth map with the visual image, pixel by pixel. In this case, it is not necessary to calibrate both cameras. The configuration supports 3 different free run acquisition modes: 5, 15 or 30 fps. Microsoft has provided an SDK to communicate with the Kinect system via a USB-C interface. Camera settings and image capture are implemented in Python.

4.1.4 Stereo System

The VarioCam HD and the Azure Kinect are physically attached to a bracket and connected via a synchronization board to form a stereo system. The Kinect is mounted on top of the thermal camera as shown in figure 4.3. The baseline between the two cameras was set as low as possible, but the housings, especially the thermal camera, are large and different, resulting in misaligned image planes. The stereo system always requires calibration to find the baseline and determine the relative position of the cameras.

[2] OptoPrecision has provided both VarioCam cameras, a pre-built acquisition system and a software development kit (SDK) to control the camera

Fig. 4.3.: The stereo hardware setup with Azure Kinect (top) and VarioCam HD (bottom) forms a fixed stereo system mounted on a tripod.

The acquisition is based on a custom acquisition software to take advantage of the hardware synchronization feature of both systems. For synchronization, the Kinect camera is set as trigger signal sender. At each image capture the trigger signal is sent via a direct wire to the attached thermal camera. The IRT camera receives the signal and also initiates the image capture process. Both cameras are capable of capturing images at 30 fps. However, in synchronous mode, the frame rate cannot be achieved and must be reduced. The Kinect only supports three fps modes, while the VarioCam HD supports more fine-grained fps modes within its fps range. Therefore, the best matching frame rate is 15 fps.

The acquisition module consists of three loosely coupled applications as shown in figure 4.4. The loosely coupled system was chosen to facilitate thread synchronization by independent processes and to avoid performance issues. Image acquisition involves retrieving data from the VIS camera and the corresponding pixel-aligned depth map. Together with the grabbed thermogram, the data is prepared for serialization, along with image dimensions, acquisition timestamp, and focus state. The constructed data is published via a publish-subscribe messaging system (see section 8.1) to the stereo image channel (figure 4.4) and consumed by a recorder application that receives all messages and asynchronously saves them to a compressed file with the lz4 algorithm [@5]. A third application also consumes the published image data for visualization for the operator. Images are displayed in a live view, and the user can send commands to the camera and recording system. The 16-bit thermogram would be unrecognizable due to the large range of values, so

the image is transformed to an 8-bit representation with manually defined clipping ranges. Live transformation does not provide temperature-calibrated thermograms.

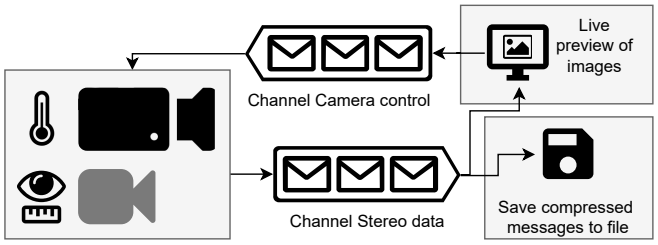


Fig. 4.4.: Three applications are loosely coupled via a message bus for image retrieval, image storage, and live image visualization.

4.1.5 Two-Point Radiometric Calibration Target

The thermal imager must be calibrated for radiometric measurements. Ring and Ammer [145] describe a standard protocol for obtaining optimal results from an uncooled thermal camera. An acclimatization period of at least 10 minutes is required. Although the cameras have an internal reference temperature source that periodically calibrates the measurement internally (NUC), it is recommended to always place an external radiator of known temperature to provide constant evidence of the reliability of the measured data. Thermograms should be constantly monitored by the operator for camera drift. Machin et al. specify more on the reproducibility of temperature measurements in [105] by defining ways to reliably calibrate, trace, and accredit to national standards. They also describe specifications for the construction of an external temperature source for a blackbody cavity. Going further, Nugent et al. [120] show how a blackbody cavity provides an external temperature source for calibration that respects the camera's internal parameters. They also describe the internal recalibration mentioned above, called nonuniform calibration (NUC). This technique is similar to the blackbody reference, but the shutter field is used as the reference instead of an external device. The advantage is that there is no lens distortion because the shutter is between the lens and the sensor. The report [91], a contribution to the National Institute of Standards and Technology of the United States of America by Lane and Whitenton, explains in detail how to calculate the true temperature of an object based on its thermal emission captured by a thermal camera. The process includes an initial calibration with a blackbody radiator, aligning the

image for consistent values for the same object, and correction. However, the authors mention that it is not always possible to obtain a true temperature due to motion blur or small object size, so signals are averaged. Measurement results and uncertainty depend on the definition of the measured object and its properties, as well as the camera characteristics. The work of Lin et al. [97] provides a novel shutterless calibration routine to overcome the problems of static NUC calibration routines for uncooled microbolometer systems when the ambient temperature changes rapidly, such as cameras mounted on unmanned aerial vehicles. Švantner et al. [161] apply the described blackbody calibration with a fixed blackbody in the thermogram. The authors are one of the few research groups to describe their calibration routine and to compare in detail the results with and without calibration in the application of a human study. Recently, Mazdeyasna et al. [113] published a comprehensive best practice guide on external factors in thermal radiation measurement of the human body. They describe the current state of blackbody correction along with the theory of infrared thermography to fully understand potential external factors such as ambient temperature, humidity, and more.

In this work, the temperature or radiometric calibration is performed with a custom-built external calibration device visible in the recent studies of this work. The device was manufactured by OptoPrecision and features a custom PID controller[3] board with NTC thermistors (TDK / EPCOS B57703M103G 10k) as sensor units (Nägele, personal communication, Jan. 22, 2024). It consists of two single black painted aluminum plates. The heating components are PID controlled with a stability of $\pm$ 50 mK and a resolution of $\pm$ 1 mK. In our projects we operate one plate at 25° C and the other at 35° C. To have similar emissivity behavior as human skin ($\varepsilon = 0.98$), we apply a black color to the plates to get $\varepsilon = 0.97$.

In addition, each plate has an identifiable marker to improve recognition during image processing. Black insulating foam is placed on top of the plate to form an ArUco marker [51] (figure 4.5c). We chose the markers shown in figure 4.5 from the set of possible markers because both have a single connected area and a less concave shape than others, which improves detection accuracy in the thermal image domain. We find the markers in the thermograms and get the average temperature deviation from the area of the markers (white area). However, in order to reduce the influence of the insulating foam, either

[3] A proportional-integral-derivative controller is a control system that has three different terms and parameters to update a controlled variable. It is widely applied in industrial applications.

by its thermal transition area at the boundaries or by the height of the foam, we take only the central parts of the marker. The selected markers combine easy detection in the thermal spectrum with a high amount of effective area for temperature averaging.

(a) Marker (id=45) associated with 35° C temperature radiation.

(b) Marker (id=37) associated with 25° C temperature radiation.

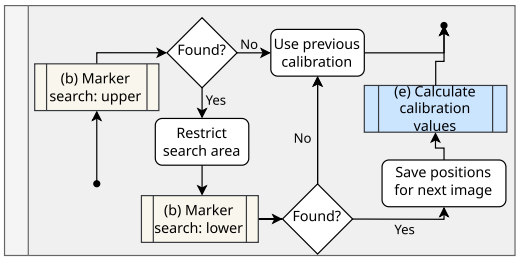
(c) Built target with insulating foam on black painted aluminum.

Fig. 4.5.: ArUco markers [51] for the temperature calibration reference system and the built device.

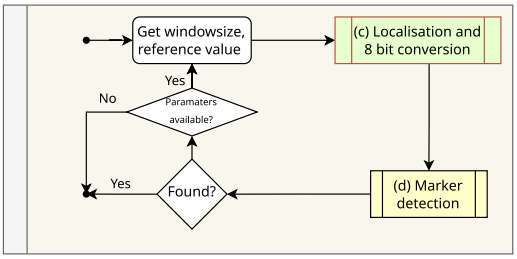
Detection of ArUco Markers in Thermograms

ArUco markers are found programmatically and receive more information about their position with respect to the camera. [51] describe the five steps of marker detection and assignment. The first segmentation step of the markers is one of the most important ones in our case. The authors employ an adaptive local threshold that works well under different lightning conditions. Since we are dealing with thermograms, which have a completely different appearance than visual images, the images have to be pre-processed and the parameters have to be optimized for successful detection. Figure 4.6 provides an overview of the marker detection. The first block (a) describes the overall routine and what steps are involved. We have a set of two fixed markers, which simplifies the marker search in later steps. In addition, we have the prior knowledge that both markers must be adjacent. Therefore, first one marker in the whole image will be found and then the search for the second marker is applied in a small area around the found marker (5$\times$ the marker size in each direction), which dramatically reduces the computation time. In addition, the found

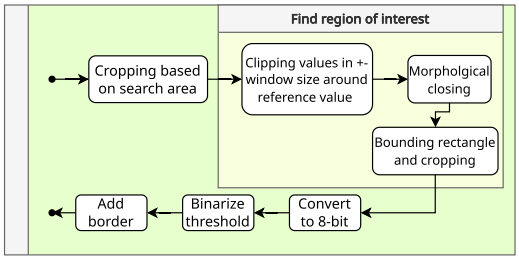
marker positions are stored as the initial search space for the next image. It is rare that a marker position changes during a measurement. However, if the search within the predefined area fails, an exhaustive search of the entire image is performed.



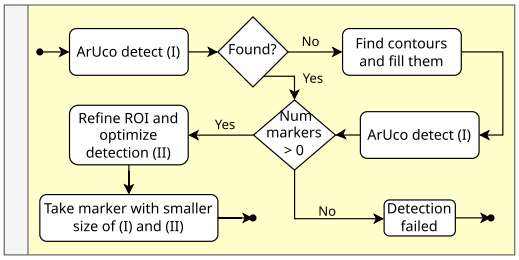
(a) General ArUco detection routine.



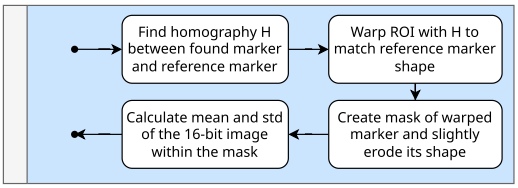
(b) Single marker search.



(c) Localization and 8-bit conversion.



(d) Detect marker.



(e) Calculate calibration values from both markers.

Fig. 4.6.: Parts of the ArUco detection algorithm.

The marker search (b) has to be sophisticated because the contrast between the foam and the aluminum isn't very high in the 16-bit image. So we have to artificially increase the contrast. This is an iterative process with different parameter sets depending on the ambient temperature or internal camera drift. The 16-bit image is clipped with a small interval. Together with a dynamic offset and the target temperature, an 8-bit image is estimated from the raw values (c). However, image features can change due to environmental or drift effects, so the interval (window size) is modified in a systematic search. The target temperature T is given by the markers. The corresponding raw pixel intensity P is not yet known, but is assumed to be $P(T) = 27315 + T \cdot 100$. To get better estimates after camera drift, the target assumption is also changed.

Therefore, a new marker search iteration will involve a different intensity assumption T. For the low temperature target (A_l) 17 values are considered[4] and for the high temperature 15[5]. The algorithm takes the assumptions in order, since the last assumptions occur less frequently than the first ones, and stops the search when a marker is found. The window size is also changed with each iteration, for the lower target the range starts with size $w_0 = 1.0$ and for the upper one $w_0 = 3.5$. The interval is changed with 15 modifiers[6]. This results in a final window size of $w = w_0 + w^*$. The full exhaustive search is built using a Cartesian product of the two sets, the window threshold w^*, and the pixel value assumptions. For the Cartesian product, all combinations with the first element of w^* are processed first.

For each of these combinations, the ArUco detection is processed, and if one of the two markers is found, the procedure is stopped. The algorithm for detecting the marker is divided into two subsequent steps: first, the rough location is found, the image is cropped, and then converted to 8-bit pixel depth with the window size and reference value (c). The second part (d) takes the localized marker and proceeds with the ArUco detection algorithm of [51]. If it fails in the first step, the image is processed more heavily. In a second step, the detected area is sharpened again and cropped around the detected marker. Marker detection is applied again to find the marker. The marker with the smallest bounding box is selected. The refinement and size criteria ensure more accurate detection results.

The last step is to estimate the calibration values (e). We are interested in the white area of the marker. However, the image size can change as well as the angle to the camera. Therefore, a homography is estimated between the found marker points and the points of an ideal marker of fixed size to normalize the real marker image. The warped real marker is converted into a binary mask and together with the generated marker the intersection areas are determined to ensure only overlapping (valid) pixels. Although the markers should perfectly match their generated counterparts, they do not. Subsequent erosion of the mask further shrinks the masks to reduce thermal effects at the foam boundaries. The final mask is applied to the original 16-bit thermogram to calculate the mean and standard deviation (SD) of that area. The mean represents the calibration value of the corresponding reference temperature.

[4] $T \in [25.0, 25.5, 24.5, 24.0, 26.0, 23.0, 22.0, 21.0, 26.5, 27.0, 27.5, 20.5, 20.0, 19.5, 19.0, 18.5, 18.0]$
[5] $T \in [35.0, 33.5, 33.0, 34.0, 34.5, 32.0, 31.0, 36.0, 36.5, 37.0, 37.5, 38.0, 38.5, 32.5, 32.0]$
[6] $w^* \in [0.0, 0.1, -0.1, 0.2, -0.2, 0.3, -0.3, 0.05, -0.05, 0.4, -0.4, 0.5, -0.5, 0.6, -0.6]$

With both calibration values, the image can be converted from 16-bit raw data to a thermogram with known radiation temperatures.

Temperature Scale Applied to Thermogram

In this work, we do not directly calculate the total emitted radiation W_{total} as described in section 3.2, since this approach does not provide a simple measurement without collecting additional information about the environment. Instead, we measure two well-known reference points with similar emissivity to humans within our target temperature range and then linearly scale the pixel intensities to the given constraints. However, our assumption only holds for small temperature scales around the ambient room temperature of 20° C and small distances from the object to the camera, so that the atmospheric influence can be neglected [172]. For large temperature ranges, such as in industry, this may not be applicable due to the temperature dependence of the emissivity or larger distances, which do not allow the removal of the atmospheric influence. In this work we do not analyze large temperature changes, our range is within 10–40° C and will be set span 10° C, where no aggregate change appears and no emissivity change for skin is reported. We define two temperature ranges: the target temperature scale, which is the lower T_{t_l} and upper T_{t_u} boundary for the 8-bit image, and the calibration scale (T_{c_l}, T_{c_u}), which defines the two reference temperature measurements as well as the real pixel values (P_{c_l}, P_{c_l}). Now the pixel values P_{t_l} and P_{t_u} of the target range are defined as $P_t = f(T_t) = m \cdot T_t + b$ with

$$m = \frac{P_{c_u} - P_{c_l}}{T_{c_u} - T_{c_l}}; \quad b = P_{c_l} - m \cdot T_{c_l} \tag{4.1}$$

The 16-bit raw image data from the thermal camera is clipped by the lower and upper ranges. The area in between is linearly scaled by 256 discrete values and stored in an 8-bit image format as the final thermogram.

In addition, we also convert precalibrated images (VarioCam hr) from larger temperature scales (e.g. 23–39° C) to a smaller target range (e.g. 25–35° C). The algorithm involves converting from temperature to pixel values, clipping the image at the lower and upper boundaries, and scaling the image to 8 bits to fit the new range.

4.2 Additional Sensors

Various sensors provide additional data during the exercise that is not measured by the cameras. For this work, we acquire, store, and transform the sensor data in a common processing system on a single machine. Many of the sensors are wireless and communicate using the ANT+ protocol. Our partner OptoPrecision has developed a device that captures all the messages from the various ANT+ devices and stores them with system time, sensor type, value, and received signal strength indication. Data packages can be received with a time window of 150 µs. Other sensors have different communication approaches and must be handled accordingly. The fusion of all data is covered in chapter 8.

Heart Rate Monitor A Polar H10 heart rate monitor, manufactured by Polar Electro Oy, Finland [@23], is worn on the chest by the participant. The manufacturer does not provide technical details of the device except the principle of measuring the polarity change of the heart. The accuracy depends on the position of the chest strap on the body and the amount of other disturbances to the electronic measurements. The sensor is able to provide its reading via ANT+ protocol.

In-ear Sensors Two Cosinuss° One sensors from Cosinuss GmbH, Germany [@7], provide heart rate, core temperature, and 3-axis accelerometer data. The device attaches to one ear and measures data inside the ear. Data is transmitted via ANT+. The sampling rate is given as 100 Hz. Heart rate is measured with an optical sensor based on photoplethysmography with an absolute middle deviation of ± 1 bpm and a range of 40–220 bpm. The temperature is measured by a resistance sensor (Pt1000) with a precision of ± 0.1° C and a range of 0–50° C. The ANT+ data rate is 4 Hz. Two sensors are attached to both ears.

Contact Sensors In addition, three CORE sensors by greenTEG AG, Switzerland, can be attached to a belt to measure and approximate core body temperature by measuring the temperature directly on the skin. Core body temperature is estimated using proprietary algorithms that have been validated by [40]. The manufacturer claims a mean absolute core temperature deviation of 0.21° C [@6]. No further technical information is provided. Data is also available via ANT+.

Environmental Sensors A wired sensor monitors the room environment with room temperature and humidity. The DHT22 (AM2302) sensor from Aosong Electronics Co., Ltd. [@12] integrates both measurements in a single device with a polymer capacitor. The temperature accuracy is $< \pm$ 0.5° C. The humidity sensor has an accuracy of $\pm$ 2% relative humidity. The measurement sampling period is given with an average of 2 s. The sensor is placed on one side of the treadmill.

Speed Sensor The existing treadmill does not have an interface that can be interacted with to get current speed and incline information, nor can it be adjusted. Therefore, our partners implemented a radar sensor to measure the current speed externally. The model is the Speed Wedge MKII from MSO Meßtechnik und Ortung GmbH, Germany [@17]. The specified speed range that can be detected is from 0.8–200 km/h at a distance of 100–700 mm under ideal conditions. In our setup, we installed it on the side above the treadmill belt at a distance of about 350 mm from the belt. The speed data is updated at a frequency of 20 Hz. The technology is based on a 24 GHz radar transmitter and utilizes the phase shift of the reflected radar signal (Doppler effect) to estimate the speed of the reflector relative to the transmitter. The sensor is connected to the control computer and messages are received through the serial port.

Spiroergometry The spiroergometry reference measurement is performed with the cardiopulmonary exercise testing (CPET) system Ergostick^{TM} from Geratherm Respiratory GmbH, Germany [@9]. It consists of three main sensors and a corresponding data processing engine “BlueCherry”. The flow sensor for respiratory flow analysis has also a differential pressure measurement[7]. The accuracy is $\pm$ 3% or $\pm$ 50 mL/s and the maximum range is $\pm$ 16 mL/s. A second sensor estimates the amount of oxygen in the gas with an accuracy of $\pm$ 0.1 vol% with an electrochemical cell. Infrared spectroscopy detects CO_2 with an accuracy of $\pm$ 0.1 vol%.

Lactate To access the lactate concentration, the study operators take blood samples (~20 µl) from the earlobe in standing phases. After the experiment, the samples are analyzed with a device from EKF-diagnostic GmbH, Magdeburg, Germany, to determine the lactate concentration. The lactate analysis for the

[7]The volume flowing through a cross-section in the specified time is given in L/s.

estimation of the individual anaerobic threshold (IAT) is performed with the application LC-Lactat from mesics GmbH, Münster, Germany [@10].

Core Temperature Pills The telemetric pills eCelsius-Performance from BodyCAP, Hérouville Saint-Clair, France [@3], continuously monitors gastrointestinal temperature. A disposable electronic capsule (pill) measures core temperature in the digestive tract. The measurement time depends on the person and how long the pill remains in the body. It is ingestible and communicates via 433 MHz to a nearby gateway that collects data for later export. The sampling period is set to 15 s. The accuracy is given by 0.1° C in a range of 25–45° C.

4.3 Computer System

The cameras and sensors are connected to a computer for data storage. The developed models are also trained and inference is applied on this device. The relevant specifications are:

CPU AMD Ryzen Threadripper 3960X (24× 3.80 GHz)

GPU NVIDIA TITAN RTX 24 GB

RAM 64 GB DDR4

Network 3× Ethernet Network (2.5G, 10G, 1G)

USB 1× USB-C 3.2, and 1× USB-C 3.1 gen 2

As storage devices for studies, two 20 TB RAID-0 devices (WD My Book Duo) connected via USB-C with a maximum data transfer rate of 290 MB/s. According to Smith [@26] the random write performance is 148.34 MB/s in RAID-0 mode. The stored data have different sizes. For all studies except the stereo studies, the camera images have about 1.2 MB compressed data size, sensor data size varies but takes about 50–100 B, speed sensor data have up to 60 B. Since the sensors do not have as high a frame rate as the camera, the camera is the limiting factor for saving data to disk. At 30 fps, 36 MB must be written to disk per second, which is far less than the storage device provides. The stereo system adds more data into a single save package, consuming 5.3 MB per file. However, the frame rate is reduced to 15 fps, which results in 79.5 MB/s, even less than the measured performance for random write speed.

5 Datasets

In addition to the methodological parts of this thesis, we also present several studies and datasets collected with the purpose of medical and sports analysis in this chapter. One of the major manufacturers of thermal cameras, Teledyne FLIR, has proposed a free-to-use dataset for autonomous driving, including people, cars, and other objects in public road environments [@27]. Although there are thousands of labeled objects in the images, they initially have only bounding boxes and no segmentation masks, and the objects in the thermal images are too small, unlike our images, as we cover the entire image with a single person. We focus on images containing single persons and do not consider materials and objects as they are too different from our perspective. While facial recognition is common in the visible domain, it can also provide insight in the thermal domain by relating thermal responses and facial expressions to emotions. Kopaczka et al. [89] published a specific dataset, ThermalFaceDB, with over 2500 faces annotated with 68 facial landmarks from 90 people. Participants vary their head positions and expressions in a predefined and free moving manner. The thermograms are captured with a high resolution camera of 1024×768 pixels. The whole head is visible in the image, the background is neutral. However, the images are taken from non-exhausted humans, which limits their relevance for medical applications. Kniaz et al. [86] presented in their work ThermalGAN, a novel work on image to image translation for visible domain to thermal domain. Along with the publication, the specially collected dataset ThermalWorld is made available, which includes over 5000 image pairs of thermal and visible images with ten classes including people, cars, and buildings in outdoor and public scenarios. The actual resolution of the persons in the image is low, and in the scenarios, people are not necessarily physically exhausted or exposing body parts for thermal imaging. Other vascular imaging datasets, such as DRIVE [160, 139] or CHASE [50] for retinal vessel segmentation, are not suitable for our purpose of analyzing thermal vessel patterns during dynamic exercise in athletes or patients. Because the properties of thermal images are completely different from those of visual images, it is not possible to directly compare images from the two domains. Hillen et al. [61] describe in their review the need for

high-quality thermograms for analysis in medical related work. However, the existing datasets do not meet the requirements of high resolution thermograms of humans during exercise when they are physically exhausted and the region of interest (ROI) is not obscured by clothing or other objects.

The following chapter describes studies for the development of our methods, which were conducted in the Department of Sports Medicine, Prevention and Rehabilitation, Institute of Sports Science, Johannes Gutenberg University Mainz, Germany. The first part describes raw datasets acquired with a thermal camera and optional additional sensors. In the second part, we present the study for the StereoThermoLegs dataset generated by the methods in chapter 7. The studies are approved by ethics committees and participants have given their consent for their data to be included in scientific research.

5.1 Medical Studies

This section describes the datasets from several medical studies involved in this work. The studies have a sports or medical research focus, but are also applicable to the development of our methods.

COMMED The study included 17 patients with cystic fibrosis [62]. The test protocol (figure 5.1) is a standard walking protocol with 3 minutes of walking, 1 minute of standing, and resuming walking at an increased speed. The stages increase until exhaustion. A final rest was also recorded. In addition, each person had three test runs on different days. The patients were 31.2 ± 11.6 years old. In addition, lactate concentration was measured with a blood probe during the breaks. Respiratory activity and calorimetry were recorded with spiroergometry. The thermal camera VarioCam hr, imaged the hind legs from a distance of about 2 m. About every minute a nonuniform calibration (NUC) was automatically applied to reset the internal thermal state of the camera. During this period of about 1 s no image is captured. The temperature scale has been set from 23 to 37° C. The images are saved as JPEG and camera information is written into the image: a timestamp, the temperature scale and a manufacturer logo. This information is blanked out before further processing. No external thermal calibration is performed.

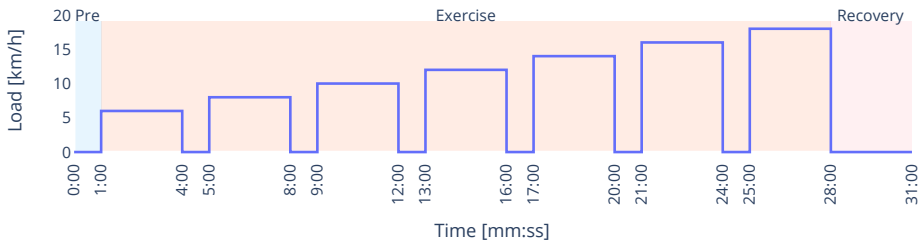


Fig. 5.1.: Example protocol with increasing speed per step and 1-minute rest between steps. Steps are increased until the user requests a stop or another stop condition occurs. There is no minimum or maximum number of steps, the number is based on the user or sensory feedback. The final stage should take place at the participant's personal maximum exertion level (`max` stage).

SPEER The SPEER study collected thermal data from 16 individuals (mean age: 23.13±4.37 years, 5 male and 11 female). The study was part of [121]. Each participant performed a single trial on the treadmill, which was captured by the VarioCam hr targeting the posterior legs. The camera was set to store no inline debug information with JPEG images at a scale of 25–35° C. There is no external radiometric calibration, but NUC was present repeatedly. Breath analysis has also been used. Respiratory and thermal camera systems are not synchronized.

LaufRad In Hillen et al. [59] the LaufRad study is analyzed. The data contains trials with 10 healthy male participants between 20 and 30 years of age. Two trials are performed for each person. A running protocol on a treadmill and a cycling protocol on a bicycle ergometer (figure 5.1), where the cycling load is given in resistance power $[W]$ and no pause is required for blood sampling for lactate measurement. The thermograms are taken with the VarioCam hr. The images are saved without any debug information overlaid. The temperature scale is set to 25–35° C, but without external radiometric calibration. The NUC is performed repeatedly. In addition, breath data, lactate and heart rate are measured and the participant's rate of perceived exertion (RPE) is assessed.

Incoreloop The Incoreloop study (not yet published) aims to analyze the thermal behavior of humans in more detail than the previous studies. Each of the 12 participants (age: 24.8±2.18) has to perform 4 trials: a standardized step protocol (figure 5.1), a long run, an alternating run followed by a steady run, and finally a steady run followed by an alternating run (appendix figure A.1).

These four trials can be compared to gain insight into the intrapersonal relationships of a single person's thermoregulatory system. The experiments also collect many other data, including respiration analysis by spiroergometry, heart rate by chest sensor, treadmill speed by radar sensor, and core temperature by in-ear sensor, lactate concentration by blood analysis, and RPE. The thermal camera is a VarioCam HD positioned approximately 2 m behind the runners. Images are stored as compressed binary files in 16-bit depth and converted to 8-bit for analysis with a temperature scale of 25–35° C. The two-point temperature calibration target is present in all images.

ThermoErgo The ThermoErgo study was conducted in the InnoSpoMed project with 15 healthy participants (age: 32.53±9.35 years) to obtain a raw dataset with the VarioCam HD along with several other sensors attached to the participants. The protocol was a stepwise running protocol on a treadmill with the camera aimed at the back legs. The temperature calibration body was present in the images. The protocol followed a standard walking protocol (figure 5.1) with a 30 s pause between each 3 minute stage for lactate measurement.

Other Trials In addition to the presented studies, we also perform individual runs or pilot studies for system analysis or image preparation. We have 10 people registered in this category. The images are taken either with the VarioCam hr or with the HD version and some of them with both cameras. The experiments with both cameras were not synchronized in time and not calibrated in a stereo system. The running protocol was different, but also follows a pyramidal design. All safety procedures were enforced, such as the RPE assessment for early termination or the medical questionnaire for pre-admission of the participant, as in the previously mentioned studies.

5.2 StereoThermoLegs

This thesis describes an additional, but not manually labeled, dataset created with the stereo transformation approach proposed in chapter 7. The StereoThermoLegs dataset (published in [6, 5]) was created specifically for the purpose of creating a new thermal dataset of people running on a treadmill while imaging their posterior legs. 14 participants enrolled in the walking protocol assessment (figure 5.2). The average age was 31.79±9.33, 8 male and

6 female participants are included. Starting with a standing phase of 30 s, followed by two walking phases of 4 km/h and 6 km/h, each lasting two minutes, continued by three running phases of two minutes each with an increase of 2 km/h per step. This was followed by a standing recovery period of 3 min. To provide a safety test procedure, the RPE was assessed at each stage and evaluated if it reached a level of 17. If it was higher, individual decisions were made to stop the study. Other medical characteristics were not evaluated. A medical questionnaire was provided to determine the participant's ability to participate. All but one participant completed the entire protocol. The outlier participant skipped the last phase due to too high RPE (18 during the second last speed phase), and we added another 4 km/h walking phase after recovery.

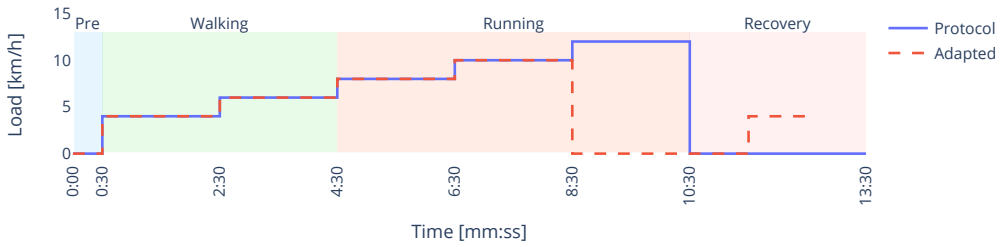


Fig. 5.2.: Exercise protocol in the StereoThermoLegs study with four regions: Pre-standing, walking, running, and recovery standing. One participant aborted the protocol due to high RPE and the adapted protocol (red dashed line) was applied.

In this dataset, the posterior legs of the participants were imaged. This is the first dataset in this thesis utilizing the multimodal stereo system with a thermal camera, a vision camera, and a depth camera (see chapter 4). The stereo setup includes time synchronization of the image acquisition from all cameras, the cameras are geometrically calibrated and their extrinsics are registered to each other. With the current hardware selection, the frame rate is limited to 15 fps. In addition, the thermal camera must perform a NUC every minute, during which no images can be evaluated. Additionally, the two-point calibration device was placed in the thermal image and set to 25 and 35° C. The Azure Kinect has a different field of view and captures the runner's entire body, while the VarioCam HD only covers the legs. A total of 178,626 infrared thermography (IRT)/color+depth (RGBD) image pairs were acquired, with an average of 12,759 images per person. For the final dataset, we selected three persons for the test set and the others as the training set, with either 3433 thermograms and 12,826 thermograms.

6 Automatic Processing of Thermograms

Object detection and classification, semantic and instance segmentation, and other high-level computer vision tasks are also applicable to thermograms, although they have a different visual appearance than common visible light (VIS) images. The algorithms do not distinguish which features they are analyzing and are adaptable to the infrared thermography (IRT) domain. However, existing methods, whether algorithmic feature analysis or inherited from data-driven optimization routines, are not directly transferable. Image features are too different to be directly comparable. Thus, parameters and optimization routines must be developed either on new custom features or on new annotated image datasets for specific tasks. As with all images, unsupervised optimization methods can be employed in future work.

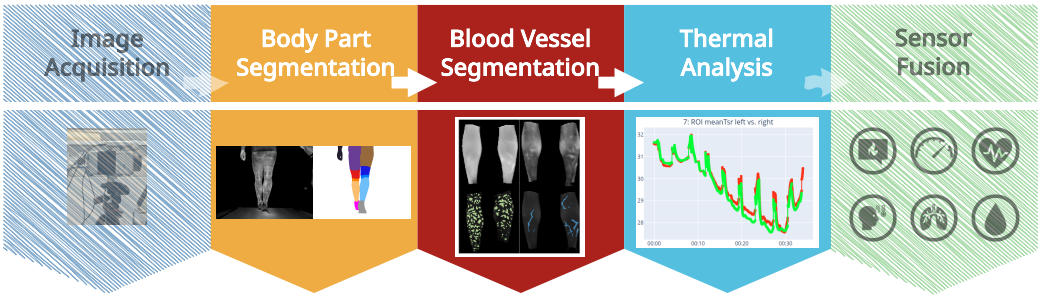


Fig. 6.1.: The focus of this chapter in the ThermoNet pipeline is on body part and vessel networks and feature extraction from thermal patterns (steps 2–4 of the full pipeline in figure 1.1).

In this chapter we will explain our methods of [60, 59, 7] in detail and extend them for more sophisticated and robust segmentation. The parts consider data from humans running on a treadmill while capturing their back with focus on the legs. The main steps shown in figure 6.1 (see also full figure 1.1) for segmenting the body parts (step 2) and finding vascular components (step 3) within the regions of interest (ROIs) are covered in the following chapters. First, the classes for each task must be defined. With the classes, it is possible to manually annotate the data for supervised machine learning algorithms.

Therefore, a specialized annotation application is developed. The next topic covers the application of machine learning algorithms to extract the body ROIs (body part network) and the vessel ROIs (vessel network). Finally, withing the ROIs thermal features are extracted (step 4). Furthermore, a concise summary of the current state of the art in human thermogram segmentation will be presented in the following paragraphs.

IRT cameras have a lower image resolution than most VIS cameras and lack low-cost, high-frequency ($>$ 30 fps) image acquisition. Real image resolution (without combining multiple images to enhance the resolution) goes up to 1024$\times$768 pixels [183, 174]. The rolling shutters of microbolometers on low and mid-cost instruments limit investigations with high speed changes. Because image analysis is similar, but IRT cameras are not widely spread, current implementations lag behind state-of-the-art computer vision methods. Nevertheless, the demand for automatic thermogram processing and applications is high. In the review by He et al. [57], the authors discuss various IRT use cases for super-resolution, object tracking, medical applications, industrial applications, and many others. The integration of deep learning methods is well recognized and adopted in the IRT field. In addition to industrial and outdoor applications, IRT is increasingly implemented in medical applications as it provides different insights of the body. The review by Parashar et al. [125] presents various aspects of machine learning in medical imaging, focusing on X-ray, magnetic resonance imaging, and other modalities, including segmentation and classification. Thermal imaging is mentioned but not explored further. Das et al. [41] developed a specialized knee segmentation algorithm, but integration and simultaneous identification of other ROIs is not possible. Bhowmik et al. [19] also work on knee inflammation detection and provide a modified region growth method for segmentation. Deep learning techniques are proposed by Magalhaes et al. [108]. A deep neural network (DNN) detects skin cancer regions and outperformed other machine learning classifiers. Another specialized thermogram single shot analysis is described in Cruz-Vega et al. [38] for the analysis of the sole of the foot in diabetic foot disease. With the work of Unger et al. [171], an automatic method based on Gaussian filters is available to analyze blood vessel perforator patterns before surgery. Again, the proposed method is not scalable and extensible to new ROIs and full segmentation masks, as a hand-crafted algorithm was developed. In addition to descriptive tasks for thermal images, generative tasks are also investigated. ThermalGAN, proposed by Kniaz et al. [86], introduces a color-to-thermal image translation method for scenes where people are performing daily ac-

tivities. The approach of Pavez et al. [126] demonstrates an image generation technique to generate face images along with different expressions in the thermal domain.

Due to the lack of suitable algorithms that meet the requirements of being applicable to moving persons, being extensible and scalable to new ROIs, segmenting multiple classes simultaneously, and being rapidly adaptable to new environments, we developed a new approach to human thermal image analysis. In our previous work by Hillen et al. [60] we describe a deep learning method for semantic segmentation of runners' posterior legs and perform statistical analysis over all images of an experiment. A comparison between the manually analyzed and the new automatic approach shows the superior capabilities of the automatic approach, including the ability to analyze thousands of images, while only about a dozen are available in the manual method. In Hillen et al. [59], the legs were further separated by their sides and ROIs within the legs. Focusing on the calves allows a more comparable result, as individual leg clothing of the persons was excluded. In addition to the ROIs of the legs, vascular-related patterns such as veins and perforators are detected. Both body region and vessel segmentation are performed by leveraging supervised deep learning on a manually annotated dataset. The publication [7] introduces the processing steps of the ThermoNet architecture (see figures 1.1 and 6.1).

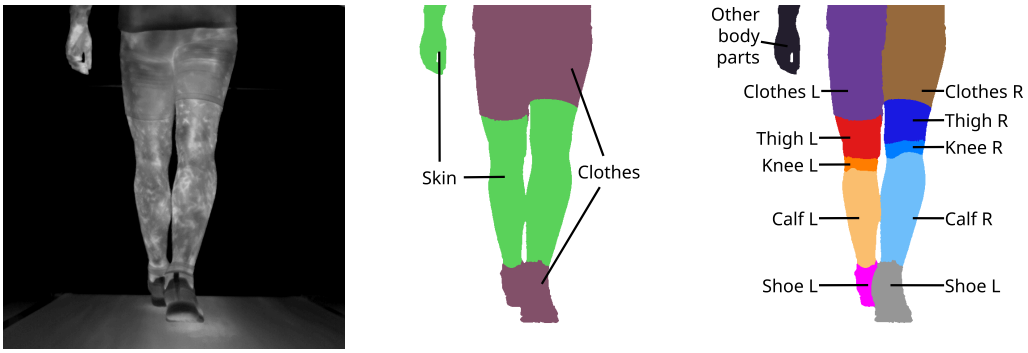
Implementation The implementation of the training and the optimization of the deep neural networks is done in `PyTorch`[8]. In addition, to structure the `PyTorch` code `PyTorch Lightning` [@8] is employed. Data loading and processing steps are implemented with `numpy` [52] and `OpenCV` [25], data augmentation with `Albumentations` [29]. The `PyTorch` library takes care of the correct automatic gradient calculation and `PyTorch Lightning` organizes and provides the training routine like the backward pass with the call of the optimizer step. The loss functions and optimizers included in this work are integrated into the `PyTorch` library (`torch.optim`) or borrowed from the code repositories of the authors of the publications. The weights for the loss functions are the reciprocal class frequencies computed once. (see appendix tables A.1 and A.3). The K-fold cross-validation (CV) utilizes `scikit-learn` [128, @24]. The hyperparameter optimization (HPO) routine is implemented with the `Optuna` framework [2]. Each trial is initialized with the same seed.

6.1 Class Labels for Segmentation

The basis of the ThermoNet processing pipeline is an annotated dataset. Two main tasks have been defined, body part analysis and blood vessel pattern detection. For the supervised tasks in our algorithms, we need to define the ROIs and associate class labels. This definition is the basis for an assistive annotation tool.

6.1.1 ThermoNet Class Definitions

In the previous work [60] we defined the classes uncovered skin, clothes (including shoes) and background (see Figure 6.2b). The class skin includes only parts where the thermal radiation is not disturbed by other clothing. There is no distinction between left and right, nor between different parts of the body such as calf, thigh, or even hands. The steps of each leg move counter-cyclically. So the legs are not in the same position at the same time. A single class for both sides would combine a straight leg and a bent leg, or a near leg and a far leg, resulting in the averaging of different occlusions, motion blur, or thermal radiation angles. To overcome this shortcoming, we introduced a more detailed definition of body part classes in [7]. First, the



(a) Thermogram. **(b)** Classes defined in [60]. **(c)** Classes defined in [7].

Fig. 6.2.: Sample class definitions (with custom color map) for a semantic segmentation mask of a physically exhausted person with multiple patterns (a). With (b), we proposed a simplified segmentation label definition in [60]. The classes in (c) describe body parts for the left (L) and right (R) sides of the posterior legs. The figure is based on [7].

left and right sides are distinguished, then shoes are separated from clothing, and the remaining skin classes are divided into body part related classes: calf

(lower leg), knee, and thigh (upper leg). The images usually do not include other parts of the body above the hip, such as the arms, head, or shoulders. Sometimes other parts of the body are visible, such as a hanging hand. We do not distinguish which additional body part is visible because there is too little data compared to the other classes. The final classes are shown in figure 6.2 as a colored mask. Each class is represented by a unique number and the masks are saved as single channel images. A complete list of classes and their color codes can be found in appendix table A.1 and appendix table A.2.

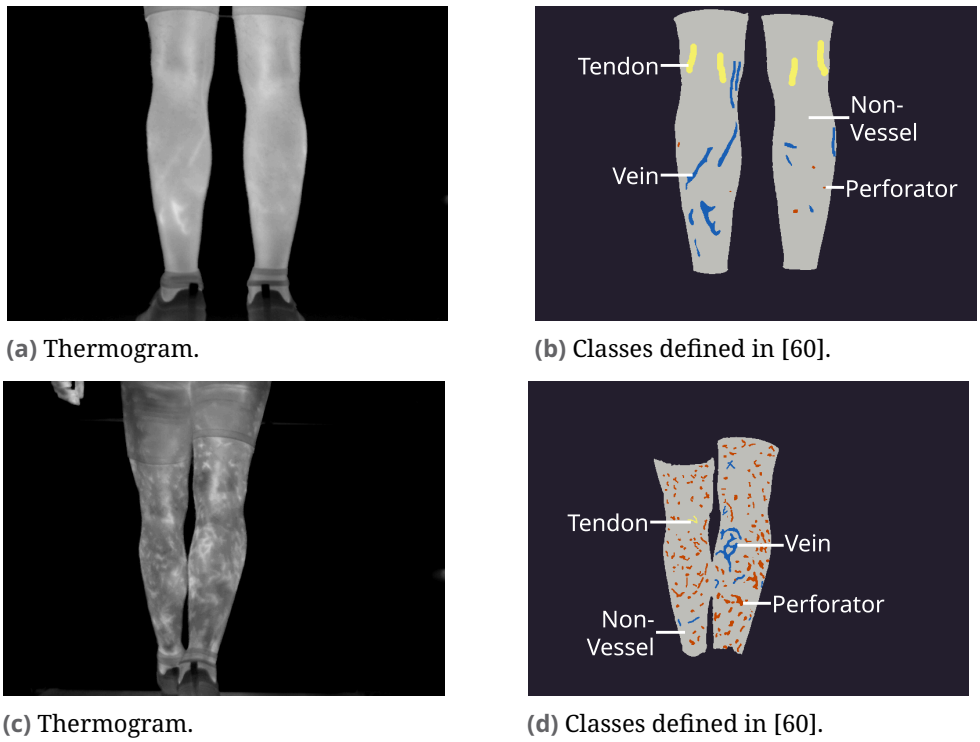
Hillen et al. [61] describe the thermal patterns associated with blood vessels (see section 2.3). Three main classes have been identified:

- Vein pattern: long and interconnected tubes
- Perforator pattern: small, snowflake or tree-like patterns
- Area pattern: indistinct area

In addition, we also define an inverse pattern as the area where no other pattern is defined as non-vessel. Based on the pattern identification, we build class labels in [60] to segment them within the thermal images. The classes are shown in figure 6.3 and the color mapping is added in appendix table A.3. Unlike the body parts, the vessels do not distinguish between left and right. In combination with the body parts, it is easy to distinguish between a single part and a specific segment. There is also no intrinsic meaning for a vessel to be on the left or right side. In addition to the vessels, we plan to integrate a tendon class, but these are not yet annotated in every image, so we do not consider them in our processing. We have omitted the areal patterns because they are most common in the shoulder regions.

6.1.2 Annotation Tools

Supervised training of artificial neural networks is the most common case when optimizing them for a specific task. Therefore, annotated datasets are crucial for the performance of the final parameter set. The more data samples available, the better the distribution of the data can be estimated and fitted into the model, thus increasing generalization performance. For many tasks, large annotated datasets with various quality control cycles already exist. However, the development of specialized applications often requires specific information from a dataset or a specialized non-public dataset, such



(a) Thermogram.

(b) Classes defined in [60].

(c) Thermogram.

(d) Classes defined in [60].

Fig. 6.3.: Example class definitions (with custom color map) for semantic segmentation of thermal vessel patterns of the same person based on [60]. (b) contains mostly vein patterns (blue) and tendons (yellow), while (d) was at the last stage of the experiment, where mostly perforator patterns are visible (red).

as medical data. Therefore, a dataset needs to be created, annotated, and curated. Depending on the task to be solved by the machine learning model, the annotation has different types. In addition, different formats are required for different neural network architectures. Both types and formats are constantly evolving. Since we are only interested in the task of semantic image segmentation, we do not discuss further inappropriate annotation types and formats such as bounding boxes, points, or categories. Segmentation associates each pixel with a category. The native label format is a 1:1 pixel class mapping or alternatively a geometric description.

There are other types of segmentation, such as instance or panoptic segmentation [84]. Figure 6.4 by [84] illustrates the types with an example of city/traffic segmentation. Semantic segmentation (b) finds all pixels belonging to a certain class, but does not distinguish between different objects of the same class. Instance segmentation (c) extends the concept by separating multiple instances of the same class into different object instances, but ignoring the background. Panoptic segmentation (d) also includes the background scene

in the segmentation, providing instances in the foreground (multiple cars, people, bicycles) and semantic clases for the background scene (like a street, buildings, the sky).

(a) City / traffic scene with several cars, buildings in the background, road and sky.

(b) Semantic segmentation of all classes.

(c) Instance segmentation with different labels for the same class.

(d) The combination of instance and segmentation type forms the panoptic segmentation.

Fig. 6.4.: Segmentation types: semantic, instance and panoptic segmentation. [84]

As the implementation of segmentation has grown, various applications for creating and curating datasets have been developed. Many include annotation workflows for different tasks, support export to different formats, and support annotation with different features such as the use of foundation models. Foundation models are trained on a large dataset and work well for many tasks, but are not optimized for custom tasks. In addition to the annotation workflow, more sophisticated data lifecycle workflows are often integrated. Many platforms are implemented as web applications to simplify application deployment and data management. A centralized system helps distribute tasks and manage multiple annotators. The Computer Vision Annotation Tool (CVAT) [39] provides an open source platform for many types of annotations, including 3D object annotations, attributes, polygons, shapes, pixels, skeletons, and more. It also includes many deep neural networks for various tasks to support annotation. LabelMe [@29] is another annotation application based on earlier work by Russell et al. [148]. It focuses on image annotation with polygons, geometric shapes, lines, and points. Classification labels and video annotation helper methods are also included. Labels are exported as structured data in JSON format or as image data. FiftyOne [@16] is an open source annotation framework that acts as a dataset management tool. It provides many integration points for custom annotation types, approaches, and export formats. It includes user management and collaboration features. Dataset analysis provides deeper insight into both annotation performance and dataset statistics (class distributions, label clustering, etc.). Direct annotation is delegated to downstream annotation applications such as CVAT. These three tools show the variety of different applications for annotation, especially in computer vision. Many other commercial platforms are also

available. However, none of the presented applications met the requirements of this work. The segmentation masks should be stored as images and as fully dense masks. However, creating dense image masks (each image pixel has a corresponding label) is very time-consuming with manual annotation. Additional sophisticated management features are not necessary for this research and were excluded for ease of development. As an additional annotation tool, the PixelAnnotationTool (PAT) of Breheret [26] allows a single annotation type: dense mask. A filling algorithm helps annotators create a dense mask from few manual annotations. The tool is not directly extensible, but its lightweight implementation allows modifications. Based on our requirements, the PixelAnnotationTool is selected and described in detail in the next section.

6.1.3 PixelAnnotationTool

The semantic segmentation task we want to achieve in supervised learning algorithms requires a dense segmentation mask as a ground truth label of the input image. However, manually painting each pixel with high accuracy to a corresponding class is tedious when labeling is done with simple painting tools. The target patterns, especially the perforator patterns described later, are small patterns in the image. Therefore, we decided to use pixel-level annotation masks. Based on our research and with ease of use in mind, we decided to use the PAT form Breheret [26]. It is based on a three-layer image painting tool: the base image, a manually drawn mask, and a dense watershed mask on top. The watershed algorithm [90] requires initial markers (manually drawn mask) for the image to generate a dense mask with multiple semantic groups. The watershed mask can be used directly as input to a DNN. In order to support the annotators in their work as much as possible, we have added more features to the software. The PAT saves the manual mask together with the watershed mask as single-channel PNG images and as colorized images for visualization. Figure 6.5 shows an example of our modified PAT. In the upper left corner, Breheret introduces the alpha setting, which controls the transparency of the overlaid manual and watershed masks on top of the original image. The resize control defines the size of the image, and the circle size defines the pen size of the label drawing. In the center of the left panel, selectable classes are placed with their names and colors. The right pane was originally just a file selector, but we added additional features.

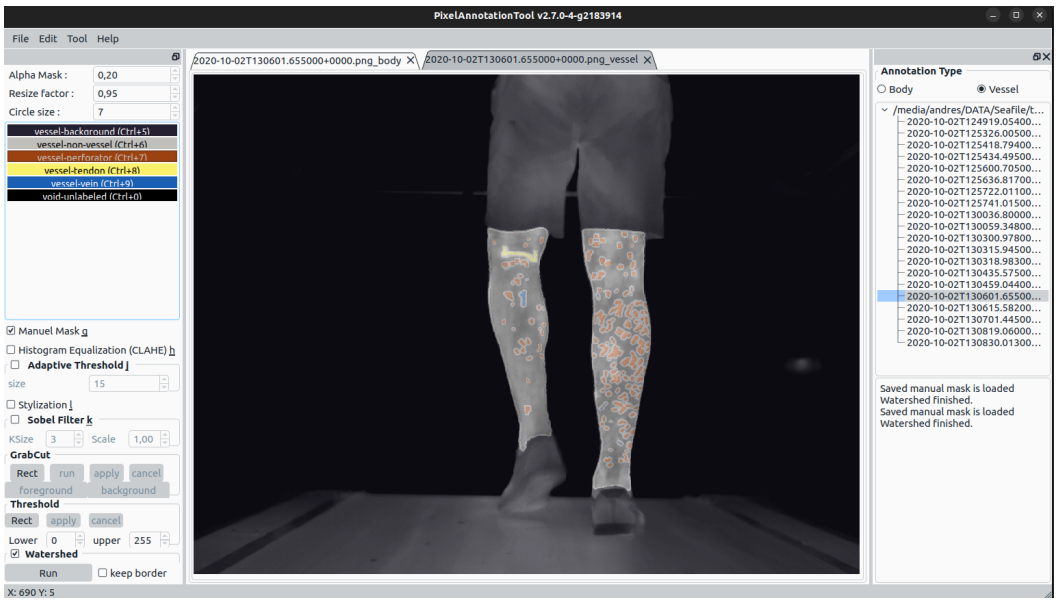


Fig. 6.5.: Modified PixelAnnotationTool with an open thermogram in vessel mode.

In our project we are interested in two segmentation tasks, the body part segmentation and the vessel segmentation. A task selector allows the user to switch directly between the two tasks, and the saved images are named according to the task. A total of eight additional files are saved per image. The task selector also automatically loads the class definitions for the task. Vessel segmentation should only remain in areas where skin is visible on the leg. However, this information is labeled in the body part mask. Therefore, the body mask can initialize the vessel mask by setting all covered leg parts and the entire background to the vessel background class. In figure 6.5 the image is loaded in Vessel mode (top right). Additional features introduced support pixel annotation. Breheret has already developed fast class label selection, image zoom, and drawing functionality. More convenient drawing features are implemented by us: A mode to remove labels, a mode to label only unlabeled pixels and leave others unchanged, which allows fast drawing of borders around other components and changing the label of a connected component. We have implemented several features to help annotators label small patterns. Previously, there were three options (bottom left): show manual mask, show watershed algorithm mask, and include/exclude watershed borders. Our application introduces several layers to increase the visibility of patterns and support algorithmic labeling (left):

- Render the image using the contrast limited adaptive histogram equalization (CLAHE) algorithm of Pizer et al. [132].

- Apply an adaptive threshold of a user-defined size (OpenCV library [25]).
- Visualizes the image by applying a stylization filter (OpenCV library [25]).
- A customizable Sobel filter [159].
- A two-point, customizable threshold mask inside a selectable rectangle allows the select class to be drawn directly on the inclusion criteria while leaving the rest untouched.
- Similar to the threshold approach, the GrabCut algorithm [147] can be applied. Within the rectangle, the GrabCut foreground and background are iteratively selected, and when the GrabCut mode is exited, the foreground is drawn with the selected class label, while the background remains unchanged.

Additional usability features have been introduced to reduce the time required to annotate images. Enhancements include drag'n'drop to load an entire folder of images (right side), a log window to verify that images are loaded and saved correctly (bottom right), and a selectable window appearance with a light or dark theme.

6.2 Deep Neural Network Segmentation of Thermograms

Segmentation algorithms can be divided into two categories: modeled algorithms and data-driven algorithms. The first category includes algorithmic analysis of images, such as edge detectors, thresholds, and connected component analysis. Data-driven techniques include deep learning algorithms and often outperform classical algorithms [116]. The review by Malhotra et al. [110] shows how image segmentation is applied in medical applications. Several algorithms are proposed to achieve segmentation of objects, structures, or other content within an image, ranging from simple processing steps to complex deep neural networks.

To automatically analyze millions of thermograms, we develop specialized but extensible DNNs to identify and segment the ROIs in each thermogram. The first task is to separate the body parts of a human in a thermogram into background and multiple body parts with the body part network (BPN). Second, blood vessel patterns are extracted and segmented by the vessel network (VN). The task is to segment the proposed ROIs of a single human in controlled

environments at once. To the best of our knowledge, direct identification and assignment of an individual vascular pattern to an anatomical structure has not been developed. Annotation of such data is also not available. Instance segmentation would come into play when multiple people need to be distinguished, but only single people are covered in this work.

6.2.1 Data Preprocessing

Data preprocessing is required to bring the images and samples into a proper format that can be handled by the systems being developed. This includes normalization and batching. In our work, preprocessing steps are performed before and after augmentation. When loading images and masks, the predefined temperature range for the thermograms is first enforced. It should be comparable within all images and is set to 25–35° C. Most of the images in the dataset are within this range. Some experiments with the VarioCam hr had a range of 23–39° C and were therefore converted to the target temperature range by linear mapping. To train the VN and focus only on the vessels, the background and clothing areas are blanked in the thermogram. To ensure that the vessel label does not have a manually incorrect background setting, we apply the body mask label for the background, clothing, and shoe areas instead. Depending on the dataset (training, validation or test), image augmentations (see below) are applied. The image is then normalized to have a mean of 0.0 and a standard deviation (SD) of 1.0 to allow for better learning performance. The normalization needs to be applied differently for body part and blood vessel segmentation. In the first case, the whole image is considered. In the second case, the background pixels are set to zero, which changes the mean and the SD. The statistics are calculated from all available study images by averaging the individual image mean and SD. For body part values, the original images are taken. For vessel values, all images must have a blanked out background. To get realistic values without the BPN, a hand-crafted algorithm (algorithm 6.1) roughly determines the legs in the image with traditional, non-data-driven algorithms utilizing thresholding and morphological operations. An example of the classic algorithm for the whole body is given in figure 6.6, along with the result of the BPN B-B (see section 9.3.1).

```
def classic_body_segmentation(image):
    """ image: Thermogram image with 8-bit depth. """

    thresh = threshold(image, method=Otsu)
    opened = morphological_open(thresh, kernel_size=(3, 3)) # Remove
        ↪ noise: 1 erosion followed by 1 dilation.
    eroded = morphological_erode(opened, kernel_size=(3, 3), iterations
        ↪ =3) # Remove larger speckles.
    dilated = morphological_dilate(eroded, kernel_size=(3, 3),
        ↪ iterations=3) # Fill the holes.
    mask = binary(dilated) # Make sure mask has 1 for leg and 0 for
        ↪ background.
    masked_image = image & mask # Bitwise AND of mask and the image.

    return masked_image, # Mask applied to the thermogram.
           mask # Leg mask.
```

Alg. 6.1: Pseudocode to roughly separate the legs from the background.

In addition to the images and masks, several additional masks are prepared for different network architectures or losses. These are a distance map of the labels to the background, the algorithmic body segmentation (see algorithm 6.1), and the original unmodified image.

(a) Thermogram. **(b)** Body skin segmentation with algorithmic approach. **(c)** Semantic segmentation of all classes with current BPN B-B.

Fig. 6.6.: Algorithmic body segmentation vs. BPN result.

6.2.2 Data Augmentation

Supervised learning approaches require large amounts of annotated, high-quality data. However, these requirements are often not met. The amount of annotated data is often too expensive to produce or not available. Manual annotation with expert knowledge is not readily available and can be subject

to human error, such as inter-annotator differences when different people work on the same dataset and do not have the same understanding of labeling and create different masks for the same true data. Also, annotators have different prior knowledge and biases that can be reproduced in the data. Another consideration besides quality inconsistencies from an annotator's perspective is the data itself. Dataset classes may be unbalanced and thus do not represent the data distribution of the real world. The imbalance may be caused by the preparation of the dataset with an unbalanced selection of data, or it may be inherent in the data if a particular piece of information is extremely rare.

Overcoming the limitations of small and poor quality datasets is the goal of data augmentation. Many methods and techniques have been discussed to augment a dataset and improve the data size and quality, thus improving the performance of a trained machine learning classifier. According to Mumuni and Mumuni [117], a review of data augmentation approaches, several categories of image augmentation have been developed. These include algorithmic manipulations like geometric transformations (shift, rotation, zoom, cropping, linear and nonlinear deformations, ...), photometric transformations (acquisition conditions and camera properties such as distortions, motion effects, lightning, weather conditions, ...). More recent approaches include transforming images with learned functions or combining/mixing multiple images. Going further with augmentation leads to feature transformations. Instead of manipulating the input space, an internal model feature space is manipulated. Another approach involves synthetic data, where again several branches of methods can be separated. Generating data can be tedious because it requires a model that represents the desired data distribution. Generative artificial neural networks, such as generative adversarial networks (GANs) or variational autoencoders, sample images from a prior data distribution. These approaches require the generation of both images and corresponding labels. More straightforward is the use of computer graphics. A world model is rendered with different textures, one representing the captured image and the other the corresponding mask. The world model holds the information. The views can be manipulated and the rendered images will be different. Choosing the right combination of data augmentation for a custom project is not a trivial task. The data augmentation methods presented here have not been systematically tested like other architectural settings (see section 6.2.6), instead we choose them manually.

The preprocessing steps are applied to all images. The augmentations are performed only on the training dataset, not on the validation or test datasets. The figure 6.7 contains an overview of the augmentations and the probability of their application. We employ geometric and photogrammetric transformations. Augmentations are applied to the image and to the loaded masks if necessary. The augmentation is included in the training procedure and is applied when loading the data. For each augmentation, a probability is given whether the augmentation will be applied to a single sample or not. First,

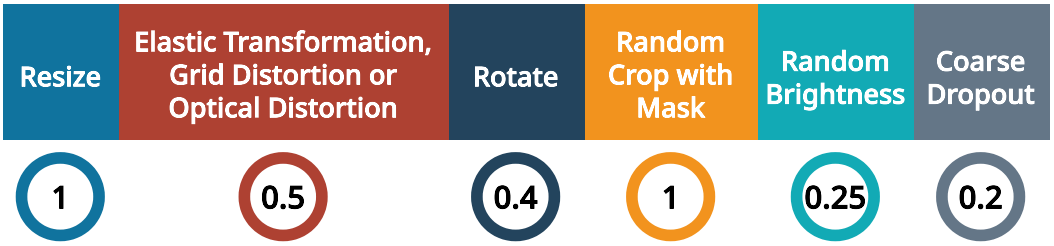


Fig. 6.7.: The pipeline shows the augmentations that will be applied with the specified probability.

all images are resized to the smallest image size in our dataset (640×480 pixels). Starting with different distortions and deformations, one of an elastic transformation [155], a grid distortion, and a simulated optical distortion is randomly selected and applied. Then the image is rotated in both directions with a probability of 0.4 up to an amount of 45°. This is followed by a random zoom with cropping. The captured thermograms often contain only the leg in the center and nothing on the sides. Therefore, random cropping could result in images that may contain nothing. Therefore, the method tries random cropping several times until a leg is included or 20 runs are evaluated. Random cropping is always applied. Random brightness changes in the images only simulate different temperature ranges and are applied with a probability of 0.25. Finally, small squares are removed from the image with CoarseDropout [42], which tries to force the network to learn structural patterns in the seen data.

6.2.3 Neural Network Architectures

In this work, we analyze several deep neural network architectures for semantic segmentation. The following sections briefly describe the architectures.

U-Net With the release of U-Net [146], the performance of semantic segmentation is greatly improved. The original application of biomedical segmentation was quickly adopted by many other fields. U-Net has an encoder with several steps of convolution, normalization, and pooling layers, and a decoder with an upscaling part. The newly introduced feature is the skip connection: the result of each encoder stage is passed to the corresponding decoder stage with the same feature map size and combined with the upscaled feature maps of the previous decoder stage. With this approach, it was possible to first obtain the main features and segment them at a high level, but also retain detailed information and integrate it to fine-tune the segmentation. The U-Net approach is a basic architecture in artificial neural networks and is often adapted and extended, e.g. Unet++ [188], Spatial Attention U-Net [82], Pyramid U-Net [184], TransUNet [35]. Each of them adds a specific feature to optimize in a particular domain.

Attention U-Net and Variants Another extension of the U-Net architecture introduces attention gates [122]. Attention gates modulate the flow of information from encoder layers to their decoder counterparts through the skip connection. The goal is to emphasize more important features and downvote unimportant features. Attention units implement a soft attention mechanism with learnable weights. An encoder result x is gated by the upcoming decoder result g from a lower level to produce a single map with weights from 0 to 1. The attention maps guide the following layers in their feature extraction. In this work, the Attention U-Net is implemented with 5 encoder and decoder stages (figure 6.8). Optionally, a pyramidal input approach is integrated by adding a resized input image to each encoder stage as described in [184]. The deep supervision approach of Pyramid U-Net [184] is also adopted. The final layer is fed by a combination of the upscaled results of all decoder stages. The loss is then backpropagated through these paths directly to the decoder stages.

Fully Convolutional DenseNet Previous investigations on leg segmentation of thermograms in a bachelor thesis [179] applied the fully convolutional DenseNet [77] to this area. Based on the promising results, the architecture was included in the design process of the BPN. DenseNet [68] introduces a network architecture block in which layers are concatenated at the end, in addition to the residual concatenation between layers in the block. With the fully convolutional DenseNet (*One Hundred Layers Tiramisu*) of Jegou

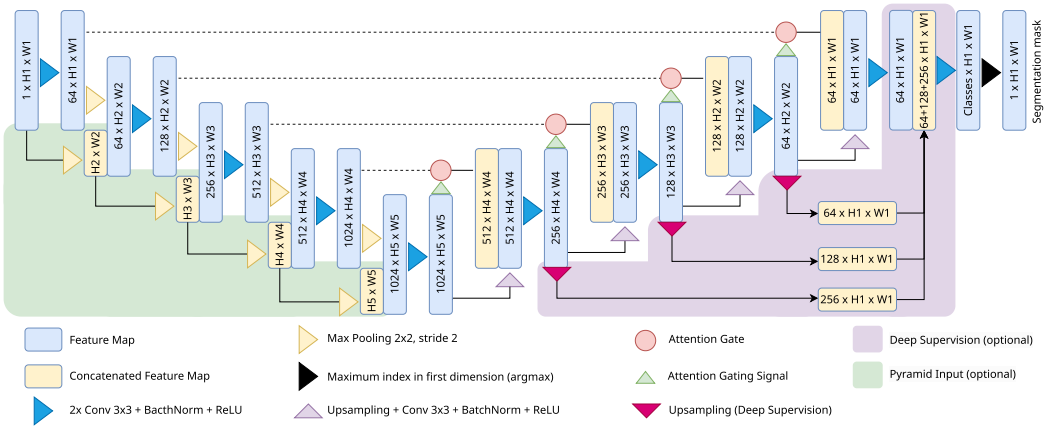


Fig. 6.8.: Implemented Attention-U-Net architecture based on [122] and [60]. Includes optional pyramid and deep supervision extensions based on [184]. The optional blocks are executed only when necessary.

et al. [77], the authors combine the DenseNet approach with the U-Net based encoder-decoder architecture. The authors propose three architectures with different number of layers, a small one (56), a medium one (67) and a large one (103). In this work we only consider the large variant, called Tiramisu103.

DeepLabv3+ The DeepLabv3+ architecture [36] combines atrous convolution with depthwise convolution to reduce the computational load while maintaining performance. Atrous convolution (also known as dilated convolution) extends the kernel with additional rows and columns. The expanded kernel has a larger receptive field with the same number of operations. (6.1) shows how a 3×3 convolution kernel is dilated at a rate of 2 to form a 5×5 kernel. Typical encoder-decoder architectures such as U-Net resize the input image for different receptive fields. Instead, the input image with the same image dimensions is processed in the encoder stage with different kernel sizes for different receptive fields.

$$\begin{bmatrix} k_{1,1} & k_{1,2} & k_{1,3} \\ k_{2,1} & k_{2,2} & k_{2,3} \\ k_{3,1} & k_{3,2} & k_{3,3} \end{bmatrix} \xrightarrow{\text{dilation rate 2}} \begin{bmatrix} k_{1,1} & 0 & k_{1,2} & 0 & k_{1,3} \\ 0 & 0 & 0 & 0 & 0 \\ k_{2,1} & 0 & k_{2,2} & 0 & k_{2,3} \\ 0 & 0 & 0 & 0 & 0 \\ k_{3,1} & 0 & k_{3,2} & 0 & k_{3,3} \end{bmatrix} \tag{6.1}$$

Transfer learning can improve performance for a task by fine-tuning a pre-trained network with custom data. DeepLabv3+ has already been trained on

the ImageNet dataset. Although the images are not from thermal spectra, the pre-trained model is considered in the hyperparameter search.

6.2.4 Loss Functions

The training objective of an artificial neural network is determined by a loss function L. Based on the task and data characteristics, such as possible class imbalance, focus on edges or coarse shape, many different loss function formulations are possible. According to the review of segmentation losses [103] they can be grouped into different basic principles, like the distribution based, region based and boundary based. In our work we consider several losses in the HPO (see below), but not all of them are described in detail because of their small impact on the final result. For the first group (distribution based) we consider the cross entropy and the focal loss. The most promising results have been shown with Dice loss and related losses Tversky, Lovasz, clDice loss from the region-based category. Boundary loss, perimeter and regional mutual information loss are considered last. A loss function L compares the ground truth label $Y = \{y_i\}$ and the predicted image $S = \{s_i\}$, where i denotes the i-th pixel of the image with N pixels. Classes are written as $c \in C$. The predicted class probabilities are estimated from the network output (logits) with either the softmax or sigmoid function.

Cross Entropy The multiclass cross entropy loss (CE) compares the two probability distributions of the class labels from the ground truth data and the logits [103]. (6.2) defines the CE based on the $\log$ probability of the predicted class $s_{i,c}$ with respect to its label $y_{i,c}$.

$$L_{\text{CE}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(s_{i,c}) \tag{6.2}$$

Additionally, a weight α_c can be introduced for each class to correct for class imbalances. Typically, class weights are estimated by class distributions from the training data.

Focal Although the weighted CE treats class imbalances with reciprocal class proportion as weights, the formulation does not consider the imbalance between easy and hard to segment samples. To extend the CE, Azhar and

Khodra [11] added a regularization factor to the loss function that emphasizes hard examples and mitigates the influence of easy to segment examples. The cross entropy term is weighted with a dynamic term $(1 - s_{i,c})^{\lambda}$ (6.3), which takes into account the probability of the predicted sample to identify hard and easy samples. High probabilities of the predicted sample are treated as easy predictions, resulting in a small dynamic factor that reduces the loss contribution of the sample. Low probabilities (hard to predict samples) will result in a larger factor and a larger contribution to the total loss. The focalizing hyperparameter λ controls the amount of regularization, with $\lambda = 0$ collapsing to the CE.

$$L_{\text{Focal}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}(1 - s_{i,c})^{\lambda} \cdot y_{i,c}\log(s_{i,c}) \tag{6.3}$$

Dice Segmentation tasks are often evaluated by the intersection over union (IoU) metric, also known as the Jaccard index/Tanimoto coefficient. It is defined in (6.4) and compares the overlap of predicted and real labels with the mispredicted areas, resulting in a score between 0 (no overlap) and 1 (perfect overlap).

$$\begin{aligned}\text{IoU} &= \frac{|Y \cap S|}{|Y \cup S|} = \frac{|Y \cap S|}{|Y| + |S| - |Y \cap S|} \\ &= \frac{\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}s_{i,c}}{\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c} + \sum_{i=1}^{N}\sum_{c=1}^{C} s_{i,c} - \sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}s_{i,c}}\end{aligned} \tag{6.4}$$

However, the direct optimization of IoU is not done in favor of the optimization of the Dice coefficient [17]. Dice and IoU are related and can be formulated in terms of each other:

$$\text{IoU} = \frac{\text{Dice}}{2 - \text{Dice}}; \quad \text{Dice} = \frac{2 \cdot \text{IoU}}{1 + \text{IoU}} \tag{6.5}$$

The Dice coefficient must be maximized, but the loss terms are minimized. Therefore, it must be subtracted from 1 to form a loss function.

$$L_{\text{Dice}} = 1 - \frac{2|Y \cap S|}{|Y| + |S|} = 1 - \frac{2\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}s_{i,c}}{\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c} + \sum_{i=1}^{N}\sum_{c=1}^{C} s_{i,c}} \tag{6.6}$$

An extension of the Dice loss is the generalized (or weighted) Dice, which incorporates reciprocal weights on class occurrences to better integrate class imbalances.

clDice With the centerline-in-volume-dice coefficient (clDice), Paetzold et al. [124] and Shit et al. [154] proposed a task-specific adoption of the Dice loss to ensure connectivity preserving detection of tubular structures, such as in blood vessel patterns. The metric deals with the centerline cl of a structure/area vol. For the ground truth (Y) and the prediction (S), cl is determined. The relative amount cl_Y within the predicted area vol_S is called $cl_Y 2vol_S$ and $cl_S 2vol_Y$ for the centerlines of the prediction within the area of the ground truth labels. The metric formulation is similar to Dice as a symmetric coefficient:

$$\text{clDice} = \frac{2 \cdot cl_S 2vol_Y \cdot cl_Y 2vol_S}{cl_S 2vol_Y + cl_Y 2vol_S} \tag{6.7}$$

To realize the metric as a loss function, it must be differentiable. Centerline estimation is essential for differentiability, but cannot be achieved directly. However, a soft centerline is determined by performing morphological operations with $\min$ and $\max$ functions. Therefore, the metric is called soft-clDice. The loss is averaged with the Dice loss.

Tversky Extending the Dice loss for a better compromise between false negative and false positive predictions leads to the Tversky loss [150], where two hyperparameters α and β are introduced to balance the false predictions. The α term in (6.8) weights the false positive examples and the β term weights the false negative examples.

$$L_{\text{Tversky}}(\alpha, \beta) = \frac{\sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} s_{i,c}}{\sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} s_{i,c} + \alpha \cdot \sum_{i=1}^{N} \sum_{c=1}^{C} (1 - y_{i,c}) s_{i,c} + \beta \cdot \sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} (1 - s_{i,c})} \tag{6.8}$$

With $\alpha = \beta = 1$, the loss Jaccard index/Tanimoto coefficient is build. $\alpha = \beta = 0.5$ refers to the Dice coefficient. Two other variants will be tested in this work: $\alpha = 0.2; \beta = 0.8$ and $\alpha = 0.8; \beta = 0.2$.

Lovasz Another way to optimize the mean IoU is introduced in Berman et al. [15]: the Lovasz-Softmax loss. The Lovasz extension estimates the Jaccard coefficient based on the sorted prediction errors. These coefficients are applied as weights to the sorted errors. The final loss is an average of all the

weighted prediction errors. The loss formulation introduces a fully differentiable approximation of the Jaccard coefficient that can be directly optimized and improves numerical stability over direct use of the IoU loss.

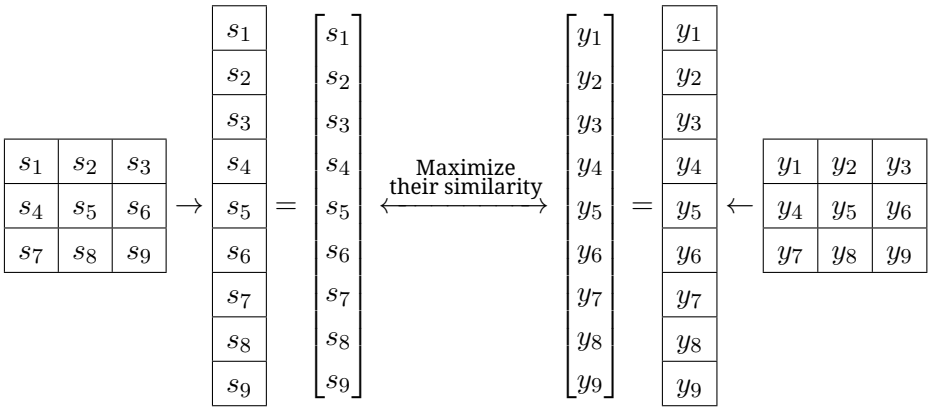
Boundary The boundary based loss [80] considers a different type of information than the previous losses. The focus is on the boundary of a shape instead of the whole area. The idea is to take the distance of each pixel to the boundary of the component to which it belongs (with distance maps). The error function is defined as the averaged distance maps of label and prediction for each class. The formulation penalizes predictions far from the label more than near errors, thus focusing on the boundaries.

Perimeter Similar to the boundary loss, the perimeter loss considers the boundaries of the class [78]. The perimeter of the regions should match between the label and the prediction. A simple approximation of the perimeter is obtained by gradients of the label and the prediction. The gradient image is written as $F(\cdot)$. In the gradient image of the labels, only contour pixels (> 0) are considered, and the sum of all gradient pixels forms an approximation of the perimeter. The squared difference of the label and prediction perimeters (6.9) can be added as a regularization term to other loss functions.

$$L_{\text{Perimeter}} = \left(\sum_{i=1}^{N} F(s_i) - \sum_{i=1}^{N} F(y_i) \right)^2 \tag{6.9}$$

Regional Mutual Information Most of the loss functions applied in segmentation tasks consider the image pixels individually. However, according to Zhao et al. [186], the local neighborhood of a pixel should also be considered. Instead of optimizing a single pixel, a high-dimensional point with the local region must be as similar as possible to the corresponding high-dimensional prediction to reduce the loss, as shown in the equation (6.10) of [186]. This approach allows the consideration of local neighborhood information around a pixel in the optimization process. The mutual information can be formulated as entropy terms of the labels $H(Y)$ and the predictions under the condition of the labels $H(Y|S)$. Reformulating the entropy as a normal distribution according to its covariance matrix allows to define a lower bound for the mutual information. Taking only variable terms further simplifies the estimation. Thus, the lower bound is $I_l(Y,S) \approx -\frac{1}{2d} \cdot trace(\log(M))$ where $M \in \mathbb{R}^{d \times d}$

encodes the variance ($Var(\cdot)$) and covariance matrices ($Cov(\cdot,\cdot)$) according to Y and S (6.11).



(6.10)

$$M = Var(Y) - Cov(Y,S) \cdot (Var(S)^{-1})^{\top} \cdot Cov(Y,S)^{\top} \tag{6.11}$$

The loss is combined with the cross entropy half and half ($\lambda = 0.5$) in the following loss formulation:

$$L_{RMI} = \lambda L_{CE} + (1-\lambda)\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} I_l^{i,c}(Y,S) \tag{6.12}$$

6.2.5 Optimization Algorithms

The parameters/weights of an artificial neural network are critical to its performance. To find appropriate parameters, the model is iteratively optimized with the data samples to minimize the loss function. The network's prediction should match the label of a supervised image pair. Typically, optimization algorithms apply backpropagation to update the weights based on the comparison between the network prediction and the label (the result of the loss function). The most common update algorithm is gradient descent (and its variant stochastic gradient descent (SGD)) [168]. To update the weights, the influence of each weight is calculated by the gradient of the loss function at its specific position. The learning rate defines the amount of correction, which is applied in the update process. The layered structure of the neural network allows the use of the chain rule for differentiation, which allows efficient computation of partial derivatives for each neuron weight. Gradient descent calculates the gradient over all sample data. However, computational reasons and other disadvantages of gradient descent discourage its use. SGD

is a variant of gradient descent, but includes only a small amount of data at a time (batch). It has been shown that SGD performs better and converges faster [168].

In practice, SGD produces a competitive result, but converges slowly. Therefore, other algorithms refine the optimizer step with additional information about the gradient and incorporate temporal statistics as well as curve analysis. The algorithm RMSProp of [63] extends the SGD method with an adaptive learning rate based on the current gradient. Therefore, the learning rate is scaled by the quadratic moving average of the current and all previous gradients. Thus, the learning rate is dynamically updated. Adam (Adaptive Moment Estimation) by Kingma and Ba [83] is a common extension of SGD that includes a momentum-based and adaptive weight decay. With two moments based on the current gradient and the moment of the previous step, the current step size is dynamically adapted. An extension of the Adam optimizer is proposed by Loshchilov and Hutter [99]. The authors show that the weight decay is not equivalent to the L_2 regularization methods used indirectly and must be decoupled. The improvement leads to better generalization performance. Zhuang et al. [189] propose another extension of Adam. Instead of the length of the gradient as an indicator of the step size, AdaBelief integrates curvature information into its update rule. This makes the algorithm suitable for the case where the gradient is large but the curvature is small.

6.2.6 Training Procedure

For the BPN (step 2) and the VN (step 3) a standard training procedure with supervised learning concepts is applied. The standard steps (figure 6.9) include data loading and preprocessing, data augmentation, model forward run, loss computation with ground truth data, and finally backpropagation to update weights. The steps are the same for both BPN and VN. Except for the different loading of labels, and for VN, the body mask is loaded accordingly to include only leg parts of the thermogram. For a fixed set of hyperparameters, a K-fold CV is applied to determine a reasonable split of the dataset into training and validation (the test set was already kept elsewhere). This partition is shared for both BPN and VN. However, BPN and VN are individually optimized in separate HPOs.

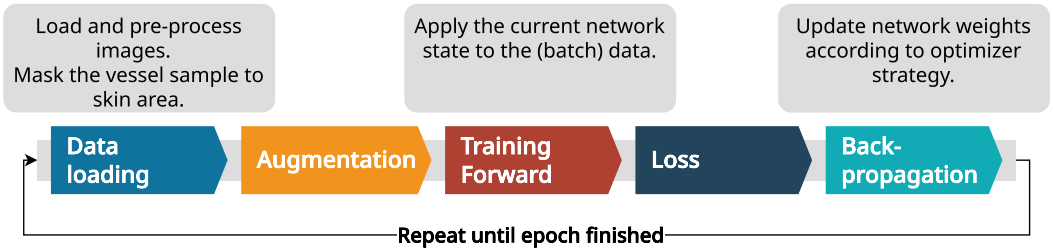


Fig. 6.9.: Overview of the training procedure with the steps applied to a single batch in an epoch.

K-Fold Cross Validation The selection of the best model is based on its validation score. However, the data selected from the dataset to train and validate the model affects the score. To maximize the score, an optimal split between training and validation datasets is required, which is achieved with a K-fold CV [55, Chapter 7.10]. In a simple case, all data samples (K) are split into N groups, where each n-th split has a different validation set compared to the other n. This is done by splitting K into a training set of size $\frac{N-1}{N}K$ and a validation set of size $\frac{1}{N}K$. However, the approach is only valid if each sample is independent of the other samples. In our case, the annotated dataset is grouped by participants. For each participant in our dataset, several thermograms are labeled. But for a single person, the appearance is similar in images from the same experiment or from other experiments. The first relates to the same clothing conditions, such as the length of pants, socks, shoes, and the impression of the background. The second considers a person's vascular structure, which does not change completely within an experiment or over several experiments. Therefore, the blood vessel patterns visible during the experiment will be similar in the samples of a single person. To mitigate the dependencies, we need to build the K-fold split on the individuals (called groups in the K-fold split) instead of on individual samples. We choose a 5-fold CV, which provides a training set of 80% and a validation set of 20%. Because we do not have the same number of samples for each person, the splits are not guaranteed to include a strict percentage of the whole data. In favor of having non-overlapping training and validation sets, the percentage is shifted dynamically. With the algorithm, it is possible to ensure similar class distributions in the training and validation sets across all splits.

Hyperparameter Search

The artificial neural network has many parameters that are optimized during training. However, the design of the model architecture is determined by the developer. The training procedure also affects the performance of the resulting model. These factors can be defined as hyperparameters. Each hyperparameter also has many possible values. The total number of combinations is huge and cannot be handled manually. However, finding the best set of hyperparameters is critical to model performance. Hyperparameter search can be thought of as a trial-and-error approach [20]. Several approaches have been developed to guide the hyperparameter selection process. A naive variant is the grid search, which executes all combinations of the (discretized) search space. More sophisticated algorithms take into account previous attempts and select a new hyperparameter set based on previous results with a high probability of achieving a better result. Bergstra et al. [13, 14] proposed the Bayesian Tree-Structured Parzen Estimator, which estimates a probability function to predict good performing hyperparameter sets. In addition to sampling new hyperparameter sets, HPO also includes early termination of trials that perform poorly compared to previous runs in the database. The successive halving algorithm (SHA), proposed by Jamieson and Talwalkar [76] and extended to asynchronous execution by Li et al. [94], compares current trials and prunes the worst trials (half of all trials). The hyperband algorithm Li et al. [95] combines multiple asynchronous SHAs with different configurations to improve results.

Search Space Since we have limited computational resources, we first select the most promising K-fold division and then apply a HPO to find all other hyperparameters. We focus on the most important hyperparameters and leave others to a predefined manual optimization. For BPN 114 and for VN 100 trials are tested. Validation epochs are evaluated with the IoU metric. The hyperparameter search will focus on the following search space:

Model Architecture U-Net, Attention-U-Net, Attention-U-Net-Supervision, Attention-U-Net-Supervision-Pyramid, Tiramisu103, DeepLabv3+, DeepLabv3+-ImageNet.

Loss Function Dice, Cross-Entropy, Lovasz, Boundary, Boundary-Dice, RMI, RMI-Dice, Soft-Dice-clDice, Weighted-Dice, Dice-Cross-Entropy, Boundary-WeightedDice, Boundary-WeightedDice-05, Weighted-Cross-Entropy, Tanimoto, Tversky-08-02, Tversky-02-08, Focal, Dice-Focal, Dice-Perimeter.

Optimizer SGD, Adam, AdamW, AdaBelief, RMSProp.

Learning Rate The learning rate has a big impact on the result of the training, because it determines the step size of the weight update during back-propagation. Although some optimizers like Adam change the learning rate internally, the initial setting is still one of the most important hyperparameters. The learning rate is sampled within $[10^{-1}, 10^{-5}]$ with a log-uniform distribution to make the smaller numbers more likely than the larger ones.

Batch Size Another hyperparameter is the batch size. A general rule of thumb is to set the batch size higher to get better results, but the batch size is first limited by the VRAM size of the hardware graphics card, and second, some combinations may perform better with smaller batch sizes. The batch size search space is also a categorical space with $b \in [1, 2, 4, 6, 8]$. From prior knowledge, for the models attention-unet-supervision and attention-unet-supervision-pyramid, the batch size must be $b <= 4$ and for tiramisu103: $b <= 2$ due to VRAM size limitations.

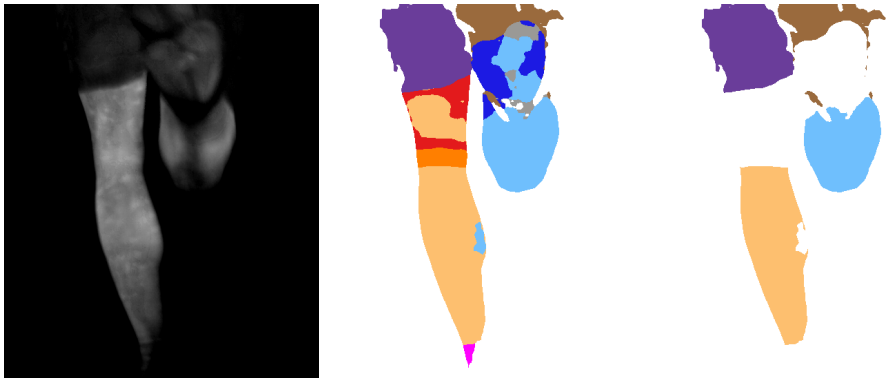
6.3 Body Part Consistency Checks for Inference

Inference of a dataset includes radiometric calibration, model prediction and thermal analysis (see ThermoNet pipeline steps figure 1.1). The first step is performed according to the calibration methods in section 4.1.5. For model prediction, the trained BPN and VN models are optimized for inference with the ONNX[1] system. The prepared 8-bit image is first processed by the BPN to obtain the body parts. The vessels are then determined within the detected calves. If there is no valid calf, it is not executed. The inference has to be done chronologically, because the radiometric calibration and the thermal statistics (see below) are based on the results of the previous positions of the segments.

Although the BPN is trained to segment and recognize single instance body parts, the results often provide multiple instances. Therefore, the training routine must effectively account for this limitation by suitable loss functions or more data samples. Alternatively, a consistency filter can be applied afterwards to improve the overall segmentation results. In this work, we decided

[1] Open Neural Network Exchange (ONNX) is an open standard for machine learning models based on an extensible computational graph model. It is supported by many frameworks for performance-optimized scenarios such as real-time model inference. [@28]

to apply consistency rules to the segmentation map from the BPN to develop the complete pipeline. Figure 6.10 shows the main problems with the segmentation masks of the BPN at the state of this work. In (b) for the left leg, the calf has two large instances that cover most of the thigh. Small instances of the right leg are also included in the left side. However, improving the masks with a hand-crafted consistency filter extracts single instances of the calf and allows applying body part segmentation in the pipeline, as shown in (c). The algorithm 6.2 outlines the main steps of filtering and extracting valid



(a) Thermogram. **(b)** Prediction mask. **(c)** Filtered prediction mask.

Fig. 6.10.: Example of poorly segmented ROIs in a thermogram segmented by BPN. The raw prediction contains multiple areas for the left and right calf. The consistency check filters unwanted areas.

calves. For each class, the predicted regions are grouped by finding connected components of the same class involving the spaghetti labeling algorithm of Bolelli et al. [22].

```
def filter_body_mask(predicted_mask):
    for shoe in ['shoe-left', 'shoe-right']:
        shoe_mask = (predicted_mask == shoe)  # Binary mask for shoe.
        # Find connected components in the shoe mask.
        components = find_connected_components(shoe_mask)
        # Find the lowest shoe component, that is bigger than 4000.
        best_shoe = find_lowest_shoe(components, threshold=4000)
    # Apply filter for each side and class.
    for side in [left, right] and cls in [calf, knee, thigh]:
        # Binary mask for class and side.
        cls_mask = (predicted_mask == (side, cls))
        components = find_connected_components(cls_mask)
        min_dist = max_value  # Initial distance
        # Filter per component.
        for component in components:
            hull = convex_hull(component)  # Find convex hull.
            # Remove small components.
```

```
18              if component.size < 10000
19              # Remove misplaced detections.
20              or component.position.y > shoe.position.y
21              # Remove if excess is present.
22              or (hull.size - component.size) > 3000
23              # Remove if convexity defects are present.
24              or component has defects and defects.size > 30:
25                  # Remove the current component from the predicted mask.
26                  remove_component(predicted_mask, component)
27
28              # Check whether the component is closer to the position of
                    ↪ the previous frame than other components.
29              current_dist = dist(component.position, pre.position)
30              if current_dist < min_dist:
31                  min_dist = current_dist
32                  # Remove previously accepted components that are now
                        ↪ invalid.
33                  remove_previous_components(predicted_mask, components)
34              else:
35                  remove_component(predicted_mask, component)
36
37      # Avoid overlapping the calves' convex hulls.
38      if intersection(left_calf_hull, right_calf_hull) > 0:
39          remove_component(predicted_mask, left_calf)
40          remove_component(predicted_mask, right_calf)
41
42      return predicted_mask  # Return filtered predicted mask.
```

Alg. 6.2: Pseudocode to filter predicted BPN-mask. The algortihm outlines the important steps. The algorithm is based on [7].

The shoe class is one of the most challenging parts because the thermograms do not visualize it well enough in the selected temperature range. Also, the rolling shutter effect occurs there at most within the image, as the shoes have the highest relative speed in motion and thus the shoes in the high position are stretched the most. Therefore, shoe classes could be segmented at multiple locations without connection. According to our definition, a class should only exist in one connected component, and we assume that shoes have the lowest possible connected component class. However, due to the rolling shutter effect, the connection could not be determined correctly. We accept shoe components for left and right if the area is larger than 4000 pixels (algorithm 6.2 line 2 ff.). Clothes are accepted without filtering because they do not affect the masks of the following steps. For the other classes (thigh, knee, calf with each side left and right), all contours are filtered individually by the following steps: a single connected component should contain at least

10000 pixels (line 18), the center should be above the lowest shoe position (line 20). Two successive filters check the contour for abnormal deformations. If the difference in size between the convex hull and the contour is greater than 3000 pixels, there are small, long artifacts attached to the component that abnormally increase the area of the convex hull. These components are also filtered because they indicate incorrect segmentation (line 22). Another check with the convex hull is based on convexity defects. This is the opposite of the previous case. Convexity defects are areas in the contour that are not convex. The size of the defect can be expressed as the distance from the convex hull to the farthest point in a defect. If the distance of the largest defect is greater than 30 pixels, it is considered a poorly segmented part and is omitted (line 24). In the last step of the individual contour filtering, a stateful check is applied (line 29 ff.). For the calves, our main target class, the center position in the previous thermogram is stored and the component (among the remaining ones) with the center closest to the previous one is chosen. If there is no previous calf, then the one with the lowest image position (largest y in image coordinates) is selected. For the other classes, the components with the highest position in the image (smallest y in image coordinates) are accepted. Finally, a check for leg intersections is applied (line 38 ff.). The analyzed calves should have similar shapes in a similar view over time. However, as the runner moves, it is possible for one leg to obscure the other, either completely or only slightly. The analysis relies on the similar view because there may be blood vessel patterns near edges. These vessels are necessary for detection and cannot be partially covered in valid images. To ensure that the legs do not occlude each other, a simple occlusion check is performed, and if they occlude, both legs are marked as invalid (with background class). We do not select the foreground leg because the boundary detection of legs with very similar surface radiation temperature (T_{sr}) is hard to distinguish and may not be correctly recovered by the BPN. To check for occlusion, both contours of the left and right calf are compared. If there is an intersection of the left and right convex hulls, then one leg is occluding the other. Figure 6.11 illustrates the principle of the occlusion check with an example. In the example, the convex hulls intersect, so both components are neglected.

(a) Left and right calves are slightly occluded and cannot be clearly distinguished in the thermogram.

(b) The BPN prediction mask shows poor segmentation in the connectivity region of both calves.

(c) Masks for left and right calf convex hulls and their intersection.

Fig. 6.11.: Example occlusion check for left (dark gray) and right (light gray) calves by finding an intersection (white) of the convex hulls of both components.

6.4 Statistical Feature Extraction

Finally, thermal statistics extracted and stored for each image. In step 4 of the ThermoNet pipeline, the segmented thermograms are analyzed for their statistical distributions within the ROIs. Only the left and right calves are examined. Other parts are omitted. The calves are the most comparable part in our field of view, as they do not normally occlude each other. The test protocol ensures that participants wear short pants that overlap only the thighs. It is not defined whether the pants should be tight or loose, which could affect the segmentation result. Therefore, statistics from upper leg classes are not comparable between different people.

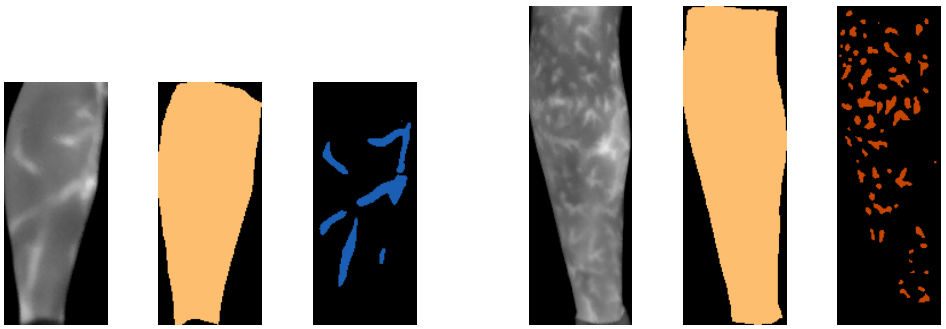
Multiple ROIs are analyzed for each side: the whole calf (BPN), the non-vessel parts (VN), the vein patterns (VN), and the perforator patterns (VN). Two BPN and VN mask examples are provided in figure 6.12 by [7]. The first (6.12a – c) shows a leg with mostly vein patterns and the second (6.12d – f) with perforators. The statistics are first calculated on the pixel intensities in the 8-bit thermogram and then converted to a temperature value with the calibration procedure of the pipeline.

For each ROI on each side, the following indicators are estimated:

numel The number of segmented pixels.

mean The mean value of all pixels.

SD Standard deviation for all pixels.



(a) Thermogram. (b) Calf mask. (c) Vein mask. (d) Thermogram. (e) Calf mask. (f) Perforator mask.

Fig. 6.12.: Two examples of segmentation masks of the calf and the vessel patterns (image cropped to left calf). (a) shows many venous patterns and (d) many perforators. Calf segmentation is represented by (b) and (e). The vein mask is represented by (c) and the (f) for the perforators. [7]

median Median intensity.

min Least intense pixel.

max Most intense pixel.

min10-mean The average of the lowest 10% pixel intensities.

max10-mean The average of the highest 10% pixel intensities.

diff10-mean The difference between max10-mean and min10-mean.

min10-median The median of the lowest 10% pixel intensities.

max10-median The median of the highest 10% pixel intensities.

diff10-median The difference between the max10-median and the min10-median.

entropy Shannon entropy with base 2.

skewness The skewness of the data distribution, whether it is more skewed to the left or to the right relative to a normal distribution. Non-skewed version based on Bessel's skew correction: $n - 1$ instead of n for normalization.

kurtosis-pearson The kurtosis (ratio of the tails to the normal distribution) of the data, based on Pearson's method. Unbiased version.

kurtosis-fisher The kurtosis of the data, based on the Fisher method (subtracts 3 from the Pearson kurtosis). Unbiased version.

probability-mean Average softmax activation for the class of each pixel (evaluated for predicted class only).

probability-min10 mean Average probability for the lowest 10% pixel intensities.

probability-max10 mean Average probability for the highest 10% pixel intensities.

blurriness ROI blurriness by variance of the Laplacian (second derivatives of the image in x and y directions) according to [127].

Additional indicators are provided for the vein and perforator classes. The whole ROI and the non-vessel ROI are usually a single large area, probably with holes. However, the other two classes are likely to have many instances within the other areas. These individual components are segmented with connected components (CCs) analysis Bolelli et al. [22]. The components are analyzed individually and the results are averaged. Since the algorithm may return components with a small area, another set of indicators analyzes only connected components larger than 8 pixels. For both, all components and filtered subset, these additional features are provided:

CC numel Number of connected components.

CC area-mean Mean area of all connected components.

CC area-median Median area of all connected components.

CC area-range Difference of the largest and smallest component sizes.

For calves, the center of the segmented ROI is extracted and the x and y components are saved. Table 6.1 gives an overview of the number of features for each ROI and analysis part (standard methods as well as connected components). The total number of features per side is 138, which adds up to 276 features for both sides of a single thermogram.

	calf	non-vessel	vein	perforator
Standard features	20	20	20	20
Body statistics	2	0	0	0
Connected components (CCs)	0	0	4	4
CCs area >8	0	0	4	4
Standard features for CCs area >8	0	0	20	20
Sum	22	20	48	48

Tab. 6.1.: Number of features for a ROI. Not all features are calculated for each ROI. The table shows the categories of indicators that are calculated. In total of 138 features are extracted for each side.

The data is saved in a CSV file for further analysis. The file name includes a unique ID and, for studies with the VarioCam HD, the timestamp. The calibration values are also saved if the image was successfully calibrated or if previous calibration values were used.

7 Stereo Transformation for Label Generation

Since automatic image segmentation of thermal images is obtained with supervised deep neural networks, a huge amount of annotated data is required. For both tasks, body part network (BPN) and vessel network (VN), we already have 870 manually labeled segmentation masks. However, this is not enough. Annotation is a time-consuming process that requires expert knowledge and is highly dependent on the annotator. Multiple annotators have different understandings of class boundaries. Another disadvantage of manual annotation, especially for the BPN, is that it focuses only on the posterior legs. There are no other regions of interest (ROIs) in the dataset. The ability to segment other ROIs of the body will greatly increase the value of the proposed system.

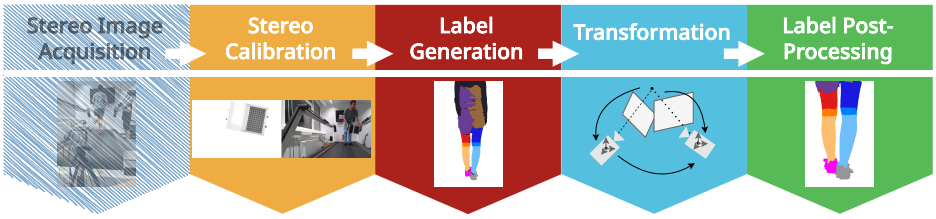


Fig. 7.1.: The steps to automatically create a custom thermal dataset with image transformation from the color+depth (RGBD) image domain to the infrared thermography (IRT) image domain based on figure 1.2: calibrate the cameras and register them in a stereo system, generate labels in the common (RGBD) domain, transform the label data to the IRT domain, and post-process the transformed label to get a dense label mask.

In this chapter, we demonstrate a system for quickly prototyping an adapted BPN to new ROIs. Therefore, we provide detailed information on the processing steps of the acquired stereo images. The first part of this chapter covers calibration considerations for joint thermal and visible camera calibration (figure 7.1 step 2). Our approach involves a knowledge transfer system from RGBD data to thermal images. In RGBD space, there are several algorithms to either segment the body parts, to separate between skin or clothing, and to

obtain the skeletal pose with joints (figure 7.1 step 3). This information can be translated into the label format we proposed earlier. To transform these easily retrievable labels, we utilize the RGBD camera in a stereo rig together with the thermal camera. In terms of projective and epipolar geometry, we can transform the pixel information from RGBD to thermal space (figure 7.1 step 4). The post-processing step (figure 7.1 step 5) fine-tunes the transformed labels and provides the final dataset. By applying this method to new ROIs of the body, we can quickly generate a new dataset with thermal segmentation masks for BPN training. Inference of the resulting network is based on thermal images only, not stereo data. The camera and stereo calibrations, rectifications and transformations are realized with implementations from OpenCV [25, @19].

The topic of transforming data between different image domains, with a particular focus on the visual and thermal domains, has been widely studied in the literature. Luo and Luo [102] introduce the topic of thermal and visible image fusion with an overview of historical, non-data-driven algorithms and current methods based on deep learning. In their paper, Rangel et al. [142] describe a system that fuses a thermal camera with a time-of-flight (ToF) camera by registration. The authors emphasize the criticality of the calibration process to ensure system stability, which has been extensively analyzed. Specifically, the design of a calibration target that is visible in all modalities, including visible light, depth, and thermal spectrum, is necessary. A heated cardboard with a circle grid pattern is recommended as the optimal choice for all three modalities. Individual camera calibration is required for stereo calibration, where images are taken in all three fields of view simultaneously. Knyaz and Moshkantsev [87] present an alternative method for creating multimodal 3D reconstructed scenes by attaching a stereo system to an unmanned aerial vehicle to capture outdoor scenes and buildings. Known building features and other recognizable patterns, such as straight lines, are employed to calibrate the stereo system in real time. In addition, a laser scanner provides accurate distance and depth information of the real world. Each camera uses structure-from-motion techniques to create a 3D model that is fused with the other cameras. Bultmann et al. [28] have developed a similar system consisting of a visible camera, a thermal camera and a LIDAR sensor mounted on an unmanned aerial vehicle. To improve the fusion process, segmentations in each domain are combined. Furthermore, new training labels from fused image labels from other domains optimize each segmentation algorithm. Manuel et al. [111] demonstrate how a depth sensor correct erroneous IRT measure-

ments caused by reflections of the object's thermal radiation at other objects. They suggest intrinsic and extrinsic calibration for better results, but did not include it in their work. Skala et al. [156] provide a thermal camera combined with a 3D scanner to model skin temperatures on a 3D human model. Therefore, the authors present different 3D scanning systems and calibrate them together with the thermal camera in a stereo system to map the image points. The authors normalize the human model and fit it to a standardized 3D model for comparisons within a database. The work of Richter et al. [144] applies the fusion for human analysis by transforming information from the visible to the thermal domain with a stereo approach. They start by detecting a skeleton pose in the visible domain and transform the skeleton keypoints to the thermal domain. Thermal radiation is measured along an automatically defined ROI. The stereo system is employed consistently throughout the experiments to capture the transformations. Our method is similar and involves transforming features from the visible to the thermal domain. In contrast, our goal is to generate a large dataset of annotated thermograms to train a deep neural network (DNN). The trained model is utilized in IRT-only applications and avoids geometric calibration for each experiment.

7.1 Stereo Calibration

As described in [53], cameras must be calibrated to remove artifacts from the projection process, such as lens distortion, or to obtain the relative position between two cameras in a stereo system. Geometric calibration can also be applied to thermal imagers. In order to apply the algorithms developed below, both cameras must be geometrically calibrated and registered in each other's coordinate system with a custom developed calibration pattern that is visible in both camera systems. The first part is to determine the internal camera calibration matrix of each camera for the projective process, as well as to model lens distortions. The second part estimates the extrinsic parameters, which consists of the rotation and translation between the two camera coordinate systems. Unfortunately, in the ThermoStereoLegs study, there was an undetected change in the camera bodies on the stereo rig mount between the time of calibration and the time of the study image acquisition. Although this change was small, the extrinsic parameters were incorrect, and we have to recover them from the study images themselves. In this section we also describe the approach to recover the original R and t.

7.1.1 Calibration Pattern

The calibration pattern requirements for intrinsic and stereo calibration are the same, and the calibrations are estimated from the same images. The calibration algorithms are based on point correspondences. A known pattern of easily recognizable feature points simplifies the matching of corresponding points. These known feature points in 2D or in 3D are matched to the found 2D features of the pattern in the image. The pattern itself is not predefined, but industry-leading software calibration tools often include patterns such as the checkerboard pattern or a grid of circles in asymmetrical or symmetrical positions (e.g. [25]).

However, for thermal images standard calibration routines may fail due to hard-to-detect calibration patterns. Luhmann et al. [101] investigate for several thermal cameras how different calibration patterns affect the calibration process. Active calibration patterns made of wood and target lamps showed poor performance because the lamp centers couldn't be measured with high accuracy. To overcome these problems, a passive test field with different thermal responses was developed. A metal plate was prepared with a special foil to create calibration patterns with different reflectance. When thermal radiation is reflected from the sky, the patterns are visible in the thermal image with higher accuracy. Another approach is a passive grid made of cardboard in front of a backplate that has either a different thermal emissivity factor or a different temperature, as described in [173]. In addition, [173] presented an adapted algorithm to find the checkerboard corners and refine the detection. In [43], an extended problem was described in which the authors calibrate the thermal camera simultaneously with a visual camera. The challenge is to find a calibration pattern that is distinguishable in both domains, which was found even with a backlit background and a non-opaque square grid.

We decided to build a simple but reliable calibration pattern made of aluminum with a size of 300×400 mm. The pattern consists of symmetrically arranged holes with a size of 10 mm and a spacing of 20 mm. The holes are made with industrial machine precision. To remove background information in the visible and thermal range, we place styrofoam behind the metal. The thermal radiation inside each hole is therefore constant, while the plate also maintains a constant one, and nothing can be seen through the holes except the white styrofoam. To improve the contrast between the metal and the styrofoam, the metal is painted black (emissivity $\varepsilon = 0.97$). The styrofoam has an emissivity of $\varepsilon = 0.60$. The difference in emissivity should be enough to detect

a difference in the thermal image data at the same temperature. However, to increase the difference and improve the grid detection performance, the aluminum is cooled in a refrigerator to 8° C or heated in the sun to 30° C, while the styrofoam maintains the ambient room temperature of about 22° C. The image acquisition should be done quickly because the metal plate will acclimate to room temperature in short time and the contrast in the thermal region will weaken, requiring a new cooling or heating round. With both concepts, black/white contrast for RGBD and temperature difference for IRT, the calibration pattern will be visible in both areas.

For better handling of the materials, we build a wooden frame around them with handles on the sides. The handles are important because touching the pattern with the hand will immediately heat it up by conduction and affect the pattern detection performance in the thermal domain. The IRT and RGBD images have different field of views and different image sizes. After building the frame, we noticed some problems with the wooden backframe and the edges of the plate. The wooden frame covering the back of the outer rows of the circle pattern results in non-uniform thermal image in these areas. As a result, the circle detection is not consistent in these areas. Therefore, the outer rows are hidden. The calibration pattern must be completely visible in the images. However, the full size of the calibration pattern of 14×19 circles is too large for easy manual handling with different positions in the images. Therefore, we covered half of the pattern with black paperwork (0.5 cm) to get a resulting pattern size of 12 rows and 8 columns. White tape around the grid increases the grid detection in the RGBD range.

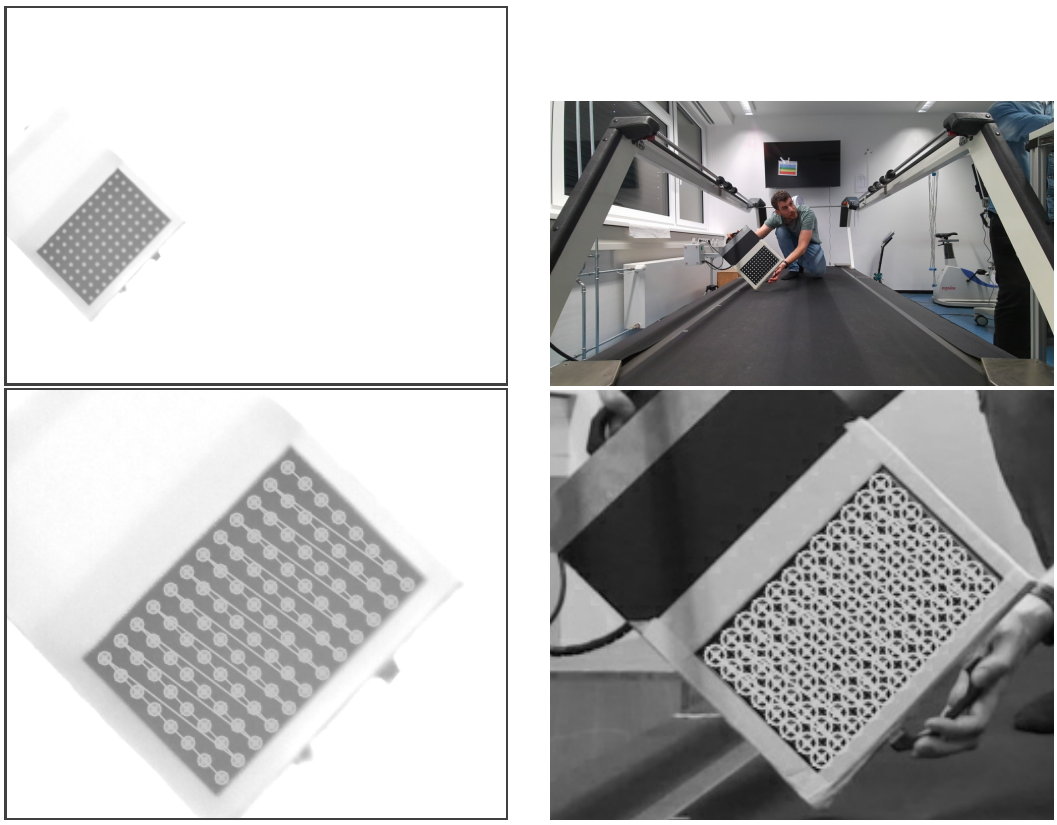
7.1.2 Calibration Procedure

This section describes how to calibrate the camera parameters for the IRT and RGBD cameras, as well as their relative position (extrinsics) in the stereo system. We assume the pinhole camera model with radial lens distortions for both cameras, and estimate the camera matrices K_{IRT}, K_{RGBD} and the distortion coefficients D_{IRT}, D_{RGBD} as intrinsic parameters. For the stereo calculations we rely on the relative pose between the IRT and RGBD cameras: $R_{\text{IRT,RGBD}}$ and $t_{\text{IRT,RGBD}}$. The fundamental matrix F and essential matrix E matrices are not used directly, but the principles of epipolar geometry are applied. The RGBD camera is a Microsoft Azure Kinect and the IRT camera a VarioCam HD. The visible light (VIS) and ToF camera should be geometrically

calibrated and registered, fortunately this is already done by the manufacturer. VIS and ToF images are time-synchronized and matched pixel by pixel. The calibration for one camera can also be directly applied to the other camera. For each camera, the board is held at different positions and angles in the image. Images where the calibration pattern is visible in both synchronous RGBD and IRT images are also considered for the stereo calibration. Since the RGBD camera has a larger field of view (FOV), there must be more images outside the thermal FOV. First, both cameras are calibrated individually, and from the intrinsic calibrations together with the same pattern points a stereo calibration is calculated.

For intrinsic calibration, the circles must be found and associated with virtual points in 3D with a distance between each point of 20 mm. Finding the grid of circles for each camera requires different approaches depending on their visual characteristics. Both images are bitwise inverted to match the conventions of dark circles and bright background of the pattern. A non-local denoising algorithm removes camera noise from the image for more reliable recognition results [27]. The IRT image is already converted to a temperature scale range (10–20° C), which increases the contrast between the background and the circles, but this contrast is further increased by applying a min-max normalization[1]. For the RGBD image, the pattern circles are barely visible in the image, so the image size is doubled. Within the prepared images, the circles are first detected by finding distinct blobs. The algorithm *SimpleBlobDetector* [@18] finds connected components that may correspond to the circles of the calibration pattern. For the thermogram, the algorithm must be very broad. A blob is searched in all thresholded binary images with a threshold value from 10 to 250 (240 different images). The blob size must be between 50 and 2000 pixels to be able to detect different circle distances. Furthermore, the blobs are not required to be perfectly circular. Therefore, a convexity[2] of between 0.87 and 1.00 is permitted. For the VIS image, we apply OpenCV's default configuration to the *SimpleBlobDetector*, which is optimized for dark blobs on a light background in visual images. The found blobs are then processed to find the centers of the circles and assign them to the given pattern employing a clustering approach [@19, *findCirclesGrid*]. Figure 7.2 shows the found circles for both, the thermogram (a) and the visual image (b). The final points for the RGBD image are halved to fit the original image size.

[1]The 8-bit min-max normalization of an image sets the lowest pixel intensity to 0 and the highest to 255, and scales the other intensities accordingly.
[2]The convexity is the ratio of the area size to the size of the convex hull area.



(a) Thermogram (10–20° C).

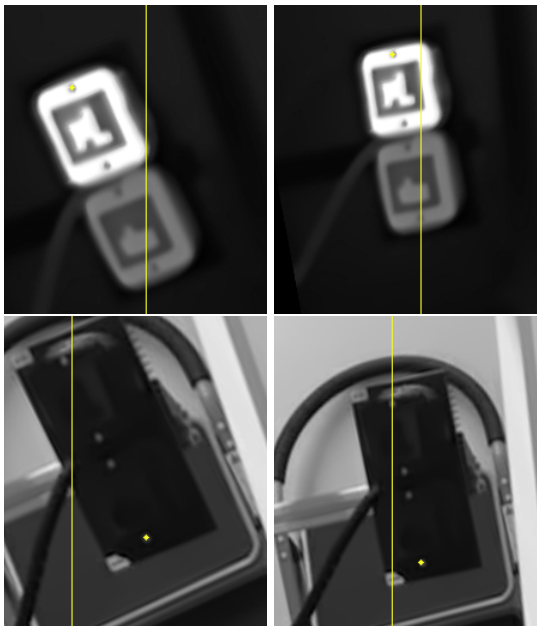
(b) VIS image from RGBD camera.

Fig. 7.2.: Circles are found in calibration pattern images in both domains. The top row shows the full images and the bottom row shows a cropped image with the pattern highlighted.

The matching grid points are stored together with the virtual points for each camera individually and, if found in both images, also for the stereo calibration. Each camera is calibrated individually with Zhang's algorithm [185] implemented in [25]. For better stereo calibration, we equalize the image sizes of RGBD and IRT by enlarging the IRT image to the size of the RGBD image. This also changes the camera matrix K_{IRT}, but not the distortion coefficients. For stereo calibration, we fix the intrinsic parameters and iteratively find the extrinsic parameters between RGBD and IRT by minimizing the reprojection error. To further optimize the result, we perform the same stereo calibration again, but now provide the previous extrinsic parameters as an initial guess. The camera matrices and stereo extrinsics can be checked for reprojection errors. They are reported and manually inspected for outliers where the circle pattern is not well recognized. These images are excluded and a new calibration run is performed until the reprojection errors are below 1.0.

7.1.3 Manual Stereo Extrinsic Correction

Due to a hardware shift after the stereo calibration that was not captured on the day of the StereoThermoLegs study, we had to correct the extrinsic calibration with manual point correspondences. The problem was detected by manual inspection of epipolar lines (see figure 7.3). We developed a manual correction algorithm based on epipolar geometry. Both images are displayed simultaneously and a user selects visually identical points manually. To ensure better quality, we perform local corner refinement and obtain a point with subpixel accuracy. The points in both images correspond to each other and are stored. Because the IRT image has few feature-rich points, the same points are found in many images. In addition, human point selection introduces



(a) Original extrinsics. **(b)** Manually recovered extrinsics.

Fig. 7.3.: Corresponding rectified IRT (top) and RGBD (bottom) images. Points and corresponding epipolar lines in other images are shown in yellow. The points are set on the screws of the calibration body. On the left, the epipolar lines are slightly off, while on the right, the epipolar lines pass through the screws correctly.

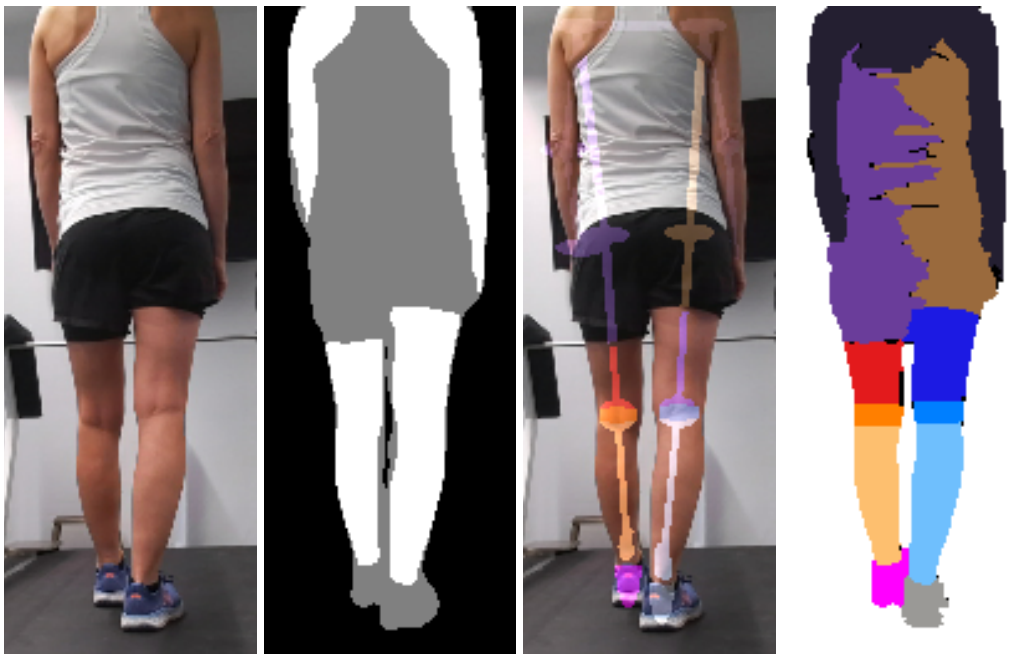
uncertainty in the positions. The five-point algorithm [119] combined with the random sample consensus (RANSAC) approach is utilized to create a new extrinsic model of both cameras. The new estimated $R'_{\text{IRT,RGBD}}$ and $t'_{\text{IRT,RGBD}}$ replace the original extrinsics. In figure 7.3 an example is shown for the original extrinsics and the manually corrected extrinsics with corresponding

point epipolar line pairs, which shows the promising correction result. The appendix figure B.3 shows the clicked points for an image pair. There are many points available for the calibration marker on the left side, but fewer in the center of the image.

7.2 Label Generation in RGBD

The goal of image translation from the visual to the thermal domain is to replace the manual annotation process with an automatic one. Therefore, labels from the RGBD domain are transformed to the IRT domain. The label description in section 6.1.1 defines several classes for the posterior legs and also distinguishes between parts with clothing and parts with skin. The generated RGBD labels should match these definitions in the visual domain and should also be represented with dense masks. The label generation process follows our published work in [6].

There are two major and several minor steps involved in generating the label. In the first part, the human pose is captured and visualized as a skeleton with joints using the YOLO pose network [109, 177]. The body mask is detected by a pre-trained DeepLabv3+ model [36], while the skin parts are obtained with another segmentation network, a FCNResnet101 [98]. The masks are fused together, as shown in the second image in figure 7.4. These three pieces of information form the basis for the second step: the application of the watershed algorithm [90]. The algorithm takes as input a mask with sparse labels. The unknown parts are filled with the labels. The creation of the initial marker mask is described as follows: along the vertices of the skeleton pose, draw either the corresponding leg label (upper or lower leg) or another non-skin mask such as clothing. The decision is based on the skin and body mask. Joints in the skeleton pose markers also define boundaries, they are enlarged by inserting an elliptical-shaped marker, which increases the known area in the watershed algorithm and improves the overall boundary detection. In the case of the knees, the watershed algorithm cannot perform well because there is no visual difference between the leg parts and the knee. However, in order to be included in the watershed algorithm, the knee joint is large enough to cover the entire area with an elliptical shape. The distinction between lower leg and shoe is initialized with vertical ellipses because there is no separating class like the knee between thigh and calf. All parts outside the body area are marked as safe background. The markers are dilated and



(a) VIS image.

(b) Mask for skin (white) and clothing (gray).

(c) Skeleton pose on VIS image.

(d) Generated label mask.

Fig. 7.4.: Components for generating labels in VIS images based on [6].

increased in area before the watershed is applied. But now some parts of the clothing may be on parts of the skin, so those pixels are eliminated. The same can happen after the watershed algorithm application, so the newly misplaced pixels are also removed. Double erosion and dilation removes the border pixels introduced by the watershed algorithm. Small clothing parts (area below 1000 pixels) are removed, as well as other artifacts are eliminated by morphological operations to ensure safe background and no clothing classes are on skin area. In our study, an operator is placed on the right side. Some parts are also labeled there, but these are not necessary and removed statically. Finally, morphological closing removes small holes from post-processing. Figure 7.4 on the right shows an example of a final label mask, along with the masks and skeleton pose, and the original image. The RGBD and IRT images have only a small overlap in their FOVs, in our case the posterior legs. The RGBD images have a wider FOV, but the label information does not need to be as accurate as in the shared FOV, which leads to non-optimal generated labels for the regions above the hip, as they are not considered in the transformation step.

7.3 RGBD to Thermal Point Transformation

The transformation process to convert the texture information (label) from the RGBD image to the IRT image is based on epipolar geometry and stereo principles. In the stereo matching system IRT represents the camera coordinate system of the IRT camera and RGBD of the VIS + depth camera. Additionally, the coordinate systems rIRT and rRGBD denote the rectified coordinate systems of IRT and RGBD. The world coordinate system W is not considered directly, in this case it is the same as RGBD, but is kept here for clarification. Other representations of the same point are named $\boldsymbol{X}_{\text{RGBD}}$, $\boldsymbol{X}_{\text{IRT}}$, $\boldsymbol{X}_{\text{rRGBD}}$, $\boldsymbol{X}_{\text{rIRT}}$ for the camera and rectified camera coordinate systems, respectively. The RGBD camera also captures depth values d with its integrated ToF module.

There are four projection matrices involved in our process: $P_{\text{RGBD,W}}$, $P_{\text{IRT,W}}$, $P_{\text{rRGBD,W}}$ and $P_{\text{rIRT,W}}$. However, intrinsic and extrinsic parameters are involved separately because we are chaining several steps together. The RGBD and IRT cameras have different image sizes: 1920×1080 and 1024×768. The intrinsic parameters are calibrated for these sizes. However, for the stereo system, the IRT images and the camera calibration matrix K_{IRT} are resized to match the size of the RGBD camera. When the transformation is complete, the thermogram resolution is restored to its original size. In the following it is assumed that the sizes match. The RGBD system is the world coordinate system, thus $[R|t]_{\text{RGBD,W}} = [I|0]$. With the stereo calibration we found the relative pose of the IRT camera to the RGBD camera, these are the extrinsics of the IRT camera: $[R|t]_{\text{IRT,RGBD}}$. The general image alignment without distortion can be expressed as follows

$$\begin{aligned} \boldsymbol{X}_{\text{W}} &= [R|t]^{-1}_{\text{RGBD,W}} \cdot (K^{-1}_{\text{RGBD}}(\text{d}_{\text{RGBD}} \cdot \boldsymbol{x}_{\text{RGBD}})) \\ = \boldsymbol{X}_{\text{RGBD}} &= (K^{-1}_{\text{RGBD}}(\text{d}_{\text{RGBD}} \cdot \boldsymbol{x}_{\text{RGBD}})) \qquad (7.1) \\ \boldsymbol{x}_{\text{IRT}} &= \frac{1}{z_{\text{IRT}}} K_{\text{IRT}} \cdot [R|t]_{\text{IRT,W}} \cdot \boldsymbol{X}_{\text{W}} \\ &= \frac{1}{z_{\text{IRT}}} K_{\text{IRT}} \cdot [R|t]_{\text{IRT,RGBD}} \cdot \boldsymbol{X}_{\text{RGBD}} \qquad (7.2) \end{aligned}$$

First we reconstruct the 3D points of the captured image in RGBD and transform them to the world (7.1). Then the coordinate system is converted to the target system IRT. Thus the original point $\boldsymbol{x}_{\text{RGBD}}$ is mapped to the IRT image plane (7.2).

Rectified Image Matching We adapt our approach to perform image matching with rectified image versions instead of the direct approach. By introducing the rectification, we are able to apply an optimization to find the corresponding image points between $\boldsymbol{x}_{\text{rIRT}}$ and $\boldsymbol{x}'_{\text{rIRT}}$. To convert our coordinate systems, the rectification rotations $R_{\text{rRGBD,RGBD}}$ and $R_{\text{rIRT,IRT}}$ are computed [54, 53]. The camera calibration matrices for the new image planes are also required. The new projection matrices P_r are decomposed by

$$P = K \cdot [R|t] = [M|p] \text{ with } M \in \mathbb{R}^{3\times3};\ p \in \mathbb{R}^{3\times1} \tag{7.3}$$

$$MM^\top = KRR^\top K^\top = KK^\top \tag{7.4}$$

$$\rightarrow \text{ Cholesky decomposition to get } K.$$

$$R = K^{-1}M \tag{7.5}$$

$$\rightarrow \text{ Rotation is reconstructed from } M \text{ and } K.$$

$$(Pc = 0) \tag{7.6}$$

$$\rightarrow \text{ Optical center } c \text{ as translation, which is the right null vector of } P.$$

to get K and $[R|t]$ for each camera. With K, D, and the corresponding K_r, OpenCV [25] provides an efficient way to compute the mapping from one image plane to the other by building a lookup table. The lookup tables will be precalculated for RGBD–rRGBD, IRT–rIRT, rIRT–IRT.

Figure 7.5 shows the five different coordinate systems and the corresponding image planes, all capturing the same 3D point $\boldsymbol{X}$. The whole process has two paths: The IRT path (red) and the RGBD path (blue). The first one is simple. It starts at $\boldsymbol{X}_{\text{W}}$ and projects it to the image plane with $P_{\text{IRT,W}}$. The next step transforms the image plane into the rectified rIRT plane. The second path contains further steps: Projection, rectification, reconstruction, transformation to IRT, projection to the rectified rIRT and inverse rectification to IRT. The second path is described in more detail below.

First, the world point is projected onto the RGBD camera image plane. The rectified version of this image is created by applying the lookup table for RGBD–rRGBD. The pixel locations and the corresponding depth values d are rectified to $\boldsymbol{x}_{\text{rRGBD}}$ and $\boldsymbol{d}_{\text{rRGBD}}$. Then the 3D points $\boldsymbol{X}_{\text{rRGBD}}$ are reconstructed by inverse projection (7.7).

$$\boldsymbol{X}_{\text{rRGBD}} = K^{-1}_{\text{rRGBD}} \cdot \left(\text{d}_{\text{rRGBD}} \cdot \begin{bmatrix} \boldsymbol{x}_{\text{rRGBD}} \\ 1 \end{bmatrix} \right) \tag{7.7}$$

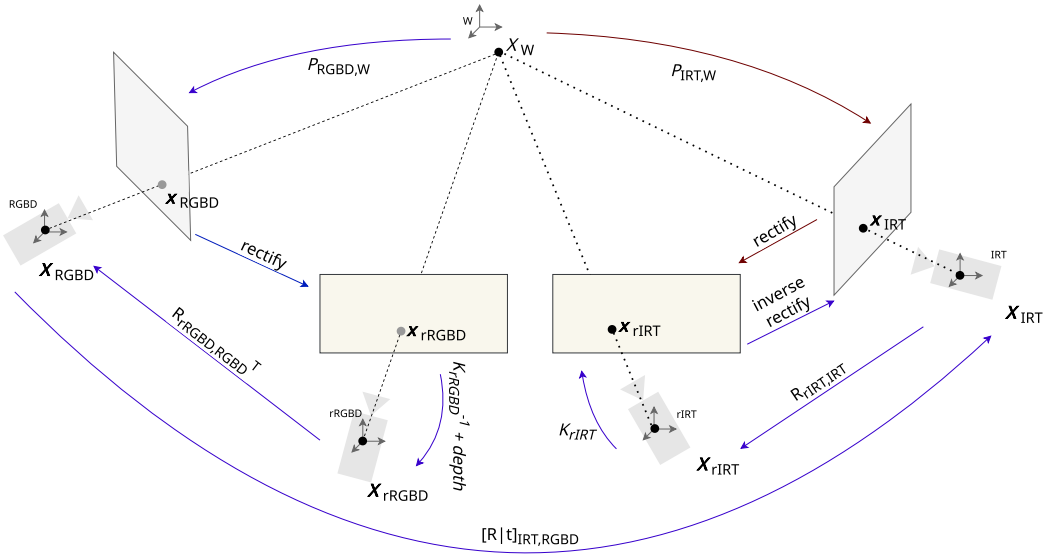


Fig. 7.5.: Simplified transformation path for matching images from RGBD and IRT cameras. Starting from $\boldsymbol{X}_{\mathrm{W}}$, the steps (blue) are to project to RGBD, rectify, compute 3D points in RGBD, transform to IRT coordinate system, go to the rectified world, project to image plane, and apply inverse rectification. Additionally, (red path) the thermal camera captures the same scene in $\boldsymbol{x}_{\mathrm{IRT}}$, which can be compared to the RGBD point in the IRT image plane.

Continuing with (7.8) we rotate $\boldsymbol{X}_{\mathrm{rRGBD}}$ back to RGBD, which allows the extrinsic coordinate transformation to the IRT system (7.9).

$$\boldsymbol{X}_{\mathrm{RGBD}} = R^{-1}_{\mathrm{rRGBD,RGBD}} \cdot \boldsymbol{X}_{\mathrm{rRGBD}} \tag{7.8}$$

$$\begin{bmatrix} \boldsymbol{X}_{\mathrm{IRT}} \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}_{\mathrm{IRT,RGBD}} \cdot \begin{bmatrix} \boldsymbol{X}_{\mathrm{RGBD}} \\ 1 \end{bmatrix} \tag{7.9}$$

From IRT the rectified system rIRT is available, so the projection places the points on the rectified image plane (7.10), which can be transformed back to IRT by inverse rectification.

$$\begin{bmatrix} \boldsymbol{x}_{\mathrm{rIRT}} \\ 1 \end{bmatrix}' = \frac{1}{z_{\mathrm{rIRT}}} \cdot K_{\mathrm{rIRT}} \cdot \boldsymbol{X}_{\mathrm{rIRT}} = \frac{1}{z_{\mathrm{rIRT}}} \cdot K_{\mathrm{rIRT}} \cdot (R_{\mathrm{rIRT,IRT}} \cdot \boldsymbol{X}_{\mathrm{IRT}}) \tag{7.10}$$

Optimizing Point Correspondences Despite the described process being expected to result in optimal point correspondences in both image planes, this is not the case. Therefore, we introduce an intermediate optimization step that takes advantage of the rectified image versions. In the rectified images, the epipolar lines l_{rRGBD} of corresponding points $\boldsymbol{x}_{\mathrm{rIRT}}$ are parallel, indicating

that the epipole is at infinity. In the present case, the cameras were arranged in a vertical configuration, with the epipolar lines representing the vertical scan lines, which share the same x coordinate with the point. The points on the epipolar line l_{rRGBD} are transformed to rIRT, forming l'_{rRGBD}. Each point of l'_{rIRT} is compared with the L2-norm to the original point x_{rIRT}. The point with the smallest distance to the original one is identified as the optimal candidate for the correspondence between x_{rRGBD} and x_{rIRT}.

Transform Texture Information We are now able to transform texture information from the RGBD modality to the IRT modality. Not only image information can be transformed, but any texture information. The collected dense labels in rRGBD are transformed to the rIRT system. For faster transformation, we build lookup maps to transform the information efficiently. Then the inverse rectification is applied to rIRT and the images are resized to the original IRT image size to get the final result.

7.4 Label Refinement in the Thermal Domain

Due to the relative rotation and translation $[R|t]_{\text{IRT,RGBD}}$ between the two camera systems, some pixels are not covered by the transformation. The different perspectives of the two cameras contain pixels in each that are not visible from the other camera. The missing information is often near edges, curves, or corners. To obtain dense transformed segmentation masks, we need to refine the masks and match the original IRT image as well as possible. In the following chapter, we explain the refinement process in detail, based on our algorithm description in [6].

The post-processing algorithm is a sequence of operations on the transformed label mask and the original thermogram. Like label generation (section 7.2), refinement is based on the watershed algorithm. The main task is to find a suitable initial marker mask, which is then filled by the watershed algorithm. The most important step is to improve the quality of the knee region. The knees are not directly distinguishable by the watershed algorithm because there is no texture change around this area that correlates with the knee itself. However, the area should be marked to separate the thigh from the calf and to cover the region where the tendons begin in the knee. We take a naive approach here by keeping only the largest part of the knee (both

left and right) and drawing a rectangular area to the left and right of the knee that is 50 pixels larger on each side and half the height of the entire knee label. This ensures that the knee covers the entire width of the leg, as shown in figure 7.6. Overlapping extensions are removed later by applying the threshold to the thermogram. For the shoe class, we need to optimize

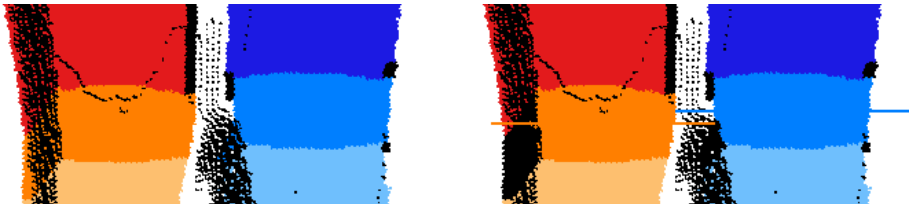
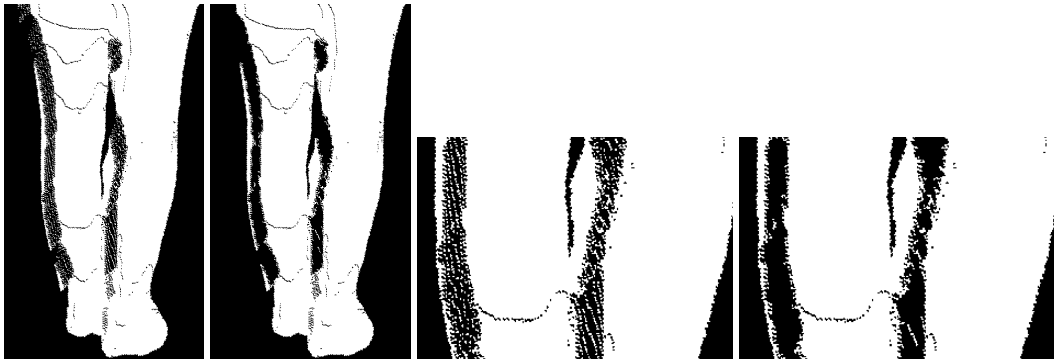


Fig. 7.6.: The largest component of each knee is taken and others are removed. The knee label is also extended to the left and right.

the mask several times. At this stage, a shoe bounding box must have an area of at least 5000 pixels, which directly filters out shoe candidates that are too small. However, we erode the area to shrink the labels because the label has problems at the edges due to less thermal information. Clothes and shoes are stored for later use. All classes are filtered to a minimum size of 40 pixels. Small artifacts from the transformation process are removed and safe background is enforced as the background class of the transformed body segmentation mask. Figure 7.7 gives an example of which artifacts are mostly filtered. The left mask contains transformation artifacts (left side of legs) and the right image removes unrelated artifacts.

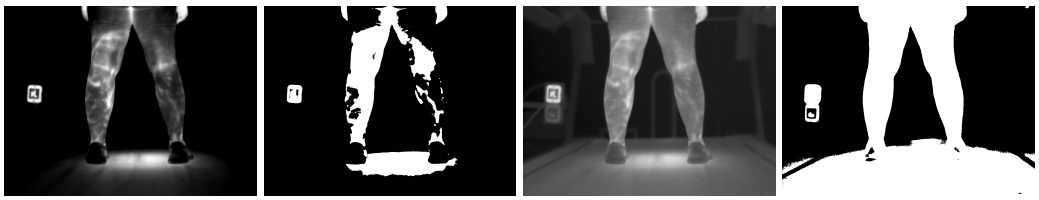


(a) Original mask. **(b)** Filtered mask. **(c)** Zoomed original mask. **(d)** Zoomed filtered mask.

Fig. 7.7.: Filtering of small transformation artifacts.

In addition to our predefined temperature scale of 25–35° C, we also use the raw image data (16-bit) to achieve better results under certain circumstances. This will be the case for a threshold of a thermogram. The 16-bit image

is normalized with a min-max normalization. Figure 7.8 shows how the fixed temperature scale thermogram is thresholded and how a 16-bit image (with min-max normalization) performs. Otsu's thresholding algorithm [123] determines the rough contours of the body. Label pixels outside the resulting



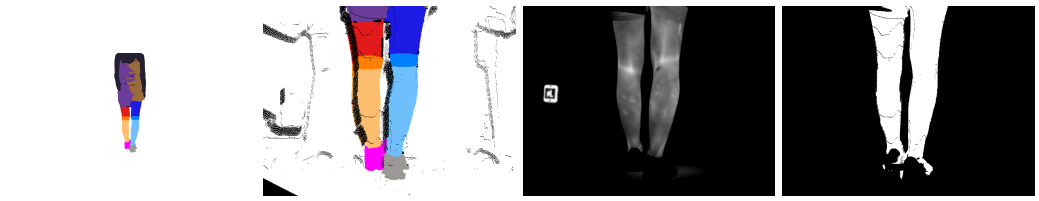
(a) 8-bit calibrated thermogram. **(b)** Threshold of (a). **(c)** 16-bit raw min-max normalized image. **(d)** Threshold of (c).

Fig. 7.8.: Example of thresholding a thermogram during a running experiment. (a) is the calibrated thermogram (8-bit) and (b) is the Otsu thresholding algorithm applied [123]. (c) and (d) show the min-max normalized 16bit thermogram and its corresponding threshold. The mask in (b) clearly shows unclosed regions within the legs, while (d) covers the legs, but also the treadmill. Image taken from [6].

mask are deleted. However, we do not fill them directly with safe background because there may be uncertainties in the thresholding result. Therefore, a safety distance to the contour is estimated by calculating a distance map to measure the distance from the boundary to each pixel outside the mask (figure 7.9d–e). Only if the distance is more than 20 pixels, we assume that the pixel belongs to the background class. The classes may contain several unrelated instances. These are leftover artifacts from the transformation process and are now removed, leaving only the largest. This filter is applied after the distance map mask is calculated to ensure that the background does not overlap the legs.

In the thermal domain, the shoes and clothes are not as detectable as in the VIS domain because their temperatures are similar to or lower than the room temperature. And because the labels are not as well aligned as a manual annotator would label them, we need to extend the distance map mask by extra space around them. Therefore, we compute a separate distance map around their convex hull and merge it with the previous one to correctly evacuate the background classes (figure 7.9f).

Another approach to reduce false positive background labels is to clear all background pixels within the bounding box of the combined classes of shoes-calf, calf-knee, knee-thigh, and thigh-clothes for each side. The bounding box is calculated for each of the pairs of classes, and no background pixels



(a) Labels in RGBD image. (b) Transformed labels in IRT image. (c) Thermogram 25–35° C. (d) Threshold of min-max normalized 16-bit thermogram.

(e) Distance map mask. (f) Distance map mask and extra space applied. (g) Erosion of shoes. (h) Applied watershed algorithm.

Fig. 7.9.: Intermediate steps for label refinement in post-processing. [6]

are allowed to remain inside this box. Subfigure 7.9f shows the rectangle evacuation spaces where no pixel is labeled around the body labels.

The previous steps estimate initial labels for the watershed algorithm by refining the area around the legs and body. But there is one part that is not yet covered. The area between the legs. Depending on the participant and the motion, this background part may or may not be connected to the outer background parts. During the transformation process, the inner background label was also not well covered because the body segmentation mask is usually oversized. For these cases, we introduce a procedure to include initial markers for the background here as well.

We assume that the area between two sides of a class is background if the line between the centers of both components crosses an edge twice. Since the edge marks the end of a class component, the area on the other side of the edge is unknown. Now, the definition of inner background is that the line connecting the centers should cross two edges. The segment between the intersection is assumed to be the background, since the labels do not touch directly. The line is drawn there. Between the shoes we also add a point with radius 2 to increase the importance of the inner background labels. The edges of the labels are found by the Canny edge detector [32] of the normalized min-max thermogram. The shoe classes still have uncertain boundaries,

so these labels are shrunk with erosion if they get too large (bounding box size $>$ 5000 pixels) (figure 7.9g). Finally, the watershed algorithm is applied. The resulting mask contains segmentation boundaries that are removed by morphological closure, and the label contours are smoothed with a median filter.

7.5 StereoThermoLegs Benchmark

In the datasets chapter 5.2 the study *StereoThermoLegs* was already introduced. The concepts of this chapter will be applied to this study. From the selected dataset 11 people make up the training/validation set and three people the test set. We train a DNN as benchmark. The purpose of the benchmark is to demonstrate the effectiveness of the newly generated dataset and to provide a baseline for further comparisons. The dataset does not contain all the thermograms captured, as this would introduce enormous redundancy. Instead, we randomly selected a small number of images from each person (~10%). In addition, we fine-tune the benchmark network and analyze the pre-training performance effect according to the amount of manual data. Therefore, several fractions of manual data are randomly selected. The following comparisons will be evaluated:

- Train the BPN with StereoThermoLegs data and report test data from the StereoThermoLegs dataset.
- Train the BPN with StereoThermoLegs data and report test data from the manual dataset.
- Train the BPN with StereoThermoLegs data, additionally fine tune with fractions 100%, 50% and 10% of manual data and report test data of the manual dataset.

The hyperparameters of the network are selected by applying the hyperparameter search described in section 6.2.6. The training procedure remains the same as for the BPN with the manually annotated dataset. Performance is measured with intersection over union (IoU) and compared between runs.

8 Sensor Fusion and Time Series Processing

The previous chapters focused on the first steps of the ThermoNet pipeline for acquiring and processing thermograms. This chapter focuses on the analysis of a sequence of images from an experiment. The individual thermogram statistics are analyzed, and additional sensors are fused together with the same time reference to allow for a combined investigation (step 5). Figure 8.1 gives an overview of the steps involved in the time series processing pipeline.

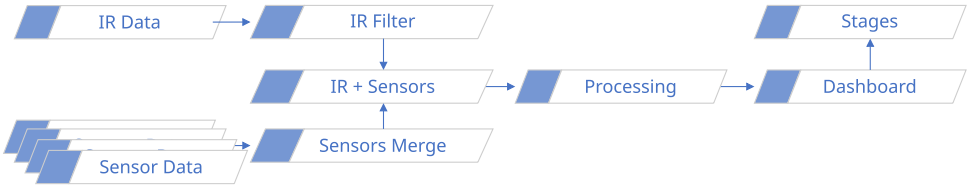


Fig. 8.1.: Overview of the steps for merging and processing time series data from thermogram statistics and additional sensors.

8.1 Acquisition System

The studies apply many different sensors, including image data. The data is collected on a single computer to ensure a common time system. The acquisition software is provided by our cooperation partner OptoPrecision. The exception is the acquisition of stereo image pairs in chapter 7, which is a custom development to ensure the special features of time-synchronous acquisition and direct control over the properties of both cameras (nonuniform calibration (NUC) rate and frame rates). This chapter only deals with non-stereo acquisition.

The loosely coupled applications exchange data over a publish-subscribe message system. Four different sensor systems/cameras are defined: the thermal camera, the ANT+ sensors, a radar-based speed sensor and the external spiroergometry system (BlueCherry by Geratherm [@9]). For each of these four groups, a capturing application receives the data and creates a protocol buffer (protobuf) [@11] message that is forwarded over a (local) network message bus (publish/subscribe architecture) for saving on a disk, separated by subsystem and capturing minute. Different message types are specialized to the specific origin, but all contain the receiver's timestamp with microsecond accuracy (if available). The timestamp is the most important part for the later fusion of all subsystems. With the approach of connecting all subsystems to one machine, we ensure that the timestamp is related to a single time reference. In subsequent steps, the captured information can be retrieved by reading all stored protobuf messages and processing them accordingly. Figure 8.2 shows the messaging approach and how the main data sources are connected. Additional services such as a graphical user interface and device control mechanisms are also integrated into the architecture, but are not shown here for clarity. Loose coupling of data acquisition and writing to the message queue system provides a flexible, easily maintainable architecture for parallel execution. New applications can be easily added as senders or consumers of message streams to introduce new functionality such as live data display or online processing and analysis. OptoPrecision includes a live presentation and control mechanism of the camera system and a live view of all sensors. The live previews are necessary for the study operator in terms of safety control and system check.

Infrared images are captured by the thermal camera with a single application and stored as protobuf messages on the storage device. The application allows certain camera controls such as issuing an autofocus cycle, manual focus setting, manual and periodic NUC. The thermogram protobuf message contains the internal camera time and image ID (since camera startup), the current focus, the raw image data with 16-bit depth, the width and height of the captured image, and the system timestamp. To reduce the file size, the serialized messages are compressed using lz4 compression [@5]. Unfortunately, it is not possible to access the current speed and inclination of the Saturn treadmill programmatically. Therefore, we have introduced a radar sensor to measure the speed of the treadmill. The inclination is not measured and taken from the experiment protocol. The sensor is connected directly to the computer and captured in an additional application that stores the

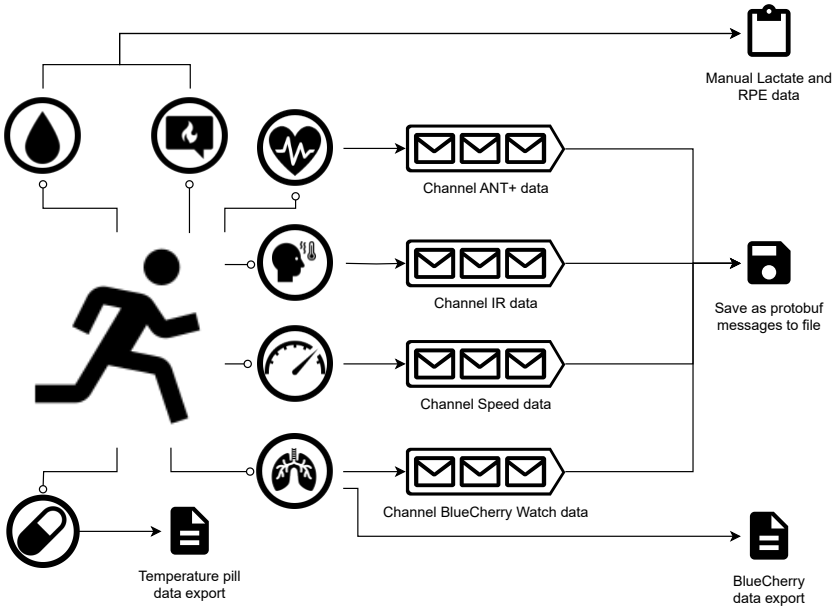


Fig. 8.2.: A person is measured by multiple sensors. Blood samples for lactate concentration (top left) and rate of perceived exertion (RPE) are stored manually. The heart rate sensor, other wearable sensors, and environmental sensors are connected to a control computer with ANT+ and forwarded to a publish/subscribe channel. The infrared thermography (IRT) images are also published on a separate channel, as well as the speed information. The fourth channel is used to monitor the status of the BlueCherry system, while the raw BlueCherry data with spiroergometry data must be exported to an extra file and cannot be published in a channel. Temperature pill data is also exported separately. All messages published in a channel are received by an extra application, which takes the messages and stores them as a protobuf message on disk.

sensor data as *timestamp*, *velocity* and *direction*. Most sensors communicate wirelessly with a controller via the ANT+ protocol. OptoPrecision has developed a capture board to receive data from many sensors simultaneously. The environmental sensors monitor the ambient temperature and humidity. The sensors CORE (including two different functions), Cosinuss One and a Polar chest band capture body core temperature, skin temperature (at one point), in-ear body core temperature and heart rate. All of these sensors are stored with the device serial number, sensor type, received signal strength indication, and sensor value. Temperature pills can wirelessly measure core body temperature from inside the digestive system. The sensing device is proprietary and not compatible with our systems. However, the exported data table contains timestamps with seconds and values approximately every 15 s. Without the global time reference system, we are unable to accurately

match the timestamp to other sensors. At the low frequency this should not be a problem, but further investigation may improve the acquisition setup.

In addition, the spiroergometry software *BlueCherry* provides various information, including breath data and test protocol stages. The proprietary software package is only available for Microsoft Windows and cannot run natively on Linux operating systems. However, our system is running on Ubuntu (GNU/Linux distribution). Therefore, we cannot execute the software directly, but we still want to make sure that the timestamps are based on the same reference time. Therefore, BlueCherry runs in a virtual machine with Microsoft Windows 10 and Oracle VirtualBox [@20]. The integration is still not comparable to the other sensors. BlueCherry has limited access to its data through a programmatic interface. Data can be periodically stored in a single file that is constantly updated, but it is not reliable for complete data retrieval. BlueCherry offers an alternative way to access data by manually exporting it after an experiment. Manual exports contain all data in a tab-delimited text file. Each row defines a new record at a new or (very rarely) the same time. The timestamps are relative to the start of the session and have no relationship to the global time system. We combine the limited data interface via a synchronization file and the manual export to derive the global timestamps for each row of data. Figure 8.3 briefly shows the implemented approach. A separate application monitors the synchronization file and captures any stage changes. Whenever a stage change is reported, the file-watching application creates a protobuf message with the new stage name and timestamp, sends it to the messaging system, and stores it like other messages from the other subsystems. After the experiment, the data must be manually exported from BlueCherry.

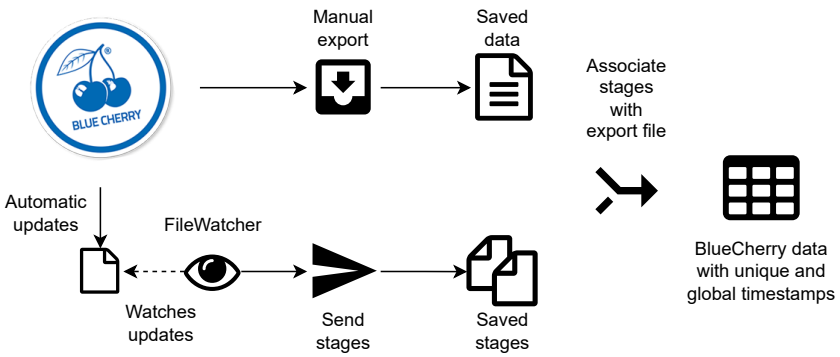


Fig. 8.3.: Data acquisition of the BlueCherry software with global time system.

Blood samples will also be taken from the ear during the pause after each stage and RPE will be reported by the participant. The samples will be analyzed in the laboratory after the study to examine lactate concentration during the study. Both lactate and RPE values are associated with the stage from the experimental protocol. The analysis can also be combined with the spiroergometry system and, if applicable, provides the physiological thresholds IAT, VT_1 and VT_2. The ventilatory thresholds (VTs) are estimated manually by a sports scientist.

8.2 Sensor Fusion

We are implementing several sensors, but they all have different frame rates and do not have the same inherent time system. Therefore, we first focus on converting and merging all time series from each sensor into a single time system. The thermal camera has a frame rate of 30 fps, which is the highest among all the sensors. Since this is our target system, and it also has the highest sampling rate, we adapt other sensors to this system and do not go beyond these timestamps. First, we combine all the external sensors, and then we merge these sensors with the IRT system.

8.2.1 BlueCherry Spiroergometry System

The data from the BlueCherry export and the state watch (figure 8.3) are combined and deduplicated before being fused with other sensors. The export file contains not only time-related sensor readings, but also static data such as height, weight, and age in a text block above the tabular data. In our analysis system, we do not include scalar, time-independent data, so we simply add these data as extra columns to our data frame. Our system relies on the timestamp as a unique value, but in rare cases the BlueCherry system reports multiple measurements for the same second. The reason is that the time precision is set to seconds only, while the underlying sensor may provide data more frequently. To overcome this problem, we take the N consecutive and duplicate timestamps and scale them from the original second t_0 to $t_i = t_0 + (0.5\text{ s}/N) \cdot i, i \in N$. The duplicate timestamps are set to be closer to the original timestamp than to the next timestamp. For the BlueCherry FileWatcher all protobuf messages contain a *timestamp* and the state. During

an experiment there are often about 4 state changes. The state *Last* determines the start of the active phase of the experiment and is set as t_0 for all timestamps from the BlueCherry export with $t_i = t_0 + t_{i,rel}$, where $t_{i,rel}$ is the relative time of the current data row. In our experience, the FileWatcher application does not always catch all state changes, but we can still recover the original start by taking a different state and calculating t_0 by subtracting the elapsed seconds. We only need one global timestamp reference, but up to 4 references are captured.

8.2.2 External Sensors

The additional external sensors can be combined with the BlueCherry data. We have encoded all values in protobuf messages with the system timestamp. The messages are read into a dataframe and the datatables of the previous BlueCherry data are merged with the sensors by a full outer join[1] based on the common *timestamp* column. In this way, we ensure that all the data is inserted into our relational data model, even if it is very sparse. Another external sensor is the temperature pill. These data tables are not in the usual protobuf messages, but they provide tabular data with an associated timestamp. When data is available, it is also merged into a common datatable in the same way. For further operations we now focus on the experiment as reported by BlueCherry. Therefore, the dataframe will be cropped within the beginning and end of the experiment, additional time points will be ignored.

8.2.3 Speed and Pause Detection

Although BlueCherry allows to set up the test protocol by setting time ranges and target speeds of the treadmill, our system has no connection to the treadmill and cannot control its state. This task is performed manually by an operator at the machine. Therefore, we do not rely on the BlueCherry speed data and concentrate on the speed detected by the radar and the later integrated step values. The speed radar sensor determines different speed phases such as stationary and moving. However, it is not reliable for direct speed assessment due to noisy and uncertain measurements. To reduce noisy signals,

[1] A full outer join refers to the combination of both tables based on a common key. The merge preserves each row, regardless of whether the key of the row is present in the corresponding dataframe or not.

the data is smoothed by a moving maximum with a window length of 5. In a moving window, the current value is set to the maximum value of the window. The maximum is chosen over the average because the signal noise is usually below the target value. Averaging would take into account the strong negative values that we want to eliminate. A Butterworth filter [30] (second order) with a normalized cutoff frequency f_c of 0.003333 Hz additionally reduces the high frequency signals while preserving the low frequency parts (noise reduction with a low pass filter). A forward and a backward pass are combined to avoid phase shifts when applying the filter. To obtain the velocity change point in the time series, the acceleration curve is first estimated and then smoothed with a moving window of length 5. We are only interested in the points of maximum acceleration (inflection points in the velocity curve). These are set as the velocity stage boundaries. Thus, the acceleration data is normalized to $[0, 1]$ and data below 0.2 is set to zero. The remaining data is grouped as consecutive values without 0 and analyzed for a single maximum within a group. Each new velocity step is assigned an increasing step number.

Next we introduce a field *velocity pause*. It's values are defined as 0 for pause and 1 for moving. To determine whether a velocity phase is a pause or not, the average of the values in the middle of the phase is considered. This avoids the phase boundaries with potentially noisy measurements and likely acceleration-related velocity differences. If the mean is below a threshold of 1 km/h, then the minimum must also be below 0.5 km/h to ensure that it is a real pause and not a slow moving treadmill. To avoid selecting the beginning or the end of a phase, the pre-phase (2) and the post-phase (3) are set to the 20 s from the beginning or the last 20 s of a moving phase. During this time, people may not be moving as fast as the target speed of the phase because the treadmill is still in the acceleration phase. The speed filtering and processing approach is designed to overcome the noisy behavior of the sensor module and may not be necessary for other speed estimates or values directly reported by the treadmill.

8.2.4 Lactate, RPE and Thresholds

The current state of sensor and external data fusion is the combination of different sensors, speed information, stage information and more. However, the lactate blood samples and the RPE are not yet integrated. The manual and external estimated values cannot be fused directly by their timestamp

because these values do not have a concrete timestamp. Data is loaded from tabular export files. Some also contain additional values such as the IAT, VT_1 and VT_2, and other manually collected values (age, sex, and others). Data are available for each phase of the protocol. The phases are given with their relative start and end times, except for the first phase *Pre*. If we apply the global start point from BlueCherry observer to the start times of the phases, they do not match well with the real moving phases in our experiment. This may be due to small delays in the treadmill configuration or other timing issues. Nevertheless, we can match the protocol phases to the correct velocity phases in our previous data frame. Figure 8.4 gives two examples of protocols

BlueCherry Phase	Ruhe	Last													Erholung	
Velocity State	P	M	P	M	P	M	P	M	P	M	P	M	P	M	P	M
Protocol Speed	0	6	8		10		12		14		16		18		4	4
Protocol Stage	Pre	1	2		3		4		5		6		max		Rec+1	Rec+3

(a) Incremental step protocol. The assignment of the protocol steps and the speed to the movement phases is straightforward.

BlueCherry Phase	Ruhe	Warmup	Last									Erholung	
Velocity State	P	M	M					M	M	M	M	P	M
Protocol Speed	0	6	10	10	10	10	10	14	8	14	8	4	4
Protocol Stage	Pre	1	2	3	4	5	6	7	8	9	10	Rec+1	Rec+3

(b) Steady run with multiple stages and alternating phases. It is not possible to directly associate speed with stage.

Fig. 8.4.: Protocol examples with BlueCherry phases, measured velocity state, protocol stages and speed definition (P=pause $\hat{=}$ speed=0 km/h, M=moving $\hat{=}$ speed>0 km/h).

with the BlueCherry stages, the measured states with the speed sensor (pause is P and moving treadmill is M). The protocol stages and the corresponding speed are also given. The BlueCherry data is divided into three groups: *Ruhe*, *Erholung* and *Warmup+Last*. The first group always refers to the *Pre* phase, the second group to the post-exercise phase, the recovery. Depending on the protocol, different numbers of speed stages are available. Two recovery stages should be defined, the recovery after 1 minute (*Rec+1*) and after 3 minutes

(*Rec+3*). If there are two speed stages, we assign them accordingly. Otherwise, the same speed is assigned to both recovery stages.

The main part of the stage assignment has to be done in the *Warmup* and *Last* phases according to figure 8.4. The distinction between *Warmup* and *Last* is also only present in the BlueCherry data, so we will only focus on the found velocity states. In the simple case (a), each stage is separated by a pause, and consecutive stages have different velocities. Matching the measured speed to the protocol definition is straightforward. For long steady runs like in (b), a moving phase may have several stages with the same speed. Furthermore, they are not separated by pauses and a direct matching is not possible. For each time frame with the same speed, the protocol definition is checked to see if there are multiple stages with the same speed or not. The first case requires further investigation, the second case continues with direct comparison. If there are multiple stages with the same speed, the duration of each stage from the protocol definition is added to the consecutive stages starting with the first timestamp. The end of the last stage is set to the next stage with a different speed.

In the incremental exercise test, the speed is not only given by its real value, which is constant during the whole stage. However, the body adapts during the time and reaches its body state according to the given speed after about 3 minutes. Therefore, each phase lasts 3 minutes. To simulate a steady increase in speed, sports scientists linearly interpolate the speed from a previous stage to the current target speed. The virtual speed related to the individual anaerobic threshold (IAT). In the experiment, we take each stage and calculate the linear interpolation from the previous target speed to the current target speed. However, the first stage should not start from 0 km/h, hence the same increase as in the following steps is assumed.

The recovery and pre stages are excluded from this assumption and have a constant speed value. If the stage *Rec+1* is missing, it is reconstructed from *Rec+3* and its previous stage. Rec+1 means that the stage limit should be set 1-minute after the exercise. However, there will be no manual measurements for the reconstructed *Rec+1*.

Two VTs can be estimated manually for a cardiopulmonary exercise testing (CPET) in the BlueCherry software. These values are not exported and only present in the manual log file. They are combined as single time points in the experiment. Additionally, the IAT is given in the form of a lactate concentration. We perform a linear interpolation between the lactate values

and find the time at which the IAT was reached. Based on the IAT and the virtual speed, training zones are defined. All four zones are marked in the datatable. After merging the sensor data with the corresponding manual data, missing values are filled in with previous values to ensure that each time point has a valid entry for each sensor.

8.2.5 Thermogram Statistics

The feature extraction of thermograms according to the defined regions of interest (ROIs) data (see chapter 6) is now further processed. First, the thermogram statistics are clipped within the times of the experiment. In the previous steps, we have already identified malicious data by prediction artifacts, but the cyclic behavior is not examined. Running with both legs is periodic and the leg movement is shifted. According to the running cycle in figure 8.5 by [180] the legs are either in a standing phase or in a flying phase. In the standing phase, the leg moves towards the camera at the speed of the treadmill, while in the flying phase, the leg is lifted away from the camera, but is also bent. The viewing angles become much larger than 40° and the relative speed also increases. In particular, the speed of the flying leg relative to the image plane has a large effect on the image, because a high translation of the leg points that is faster than 8 ms introduces motion blur. In addition, there are two effects caused by the rolling shutter: first, an object (here the lower leg and especially the shoes) is enlarged, and second, information is superimposed in the enlarged areas. On the original point another point has appeared and both radiations are accumulated, leading to false information.

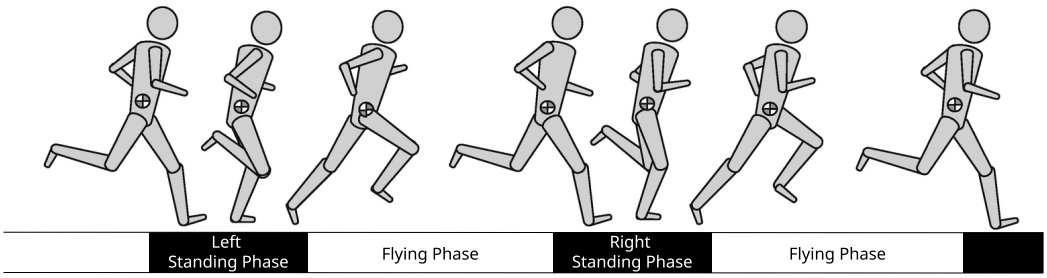


Fig. 8.5.: Schematic double step cycle of a runner. One leg touches the ground during the standing phase, while both feet are in the air during the flying phase. The left and right legs alternate the standing phases. Modified and translated image from [180].

The different positions result in varying sizes, leading to different measurements of the thermal signal. The angle plays a role in the thermal radiation detection when it exceeds more than 30–40°. We do not measure the angle of the legs because the stereo rig with color+depth (RGBD) camera was not available during the study recordings and has a lower frame rate. Nevertheless, the analysis of the cyclic movement can recognize the regions with the largest detected ROI, which we assume to be the phases with an uncovered view from an almost straightened leg in the standing phase. The data are filtered separately for each side, since the left and right legs have opposite movements. We find local maxima of the number of elements in a ROI in the data series. There should be at least 5 different points between two local maxima, which ensures a robust finding of local maxima in the cyclical movement and in the standing phases, where the area has similar values over time.

The thermogram statistics are merged with the previous sensor data frames based on the common global timestamp. We do not apply a full outer join here. The camera data has a nominal frequency of 30 Hz, while the sensors have a lower frequency. However, the actual capture times are different, and joining based on microseconds will result in very few matches between the camera and sensor timestamps. The few matches would result in a data frame with many empty and mismatching values. The goal is to match the lower frequency external sensors to the higher frequency camera data. The sensor data also contains invalid values that have been augmented with the last valid value to preserve the sensor state until new data is available. For each camera timestamp, the best matching sensor timestamp is found. Both rows are merged. Data loss caused by a sensor timestamp that is between two others that are best matches of two adjacent camera timestamps is very unlikely. Except during camera NUC phases and during phases where both legs have an unusable ROI data loss may occur. The goal is to always have reliable and non-repeating camera data, so some data loss from other sensors must be considered. Since there is a potential loss of sensory data, it is possible that the threshold positions IAT, VT_1 and VT_2 will also be lost. These positions are recovered by separately matching the VT timestamp to the IRT data. The thresholds are reconstructed from low-frequency sensor data, the small possible time shift in our approach is negligible for the threshold accuracy.

8.3 Time Series Post-Processing

The fused data is further processed to gain more insight and present it well to analysts. The data points are noisy, whether they come from the ROIs of the calves or from the vessels. The external sensor data such as spiroergometry and other sensors are also affected by noise. For better visualization, we are only interested in the lower frequencies of the data. Therefore, the data is smoothed with a Savitzky-Golay filter [152] with a window length of 151 data points. The Savitzky-Golay filter fits a polynomial function to the data points within the window around the target. The target point is fitted to the curve. The filter preserves the characteristics of the curve while reducing noise.

For medical applications, the maximum values of heart rate, oxygen uptake ($VO_{2,peak}$) and the maximum number of perforators are determined. In addition, a comparison between predefined values must be estimated: Pre vs. IAT, Pre vs. Post, IAT vs. Post, Post vs. Rec, Pre vs. Rec. For these phases, the difference of the mean surface radiation temperature (T_{sr}) of each side's calf is calculated. Pre is the time just before the start of the exercise, Post is immediately after the exercise and Rec is at the end of the recovery period. For the Post vs. Rec comparison, reperfusion is also examined by comparing the number of perforator components within a calf. To analyze whether the overall effects can be measured when calculating at both sides at the same time, those are also averaged. The ratios of T_{sr} from vein (V_{sr}) to non-vessel (NV_{sr}) and perforator (P_{sr}) to non-vessel supports the analysis whether the specific vessel patterns have lower or higher T_{sr} than the non-vessel parts. The time is converted to timestamps relative to the beginning.

The prepared data represent the whole experiment. However, at some points only a compressed version of each stage is needed. Therefore, for each stage from the corresponding manual "stage" the last 30 data points are averaged. Additionally, the pause stage is taken into account. In the pause, however, the values are taken directly from the standing phase, i.e. at the beginning. This makes it possible to compare the values during and immediately after each phase. A representative thermogram is exported for each phase. Pictures taken at the same relative time during the running phase can show completely different leg positions because the runners move differently, so for each phase the first picture is taken during the following standing phase (pause). The thermogram is found by first getting the next pause phase and taking a thermogram about 5 s after the start of this phase. It may not be the first

image because the treadmill is still in the braking phase and the runner has not completely stopped.

Data visualization is provided within a dashboard as with multiple views. Nine different views provide an overview of the individual insights of an experiment. Not all information is available for all experiment types, e.g. the incremental walking test has defined thresholds, the others do not. In some fields, only one side of the leg is evaluated. Therefore, we estimate the more qualitative results by checking the noise in the NV_{sr} for each side and taking the less noisy one. The noise is compared to the mean squared error (MSE) of the difference between the original values and the smoothed values. The side with the lowest MSE is employed for visualization. These fields are defined:

1. The first plot compares the heart rate with the smoothed NV_{sr} of the selected side over time (x-axis). Additionally, the three thresholds IAT, VT_1 and VT_2 are marked as well as the training regions (REG, GA1, GA2, EB).
2. The second panel shows the smoothed NV_{sr} and the smoothed respiratory frequency (Bf) over time. The visualization includes only phases in which the person is moving (excluding acceleration and deceleration phases). The initial T_{sr} is also marked as a horizontal line.
3. Rest and recovery images are shown, as well as images with the highest number of vein and perforator patterns on the left leg. The detected ROIs are superimposed on the calf.
4. The core temperature from the in-ear device, the core temperature from the pill sensor are compared to the ratio P_{sr} to NV_{sr} over time.
5. The virtual treadmill speed is plotted along with both sides of the ratio of V_{sr} to NV_{sr} over time.
6. Figure 6 shows the environmental settings with room temperature and humidity over time.
7. Field 7 shows the left and right NV_{sr} over time.
8. Field 8 shows the total number of pixels of the left and right perforator components over time.
9. This field shows the average area of the veins on each side over time.

Implementation Time-series processing is powered by `Pandas` [114, 164], `SciPy`[175], `OpenCV` and `numpy`. For convenience and data analysis, all trials of a study are combined into a single data frame and exported as a comma-separated value (CSV) file. The appendix table A.4 displays and explains many of the available sensor data fields along with their units or type. The interactive plots and visualizations are implemented with plotly [@22]. The plots and graphs are saved as HTML files for further analysis, allowing application users to interactively navigate within a plot.

PART III

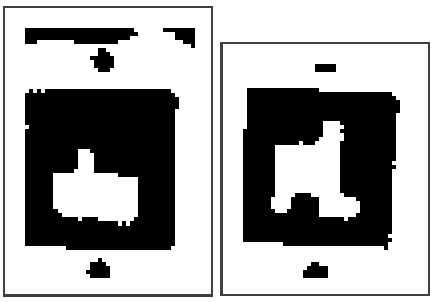
OUTCOMES

9 Results

This chapter presents the results of the methods proposed in the previous chapters, analyzing each step: starting with the thermogram acquisition and the radiometric calibration device, an overview of the validity and stability of the calibration performed is given. This is followed by the manually annotated data set, including the results of the K-fold cross-validation (CV). Then the body part network (BPN) and vessel network (VN) are evaluated. For the automatic dataset generation of the stereo approach, the calibration, the resulting dataset, the benchmark network training, and a comparison of the performance when applied to thermal analysis are explained. In the last section, the outcomes of the sensor fusion and time series analysis are presented.

9.1 Two-Point Radiometric Calibration Target

When taking thermograms, we place a temperature reference object in the field of view (FOV) of the camera at the same distance as the target object. A pixel-temperature mapping is obtained from the pixel values of two known areas. The areas are found and identified with two ArUco markers without prior information about size and position within the image. The detection first performs an exhaustive search, but keeps the previous locations and performs the search in the next thermograms first within the previous region of interest (ROI). With this implementation, continuous detection can be performed in successive thermograms at up to 80 fps. The thermograms are pre-processed for better detection results. For all combinations of different thresholds to find a marker in the full 16-bit image, the processing takes up to 80 s per image. Figure 9.1 shows two detected and preprocessed thermograms. The detected marker contour does not perfectly match the ArUco definition, but the recognition algorithms can correctly associate them.

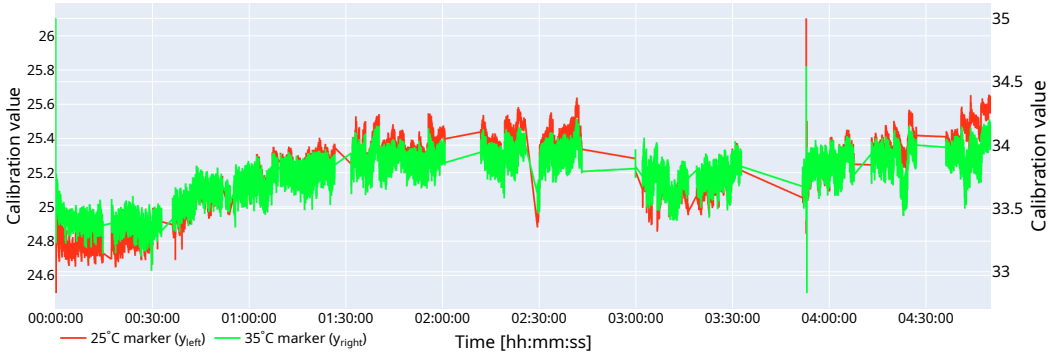


(a) 25° C-marker. **(b)** 35° C-marker.

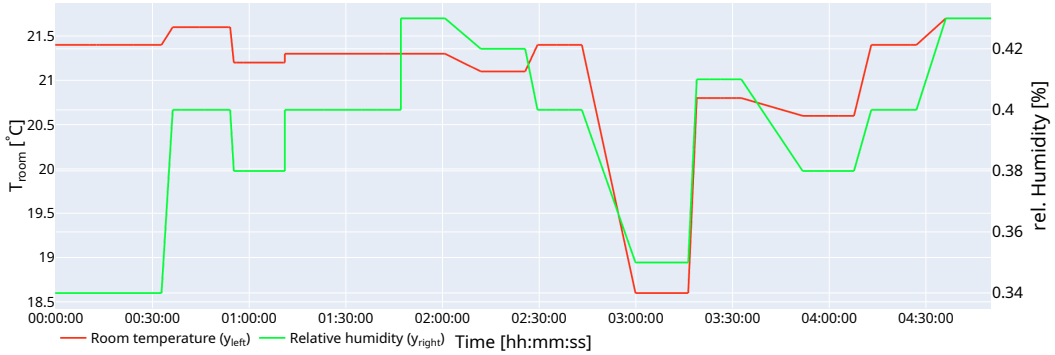
Fig. 9.1.: Examples of markers found in preprocessed thermograms.

The calibration plates are controlled by a PID controller, so the temperature varies slightly. Figure 9.2a shows the calibration values for both markers over time during the entire StereoThermoLegs study. It can be seen that both calibration targets show similar behavior over time and that the calibration values increase over time, indicating camera drift. In (b) the corresponding room temperature and relative humidity are plotted. Pixel intensities change as room conditions change. However, the intensities do not directly reflect the change in room conditions. The internal camera temperature may change differently and have a greater effect on the observed intensities. The variation of calibration values over long time measurements and the noise in the short term show the need for radiometric calibration in each image.

OptoPrecision provided an additional calibration device with a single temperature controlled plate based on the TC-XX-PR-59 temperature controller from Laird Thermal Systems Inc., Rosenheim, Germany [@14]. The single target has a stability of $\pm$ 0.05 K and, like the two-point target, an unknown offset error. The device is similar to the two-point target and is utilized in experiments to measure the influence of different angles and distances in the calibrated image. The experiments were conducted in the Department of Sports Medicine, Prevention and Rehabilitation. The ROIs for both the target and reference plates are determined manually by labeling the plates and extracting the center circle with a pixel size of 29. Figure 9.3 shows the results at different angles. The angle is measured between the optical axis and the normal to the target plane. In the first graph the target is rotated horizontally, in the second graph it is rotated vertically. The PID signal is given by the internally measured temperature of the target plate. The mean values for the PID signal and for the target surface radiation temperature (T_{sr}) are 29.99° C and 31.12° C for both graphs. The observed T_{sr} of the target shows a similar



(a) Calibration values for upper and lower marker.



(b) Room temperature and relative humidity.

Fig. 9.2.: Calibration values for both markers during the StereoThermoLegs study and ambient temperature and humidity.

curvatures as for the PID signal curve, but has an offset error of about 1.13° C. However, if the angle is 40° in vertical direction, the observed T_{sr} is changed more than the expected noise. The experiment shows the influence of the emittance of the object due to the angular dependence of the emissivity of the material, as described in section 3.2.1.

In addition to angle, distance also affects the measured pixel intensity. To check how much this affects our experimental setup, we set up a test with the single point target and change the distance of the target while leaving the calibration targets in the same place. The PID controller was not accessible for this experiment. The calibration targets are placed at a typical distance for runners on a treadmill of 2.1 m. Figure 9.4 shows several measurements at different distances. The variance of the target T_{sr} is 0.0034° C, which is less than the radiometric resolution of the camera. This shows the stability of the measured region with the 29 pixel over different distances, although the real size covered by the ROI changes.

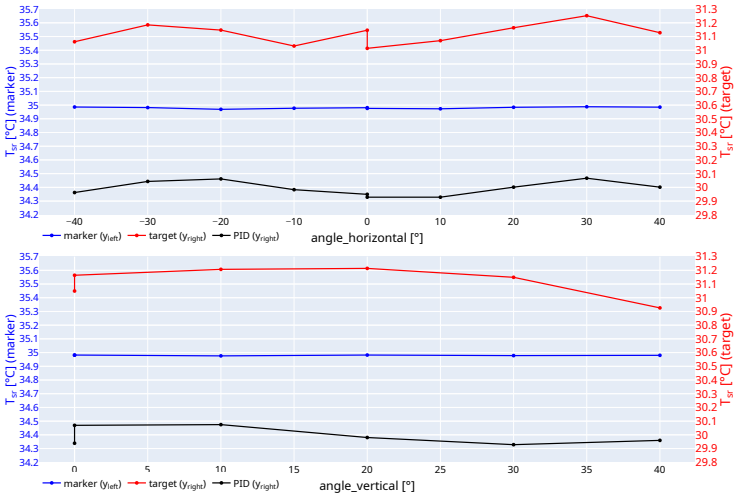


Fig. 9.3.: Influence of the angle of a thermal radiator with our calibrated setup. The left y-axis denotes the T_{sr} for the calibration target, while the right y-axis indicates the target T_{sr} and the PID control value of the heater.

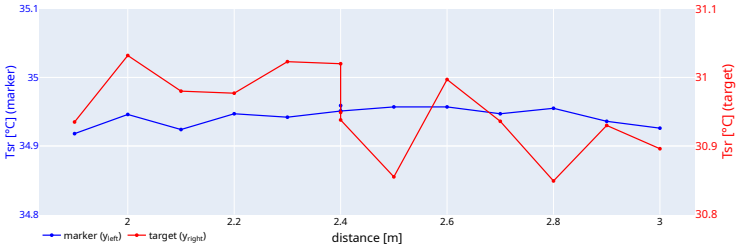


Fig. 9.4.: Influence of the distance of a thermal radiator with our calibrated setup. The left y-axis denotes the T_{sr} for the calibration marker, while the right y-axis indicates the target T_{sr}.

9.2 Annotated Datasets

In section 6.1 we described the method of collecting annotated data and the proposed class definitions. As a result, we curated two datasets, one for the body parts of the posterior legs and one for the vessel structures. The data were manually annotated by six people, each with a different amount of work, utilizing the extended PixelAnnotationTool (PAT) (section 6.1.3). These datasets are used to train, validate, and test the proposed models. We are aware of the fact that the test set will only be used in documents to publish results and not to study the data in advance.

The first manually labeled dataset consists of dense masks with body parts from several medical studies with the two different thermal cameras. 17 persons are from the COMMED study, 5 from LAUFRAD, 15 from SPEER and

9 other persons are not included in any study. A total of 46 humans with 870 annotated thermograms are included. For most people, a range of 10 to 20 images was labeled, except for three persons who have 30, 60, and 80 annotated images. The images are selected from the entire experiment. However, many images are labeled in the standing phase because these images were already selected during the analysis according to the manual strategy, as in [61]. The test set includes 15 of the 46 individuals with 200 of the 870 images. This results in 77% of the images as the training set and 23% as the test set. People from the LAUFRAD study are only included in the test set. Images of a single person are not separated into different dataset parts. Some of the experiments from the unregistered studies are captured in a vertical format that covers the entire back of a person, not just from the hip down. Since there are only a few examples with this format, we crop the images at the hip to achieve the same visual image perception as the others.

The second dataset is about the segmentation of blood vessel patterns of the posterior legs. Each vessel label has a corresponding body label and vice versa. The vessel background matches the body classes of background, clothing, and shoes, leaving only body parts with skin for the vessel labels. For each labeled image in the body parts dataset, a vessel mask is also annotated.

The training set also needs to be split into a training part and a validation part. To find an appropriate split, we compare the training of the applied models with K-fold CV with fixed hyperparameters. The split is applied for both the vessel and the body part networks. Figure 9.5 shows the five splits for the data set (each color in the top row represents a different person) that are included in the deep neural network (DNN) training to select the best performing split. The best validation intersection over union (IoU) achieved was in the K-fold 3 with a result of 0.65 IoU. The others performed worse (table 9.1). In the final chosen split, the training set has 540 samples from 27 participants and the validation set has 130 samples from 4 persons.

K-Fold	Best validation IoU	Best epoch
0	0.59	94
1	0.63	64
2	0.57	44
3	0.65	83
4	0.59	103

Tab. 9.1.: CV results for 5-fold dataset split. Trained model: Attention-U-Net with Dice and AdaBelief, learning rate 0.001.

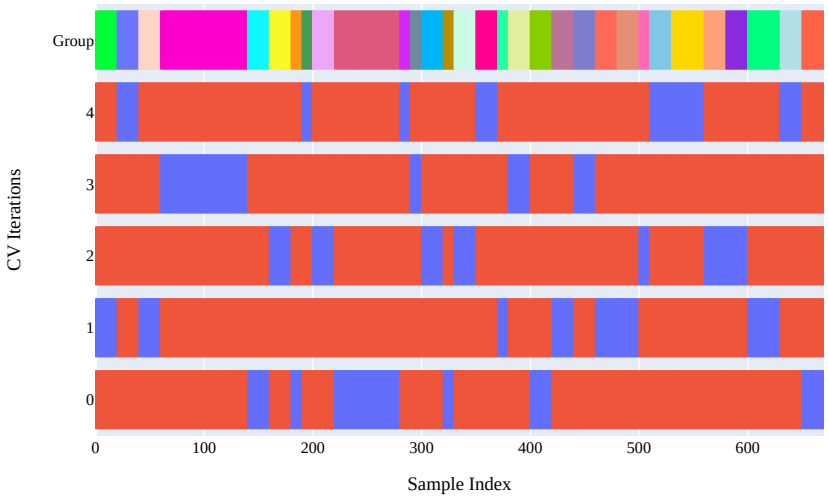
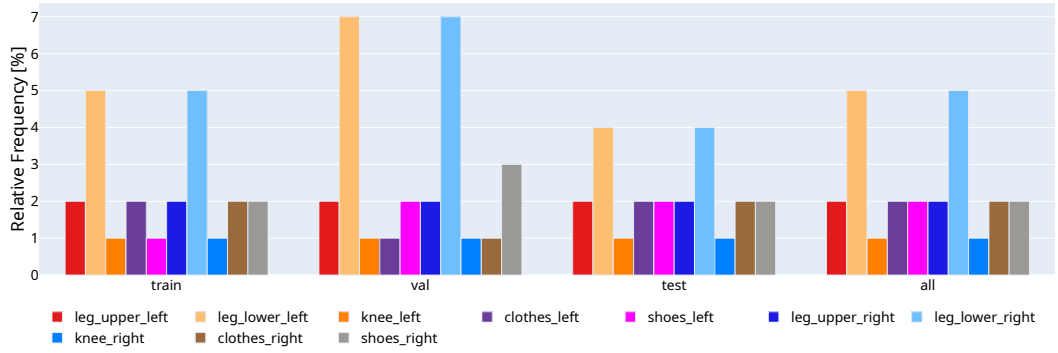


Fig. 9.5.: Example of how stratified groups work in the case of our project: 5-fold CV. The first row *Group* shows the samples grouped by person in different colors, and each following row shows a different split of the 5-fold split.

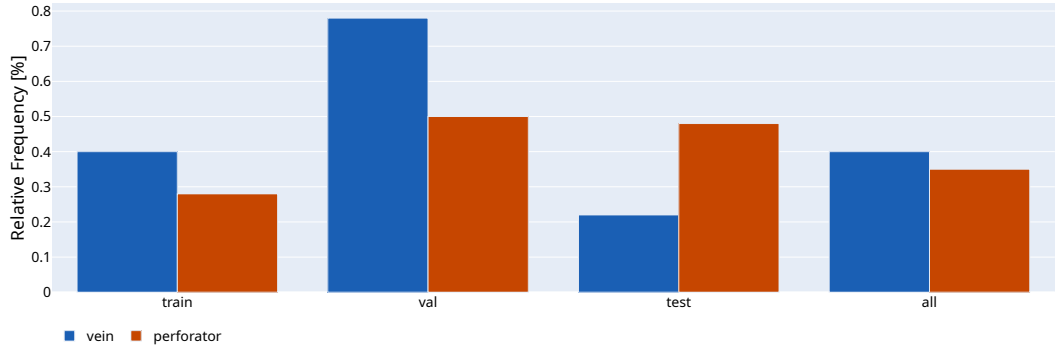
In figure 9.6 the dataset is analyzed for class imbalances. The background class and the non-vessel class are not shown because they are orders of magnitude larger than the other classes. For each of the body parts and the vessel, the relative class occurrences among all images in the dataset are shown. The comparison of the three subsets shows that the training set represents at best the total class occurrences, while the validation set has a higher number of calves as well as a higher number of veins and slightly more perforators. The test set underestimates the frequency of calves and has a different ratio of veins to perforators.

9.3 Thermogram Segmentation

The automatic thermogram processing pipeline has its main steps in the segmentation of body parts and vessel patterns within each thermogram. Therefore, the two developed deep neural networks are optimized to achieve high performance. This chapter first presents the results of the models in our publications [60, 59, 7] and then the results of the hyperparameter optimiza-



(a) BPN classes, background excluded.



(b) VN classes, background and non-vessel excluded.

Fig. 9.6.: Dataset class frequencies for body and vessel datasets, separated by training, validation and test sets.

tion (HPO) for the BPN and VN. The HPO includes a larger manual annotated dataset. The overall frame rate for loading data, applying radiometric calibration, inferring body parts, applying body part filters, inferring vessel patterns, and computing thermal features is about ~4 fps, but the individual steps are not yet optimized for fast processing.

9.3.1 Body Part Network

The BPN has been analyzed for its performance in our papers [60, 7] with different training dataset sizes and different manually optimized hyperparameters. The first paper, employing the Attention-U-Net architecture, Dice loss, AdaBelief optimizer, batch size 8, and a learning rate of 0.0001, achieves a total test set IoU of 0.8. However, the classes in this paper differ from the definition proposed in the second paper. This BPN only separates the skin parts of the legs from the clothed leg and other parts of the image (see figure 6.2b). The

model was selected based on the highest validation IoU for all classes (0.92 after 324 epochs). The results per class are shown in table 9.2. However, they are not comparable due to different class definitions.

Class	Test IoU
All classes	0.80
Background	0.91
Leg	0.93
Clothes	0.76

Tab. 9.2.: IoU results for the test set for a reduced BPN class set and reduced dataset size (263 training, 75 validation, and 87 test). [60]

In [60] the IoU was calculated with micro-averaging. That is, first the predictions and labels are summed over the batch and then the IoU is calculated. This is in contrast to the IoU in this work, which is based on macro-averaging, where each class IoU is determined individually, averaged per image, and then aggregated over the batch. As a result, the total and mean IoU values per class differ.

Network	B-A		B-B
Train data	472	472	540
Validation data	164	164	130
Test data	160	200	200
Class			
Mean IoU	*0.6881*	*0.6644*	***0.6752***
Background	0.9738	0.9690	**0.9795**
Left upper leg	0.6797	**0.6679**	0.6489
Left lower leg	0.8440	0.8061	**0.8255**
Left knee	0.5565	**0.5407**	0.5288
Left clothes	0.5078	0.4573	**0.4701**
Left shoe	0.6163	0.5923	**0.6581**
Right upper leg	0.7511	**0.7456**	0.7151
Right lower leg	0.8872	0.8538	**0.8619**
Right knee	0.6268	**0.6114**	0.5928
Right clothes	0.4482	**0.4353**	0.4318
Right shoe	0.6775	0.6289	**0.7141**

Tab. 9.3.: IoU results for BPN: BPN B-A has the hyperparameter configuration from [7] and B-B has the best hyperparameters from HPO. For comparability, B-A and B-B are tested against the full manually annotated test set of 200 images.

The improved dataset and results were presented in [7] with the extended class definitions, but still with a reduced dataset (472 training, 164 validation and 160 test images). The IoU results are shown in table 9.3 (network B-A). The

network architecture is DeepLabv3+, the loss function is Dice, the optimizer is AdaBelief, the batch size is 4, and the learning rate is 0.001. The chosen model achieves a test score of 0.69 IoU. However, the important classes are left and right calf, where the IoU gains per class are 0.84 and 0.89. In this work, we further expanded the dataset to a total of 670 training/validation images and 200 test images. As a continuation of the previously published results, we perform a HPO to optimize the result. For the BPN, 114 trials were performed (see full results in appendix table B.1). The best hyperparameter set is the combination of a DeepLabv3+ architecture and the loss function RMI, optimizer AdamW, batch size 8 and learning rate 0.000686 (network B-B). The combination achieves a best validation IoU of 0.844 after 47 epochs. With less test data in [7], the results of B-A are not directly comparable to B-B. Comparing both networks B-A and B-B with 200 test images, the newly found network configuration B-B improves the overall performance over B-A and the individual class results for the left and right calves. The test result for B-B is 0.68 IoU in general and 0.83 and 0.86 for the calves (table 9.3).

Figure 9.7 shows a parallel plot[1] for the BPN hyperparameter search. Each value in a column represents a different value of a common category. On the left is the objective value (validation IoU), followed by batch size, loss function, learning rate, model architecture, and optimizer. In addition, the validation metric is also color-coded, with higher values in a darker color. The connections between the values represent a single hyperparameter configuration. The graph provides insight into the individual performance of a single hyperparameter and it's influence on the result. The more times a value is crossed by a path, the more often it has been selected by the HPO algorithm and therefore has a positive influence on performance.

The normalized confusion matrix (figure 9.8) for the validation set visualizes the individual validation performance of each class to further evaluate the best model. The best performing class with the fewest false positives and false negatives is the background class. All other classes have a small number of false background predictions. The left and right calf follow with a high true positive detection rate of over 0.97. The predictions of the shoe classes have more false positives for the background than for the lower leg due to the low contrast of the shoes T_{sr} compared to the background and the lower temperature scale limit. The thighs have many false negatives and false

[1] A parallel (coordinate) plot provides a visualization for multivariate data to get an overview of the analyzed data. However, it is not easy to extract individual information. It is often applied to high-dimensional representations.

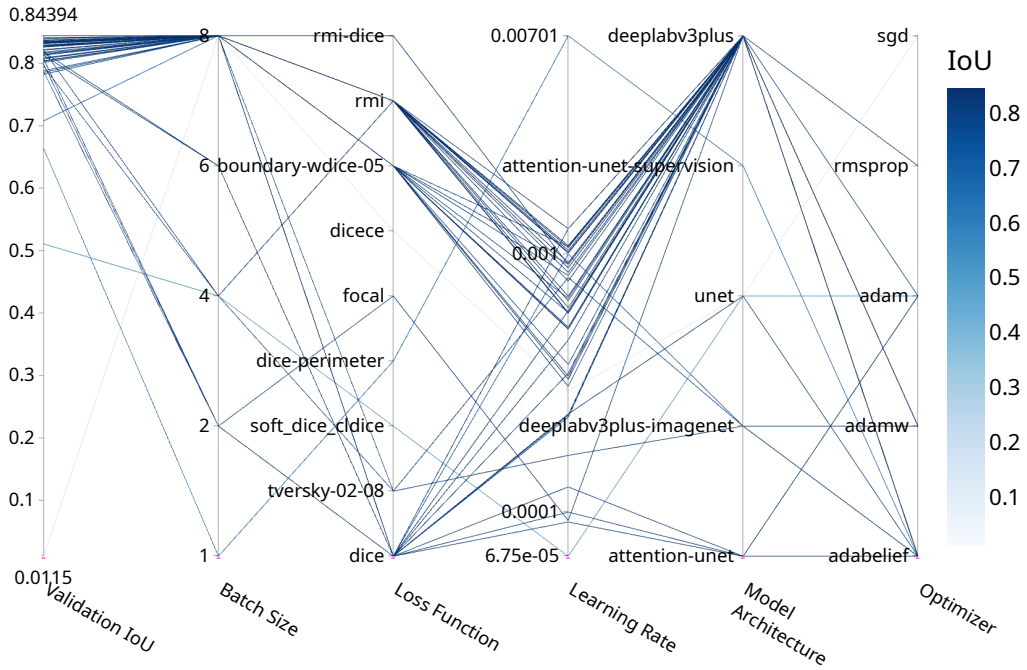


Fig. 9.7.: The parallel plot of the HPO for the BPN shows the different trials according to the final alidation IoU and categorized by the different hyperparameters. The figure gives an overview of the performance when a specific hyperparameter is chosen, and thus shows whether a hyperparameter is a good choice for a high IoU or not. For example, the RMI, Boundary-WeightedDice-05, and Dice losses are good choices, while Dice-Cross-Entropy has no trial with good performance.

positives in the knee and clothing regions, but also to the other side. The knee has similar behavior, its pixels are misclassified in the neighboring classes calf and thigh, but also in clothing. Side misclassifications also occur, but less frequently.

The BPN results do not directly generate a final segmentation mask. There is further post-processing to filter out bad results and perform consistency checks. Figure 9.9 shows the results of the filter method on samples from the test set. The goal is to analyze the calves, so the method is optimized for calf detection. In the first images, they are processed as intended. Legs that are too small, too bent in (b). In (c) segmented parts are wrongly removed and tight shorts are also recognized as legs. The case (d) gives an example where both legs overlap, i.e. they cover each other. Often, the network is not able to distinguish the correct shapes in this case, so both legs are removed. However, in (e) and (f) the legs are removed even though they should meet the criteria. In both cases, the filter methods are appropriate for these situations,

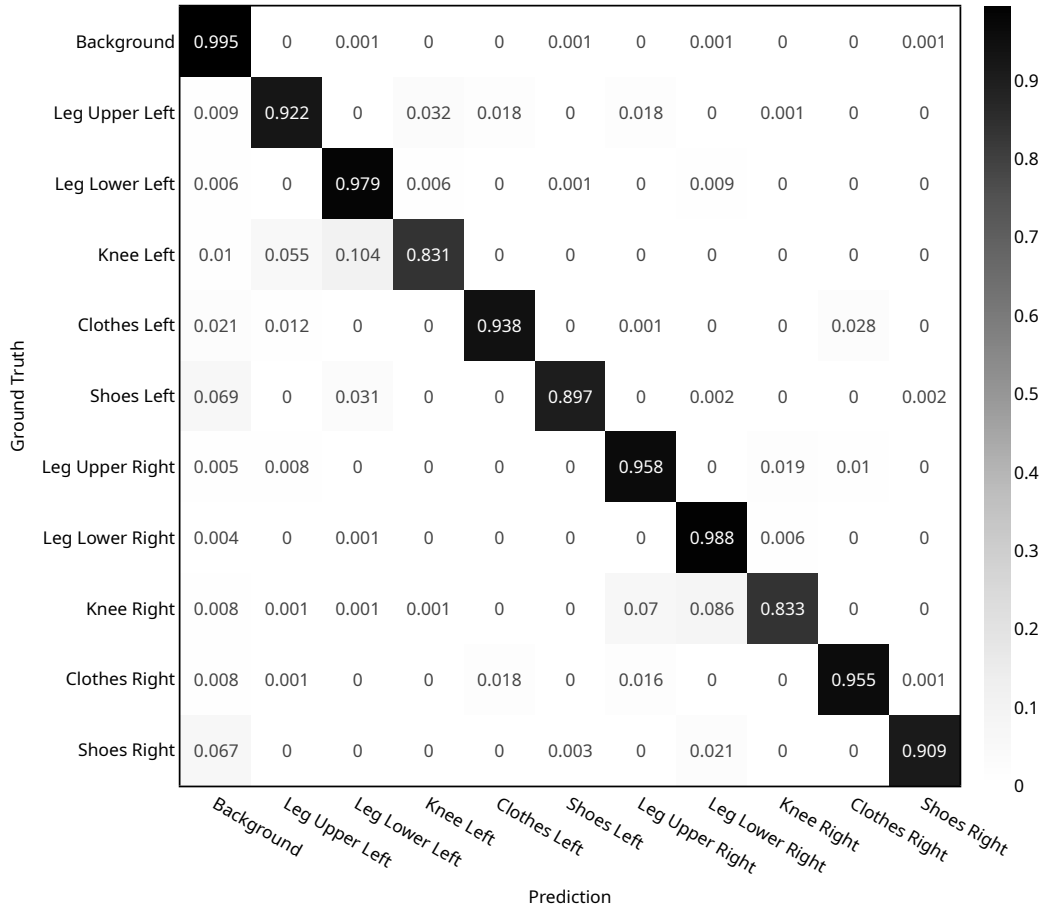
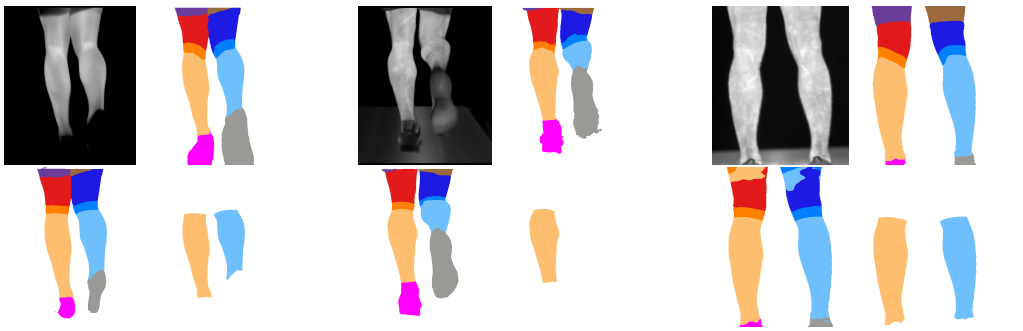


Fig. 9.8.: Normalized confusion matrix for the best epoch (47) of the best model in BPN HPO for validation data. The main diagonal represents the case where the model correctly predicted the actual class at a given pixel. The other cells represent false predictions. Depending on the view, false negative detections for a ground truth class are denoted in the row (except for the i-th value) and false positive detections are denoted in the column (except for the j-th value).

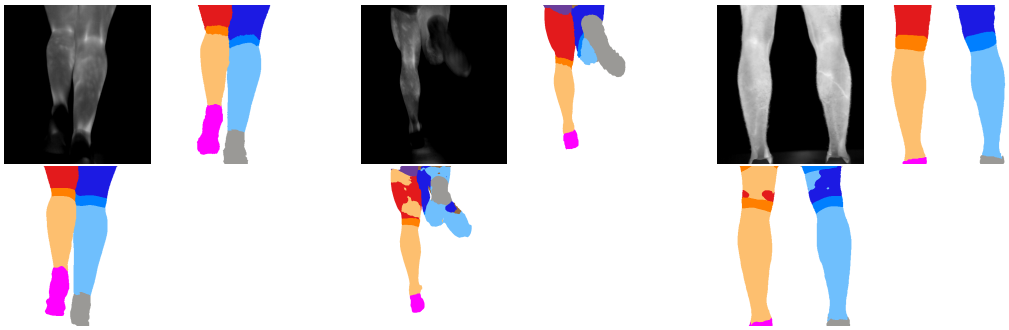
but return incorrect results due to their configuration. In (e) the left shoe was recognized as the right shoe, which is wrong. Therefore, further processing filtered the left calf because it is below the incorrectly recognized shoe. In (f), the calves are removed because the size threshold for the convex hulls is too low. In this case, not only outstanding parts are filtered, but also the valid shapes.



(a) Both legs are accepted.

(b) The right leg is removed because it is too small.

(c) Small misclassifications will be removed. Tight shorts are considered legs.



(d) Legs are touching in the thermogram. Even though the shape is recognized correctly, they need to be removed because one leg covers the other a bit at the intersection.

(e) Wrong removal of the left leg, correct removal of the right leg, because of wrong shoe detection and therefore the left leg was compared to the right shoe, which is filtered.

(f) Wrong removal of both legs, because of convex hull area difference filter.

Fig. 9.9.: Examples of filtering the BPN predictions with post-process consistency checks. Labels and prediction are reduced to left and right calves + background. Each image consists of four parts: top left: thermogram; top right: label; bottom left: prediction; bottom right: filtered prediction (only calves).

9.3.2 Vessel Network

The parallel plot for the VN HPO in figure 9.10 gives an overview of the performance of the hyperparameter categories (loss, architecture, optimizer, learning rate, batch size) and how the possible hyperparameters perform in each category. According to the complete results (appendix table B.2), the combination of Attention-U-Net input together with Tanimoto loss, AdaBelief and a learning rate of 0.000052 achieves the highest validation IoU (0.666) after 64 epochs. Remarkable performance is achieved with Soft-Dice-clDice (trial 89, IoU 0.6619) and the combination used in [60] (trial 1, IoU 0.6579).

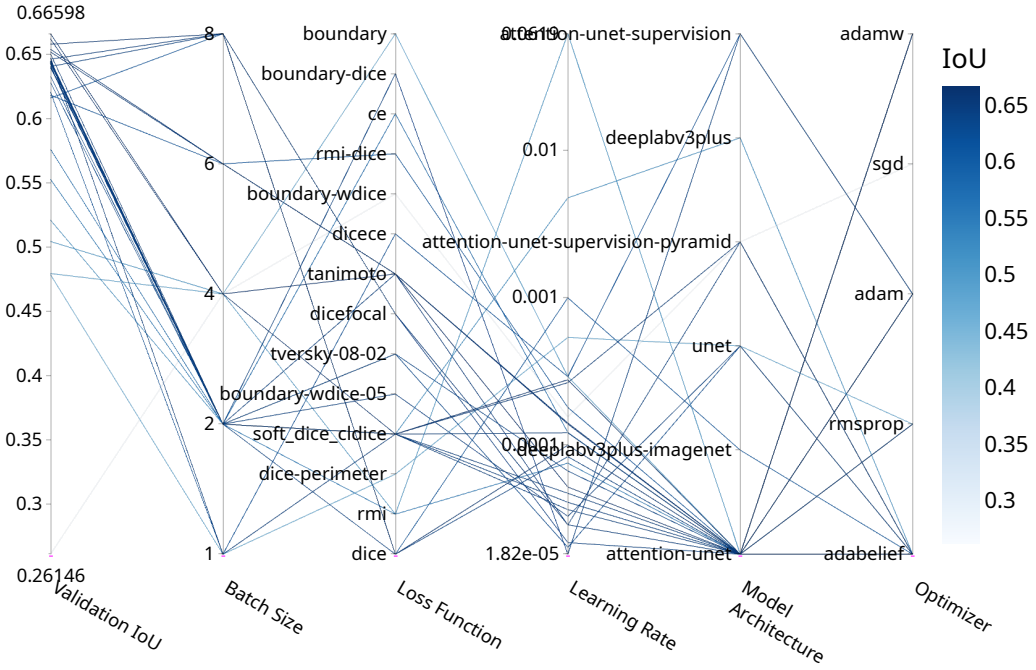


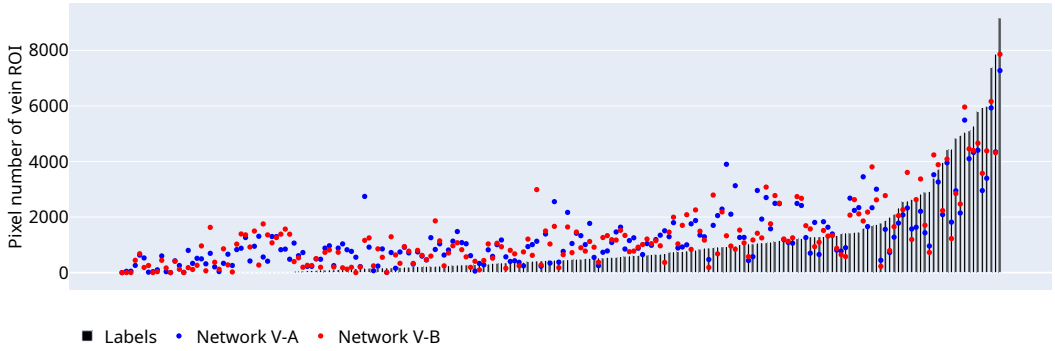
Fig. 9.10.: The parallel plot of the VN HPO combines the objective value (validation IoU) with the presentation of individual hyperparameters based on their category to give an impression of the influence of each value to a high IoU. For example, many combinations with high performance utilize the Attention-U-Net, while other architectures were tested in fewer trials due to poorer performance.

The vessel network V-A (see table 9.4) trained an Attention-U-Net with AdaBelief optimizer and Dice loss [60]. The learning rate was set to 0.0001 and the batch size to 8. The dataset contained 263 training, 75 validation, and 87 test images. All images were captured with the VarioCam hr. The highest validation IoU was reached after 549 epochs. Additionally, in the training phase the thermograms are cropped and resized to 256×256 pixels for computational reasons. In this work, we increased it to 640×480 pixels. Network V-A was also tested with the current test set of 200 images. Network V-B is trained with the hyperparameters found in HPO. The full manual dataset was included (540 training and 130 validation images). 100 trials were run with all combinations. For the vein and perforator classes (network V-A in table 9.4) the IoU is obtained with macro-averaging instead of micro-averaging as in the publication [60].

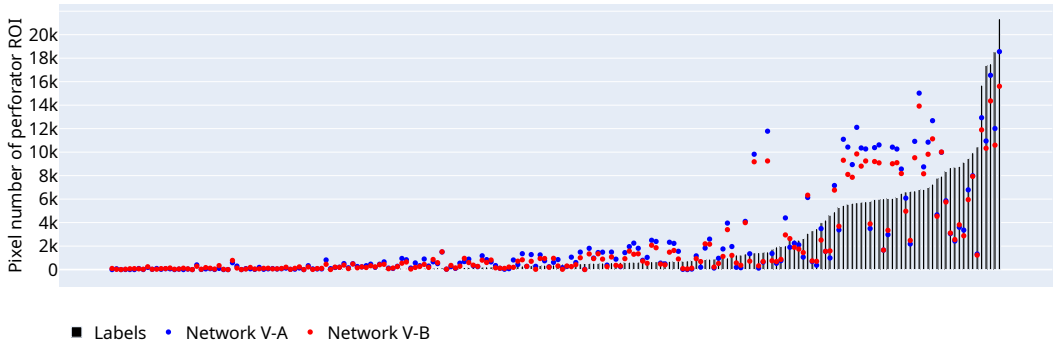
Regarding the comparison of network V-A and network V-B, the former outperformed the new network with optimized hyperparameters. Network V-A has slightly better performance than network V-B. Figure 9.11 compares the

Network	V-A		V-B
Train data	263	263	540
Validation data	75	75	130
Test data	87	200	200
Class			
Mean IoU	*0.5771*	***0.5734***	*0.5568*
Background	0.9924	**0.9971**	0.9960
Vein	0.2272	**0.1893**	0.1680
Perforator	0.1978	**0.1825**	0.1384
Non-Vessel	0.8909	0.9246	**0.9248**

Tab. 9.4.: IoU results for VN: V-A has the hyperparameters as defined in [60] and V-B has the best hyperparameters from HPO. Test results for V-A are reported with the test set from the publication (87 images), but for comparability also with the extended data (200 images).



(a) Vein ROI.

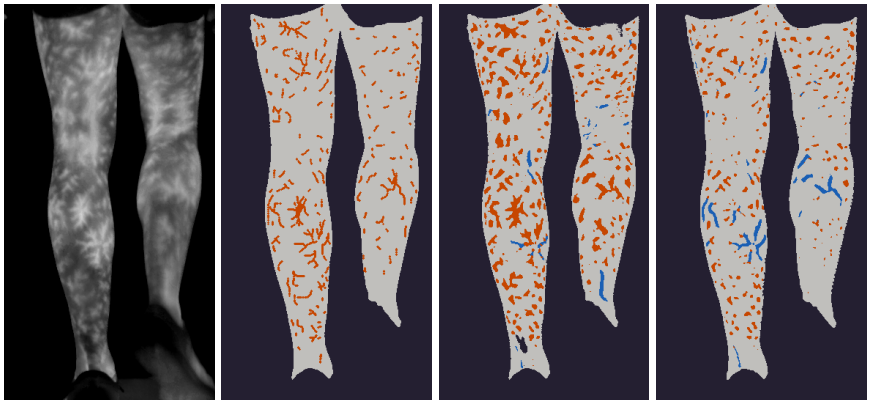


(b) Perforator ROI.

Fig. 9.11.: ROI segmentation results for the test set for the networks table 9.4 V-A (blue) and V-B (red). The black bars show the ground truth data for the ROI. The values represent the number of segmented pixels for that class. The plot is sorted by the label size of the ROI.

performance in more detail. For the vein and perforator classes, the size of the ROI in each test sample is plotted. The bars represent the ground truth

data. Both plots are sorted individually by the size of the label ROI. The blue dots represent the network V-A and the red dots represent the network V-B. It can be seen that the larger the ROI labels, the worse the result. The phenomena is demonstrated in figure 9.12, where the label is shown together with the predictions of both networks. The predictions estimate the rough shape well, but miss the correct boundaries, its size, and connections to other labels. This leads to an under- or overestimation of the region. In addition, both networks introduce the vein class even though there is no vein in the label mask. The individual IoU values per image show the low performance of the vein and perforator patterns, including the 0 for the newly introduced veins. The appendix figure B.1 shows more examples from the test set where the basic form of the vessels is correct, but the concrete shape is over- or underestimated and thus lead to low IoU. Nevertheless, the individual image IoU is higher than the overall network result, e.g. appendix figure B.1d–f.



(a) Thermogram. **(b)** Label. **(c)** V-A prediction. **(d)** V-B prediction.

Fig. 9.12.: Example results for the VN from the network V-A and V-B compared to the ground truth. IoU values for (c): mean: 0.5426, background: 0.9993, vein: 0, perforator: 0.3146, non-vessel: 0.8564. IoU values for (d): mean: 0.516, background: 0.9978, vein: 0, perforator: 0.1925, non-vessel: 0.8736.

9.4 Label Generation

For label generation, a specialized study was obtained: StereoThermoLegs (section 5.2). 14 participants were measured and their images were processed in the proposed label generation method for infrared thermography (IRT).

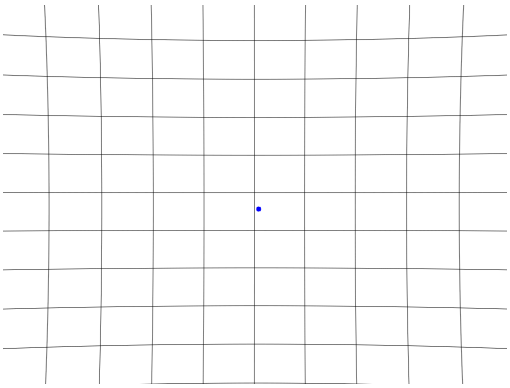
9.4.1 Calibration

The calibration pattern was acquired in different positions to calibrate the stereo cameras. Individual reprojection errors are manually inspected and images with high reprojection errors are excluded in another calibration run. For the IRT camera calibration, 64 images are valid and have a positive influence on the calibration. The color+depth (RGBD) camera has 199 images for calibration. The stereo calibration includes 60 time-synchronized image pairs from both cameras, where the reprojection error is < 0.4 for the RGBD image and < 0.6 for the IRT image. The average reprojection errors of the calibrations are reported in the table 9.5.

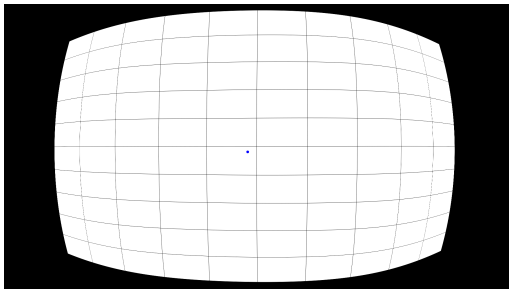
Calibration	Images/Pairs	Excluded	Avg. reprojection error
IRT camera	69	5	0.163069
RGBD camera	204	5	0.159427
Stereo run 1	65	5	0.216671
Stereo run 2	65	5	0.216671

Tab. 9.5.: Average reprojection error when calibrating a stereo system with IRT and RGBD cameras. The stereo calibration is done in two passes, with the second pass using the extrinsics from the first pass as an initial guess and continuing the optimization.

A lens is attached to both cameras. The lens distortion effects are visualized in figure 9.13. The part image (a) shows a pincushion distortion for IRT and a barrel distortion for RGBD (b). The optical center is estimated to be at the rounded pixel position $\begin{bmatrix} 519 \\ 413 \end{bmatrix}$ for the first camera and $\begin{bmatrix} 927 \\ 560 \end{bmatrix}$ for the other.



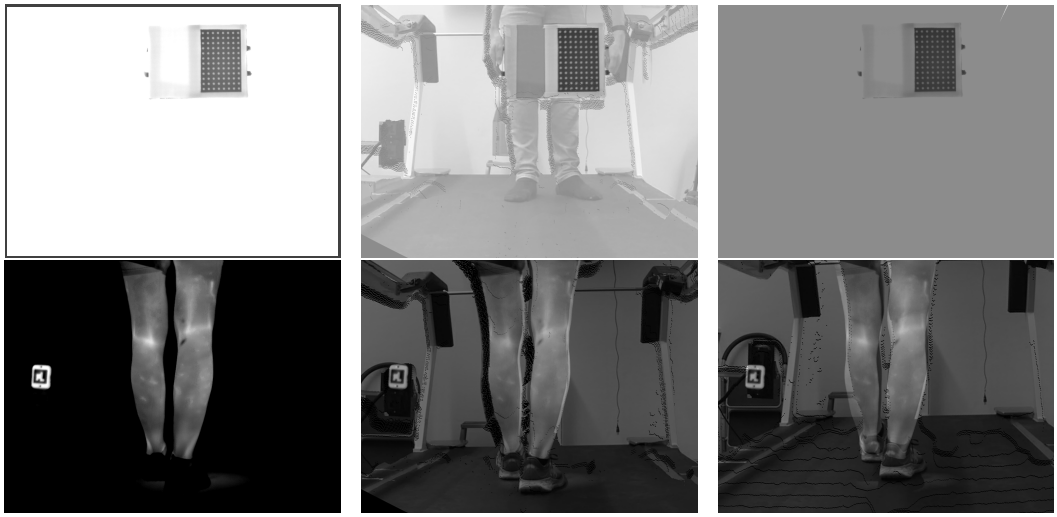
(a) IRT camera with image size 1024×768.



(b) RGBD camera with image size 1920×1080.

Fig. 9.13.: A grid pattern rendered at full image size and undistorted visualizes the distortion effects of the lenses used on the camera. Also, the blue dot represents the optical center.

Manual Extrinsic Correction With our proposed method of manually finding point correspondences in both image domains and recovering the extrinsic parameters, promising improvements in epipolar lines and point correspondences are achievable, as shown in the section 7.1.3. However, when the new extrinsics are applied to full images, the transformed image points are worse than with the old extrinsics. In figure 9.14 examples of the transformed images from RGBD to IRT are superimposed on the original thermogram. The first row shows an example for the calibration images and the second row shows an example from the experiments. In the column (b) the original extrinsics have been applied. For the calibration image, the pixels match well. While the lower image shows the small offset between the two domains. But with the corrected extrinsics, the transformation results in an even larger shift. For the calibration image, the transformation fails completely. Both show that the new extrinsics are locally good, but globally no improvement at all. The baseline for the original extrinsics is 13.66 mm and for the corrected extrinsics 50 mm. Aligning the baseline to the original extrinsics produces even worse transformations with larger shifts. Based on these results, the dataset generation was performed with the original extrinsics.



(a) Thermogram with different scales.

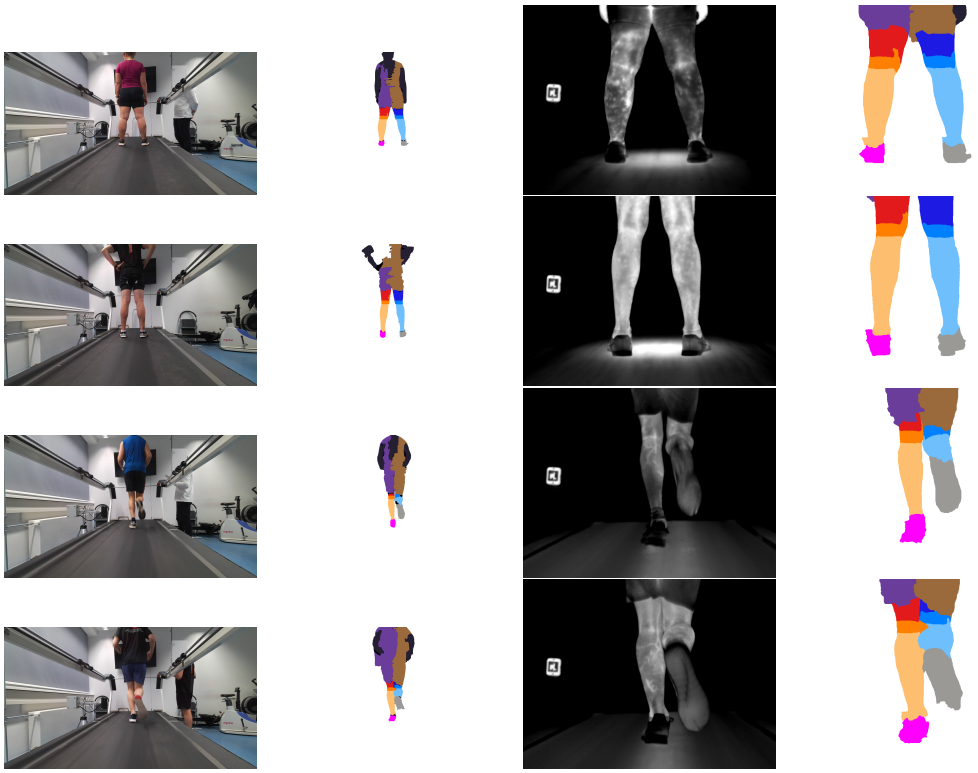
(b) Transformation with original extrinsics.

(c) Transformation with corrected extrinsics.

Fig. 9.14.: Stereo transformation with calibrated and manually corrected extrinsics. The first row shows a calibration image and the second row shows a study image. The transformed RGBD image is superimposed on the IRT image in the IRT coordinate system. Transformation with corrected extrinsics fails completely for a calibration image. The transformation for the study shows a greater shift with corrected extrinsics than with the original.

9.4.2 Dataset

Each study was cleaned for nonuniform calibration (NUC) phases. For each of the 14 studies, 10% of the image pairs were randomly selected. In addition, the test and training allocation among participants was also randomized. In total, 11 persons form the training/validation set with 12,826 image pairs, while the remaining 3 individuals form the test set with 3433 image pairs. On the training set, the image normalization parameters are computed and set to mean$= 0.084$ and SD$= 0.165$. These image pairs are processed for automatic label generation in IRT. Figure 9.15 shows example images from the resulting dataset (test set) at different stages of the runs. One line shows a corresponding pair of grabbed images, the generated label in RGBD and the result label in IRT.



(a) RGBD image. **(b)** Generated RGBD label. **(c)** Thermogram. **(d)** Transformed and fine-tuned label for thermogram.

Fig. 9.15.: Example images from the StereoThermoLegs dataset at different stages. [6]

9.4.3 Benchmark Results

The network S with DeepLabv3+, Dice, and the AdaBelief optimizer reached its optimum after 17 epochs (27 epochs trained with early stopping) and achieved an overall test IoU of 0.6630. The table 9.6 shows the IoU per class. The first result is reported for the training and test data from the automatic labeling approach. The results per class also show the side differences, which are similar to each other. However, the differences between different body parts are higher. The main target class to evaluate is the calf/lower leg. For both (left and right calf), the network achieves the highest results. Inspection of the thermograms shows that these ROIs of the body are mostly visible during the whole experiment and only sometimes occluded during fast running phases. They are usually not covered by shoes or clothing ensured by the study design. The opposite is true for the thighs. With different types and sizes of pants and overlapping legs, the upper part is not visible and is clearly separated from the lower part, resulting in a lower IoU. Also, some types of clothing are tight and some are loose. Tight pants allow thermal radiation to pass through and also appear similar to naked skin. Loose pants move unpredictably with movement, and the underlying thermal activity of the skin does not pass through them. The knees, which are the boundary between the upper and lower legs, also have a worse IoU. There is no natural boundary of the knee in our labels and less clear definition for different annotators. Shoes and clothing are even more difficult to detect because they may have a similar or lower temperature than the background and therefore the same intensity in the image, making them indistinguishable. For shoes, the high speed is also irritating as well as overlapping with other parts.

The table also compares the test results with other networks, but with the manually annotated test set. First, the baseline network B-B and the new approach S are compared with the manually annotated test set. In addition, the network is further fine-tuned with the fractions of the manual dataset (S-100, S-50, S-10) and also tested. For S, the results are lower in all classes than the results from network B-B. The stereo training set and the manual test set are from two different distributions. The manual dataset also contains images from the VarioCam hr, while the StereoThermoLegs dataset consists only of thermograms from the VarioCam HD. There are several participants for training, but all are from the same study design with the same camera configurations. Variations were introduced by forcing different types of shorts and different lengths of socks. The limited variation in the stereo training

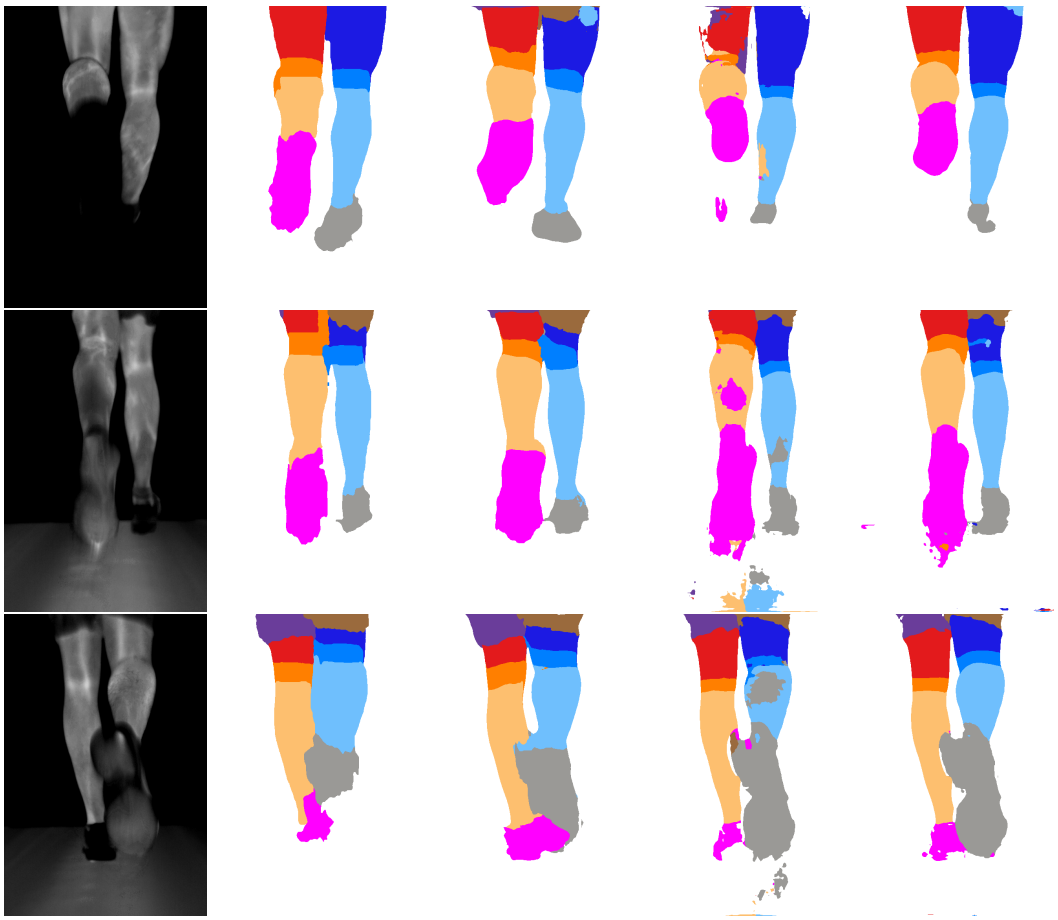
Network	S		B-B	S-100	S-50	S-10
Train data	stereo	stereo	manual	stereo	stereo	stereo
Fine tuning	-	-	-	100% manual	50% manual	10% manual
Test data	stereo	manual	manual	manual	manual	manual
Class						
Mean IoU	*0.6630*	*0.5067*	*0.6752*	***0.7166***	*0.7088*	*0.6076*
Background	0.9783	0.8897	**0.9795**	0.9790	0.9783	0.9684
Left upper leg	0.6678	0.5065	0.6489	**0.7640**	0.7428	0.5984
Left lower leg	0.8027	0.7315	0.8255	**0.8650**	0.8440	0.7846
Left knee	0.5989	0.3828	0.5288	**0.6078**	0.5881	0.4987
Left clothes	0.4796	0.3457	0.4701	0.5063	**0.5092**	0.4427
Left shoe	0.6357	0.4198	0.6581	**0.6804**	0.6643	0.5568
Right upper leg	0.6077	0.5101	0.7151	0.7738	**0.7758**	0.5721
Right lower leg	0.7758	0.6231	0.8619	**0.8806**	0.8752	0.7833
Right knee	0.6205	0.3657	0.5928	0.6386	**0.6420**	0.5145
Right clothes	0.5099	0.3175	0.4318	**0.4734**	0.4623	0.3930
Right shoe	0.6161	0.4808	0.7141	0.7143	**0.7147**	0.5713

Tab. 9.6.: IoU results for BPNs with different training and test sets. All BPNs have the same hyperparameters. The comparison with the network B-B (trained only with manual data) is performed with the manually annotated test set. The network S is also tested with stereo data. Fine-tuned networks take the network S and train the parameters with a fraction of manual data. The stereo data refers to the StereoThermoLegs dataset [6] with 3433 test images and the manual data refers to the 200 test images presented above.

data does not generalize well to multiple scenarios tested with the manual test set.

In addition to the trained stereo BPN, the network can be tuned with manual data: all manual data S-100, half of the data S-50, and finally 10% S-10. Fine-tuning with all data increases overall performance. To see if less manual data still improves performance, the 50% and 10% data sets were tested. The 10% fine tuning achieves a score of 0.6076 after 4 additional epochs, with 50% of the manual dataset an average IoU of 0.7088 is obtained (28 additional epochs), and the complete dataset has 0.7166 (25 additional epochs). Compared to the BPN trained only with manual data (B-B, IoU: 0.6752), the fine-tuned networks with 50% and 100% manual data perform better. Improvements are observed in all classes except the background class. With the 10% fine tuning, there is still an improvement over S, but not over B-B. This shows the potential of the stereo dataset to start the development of new networks with new ROIs. In addition, the amount of manual labeling could be halved to achieve similar results. The outlined stereo approach is far from perfect in label generation

and transformation, but still improves the results for fine-tuning with a small, high-quality, manual dataset.



(a) Thermogram. **(b)** Generated label. **(c)** Prediction BPN S. **(d)** Prediction BPN B-B. **(e)** Prediction BPN S-100.

Fig. 9.16.: Thermograms together with their generated label (b), the results of the BPNs S (c), B-B (d) and S-100 (e) in different stages. The thermograms are taken from three people in the StereoThermoLegs test set.

Figure 9.16 shows samples from the three people in the stereo test set along with their segmentation of the networks S, B-B, and S-100. As the IoU points out, the major drawbacks are seen mainly in the knee parts, the clothes and shoes. Although the performance of IoU is not perfect, manual inception identifies some predictions that are better than the generated labels. In the last row, the area between the legs must be background. This is not the case for the generated labels, but all networks generalize well enough to correctly determine the background. Compared to the performance of the network B-B, the visual results are better because the data distributions between the two datasets are different. However, fine-tuning the S network with manual

data (=S-100) improves the results. In the appendix figure B.4 the comparison is also applied to five people from the manual test set. The visualizations show that the network S roughly segments the body, but with many artifacts. Visually, the main improvements of the fine-tuned network S-100 over the manually trained network B-B are the better segmentations with less noise in other ROIs than the calves.

9.4.4 Applied Thermogram Analysis

The study data can also be analyzed by comparing their results in thermal statistical analysis. Therefore, we evaluate the statistical properties of the study data either with the transformed labels and compare them with different network results from the benchmark S, the previous work B-B, and the fine-tuned network S-100. This analysis shows the results for the left calf class. The comparison between the four methods in figure 9.17 shows that the networks trained with stereo data (S/green and S-100/magenta) provide similar thermal statistics as the generated labels (red). For person 1, there is a larger discrepancy between minutes 5 and 11. For the other two participants, there is only a small offset between the methods. However, network B-B with manually annotated data has a different characteristic, e.g. more noise and a different curvature in person 2. The overall temperature behavior over time is similarly covered, but the method predicts areas with higher mean temperatures. The fine-tuned network S-100 still preserves the main thermal features, but is slightly shifted towards the results of B-B.

9.5 Sensor Fusion

The sensor fusion part is crucial to compare the new processing unit of thermogram analysis with well-established methods like breath analysis (spiroergometry) or heart rate based insights. The presented processing pipeline shows a valuable workflow to merge all data into a common time system and to enable further analysis like statistical methods to gain physio- and pathophysiological information. In this chapter we present the results of different steps of the pipeline. For both acquisition systems (section 8.1 and

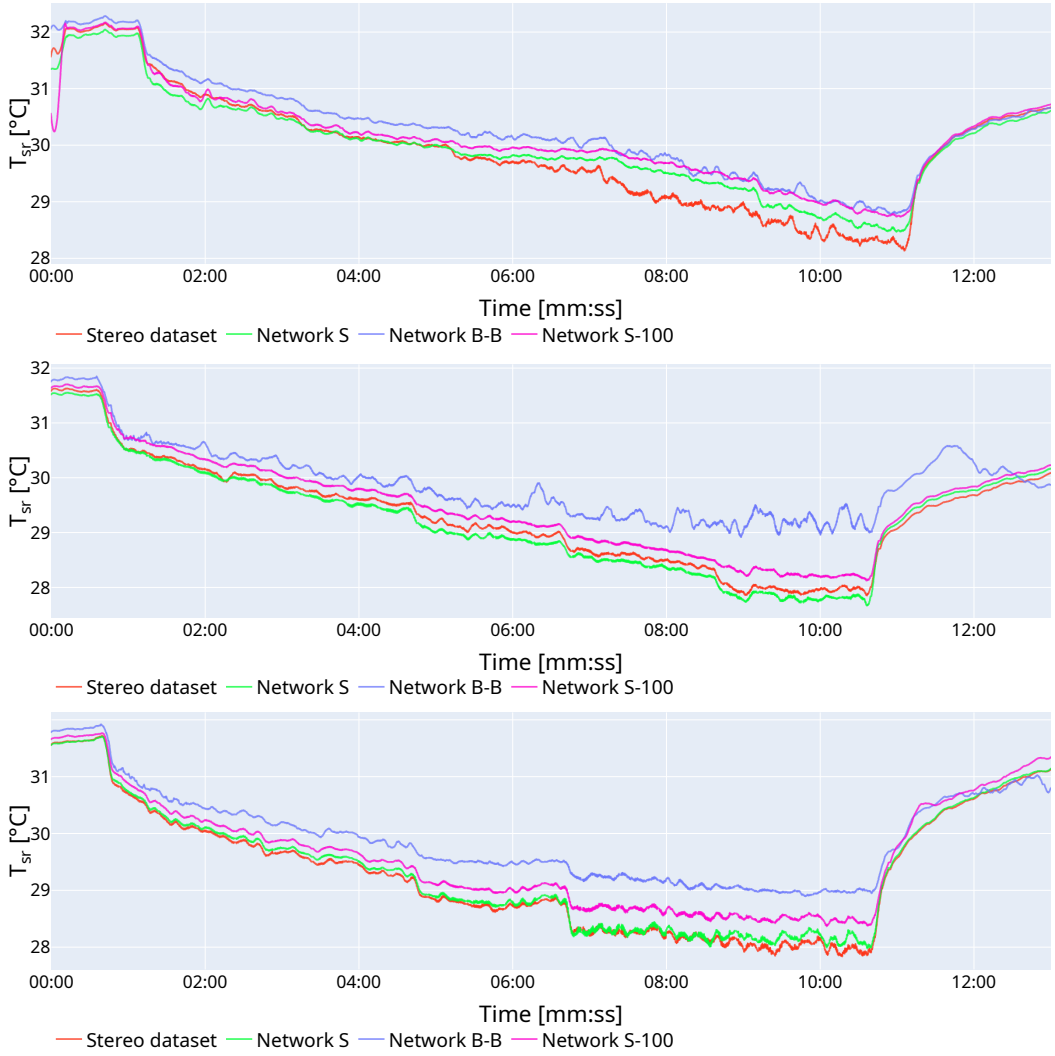
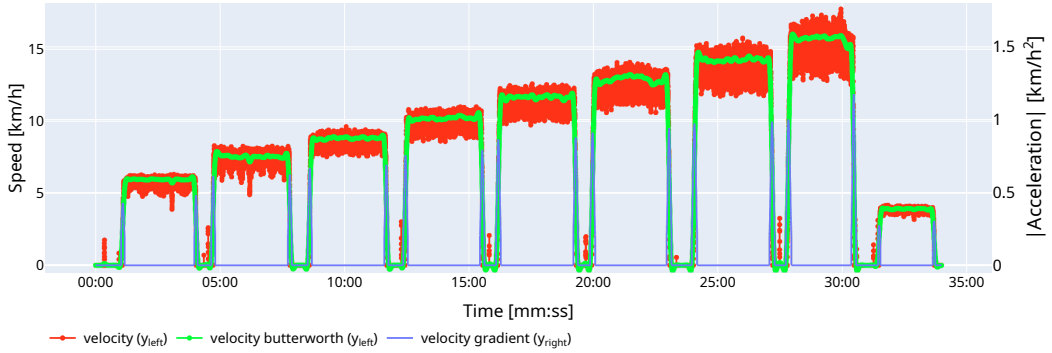


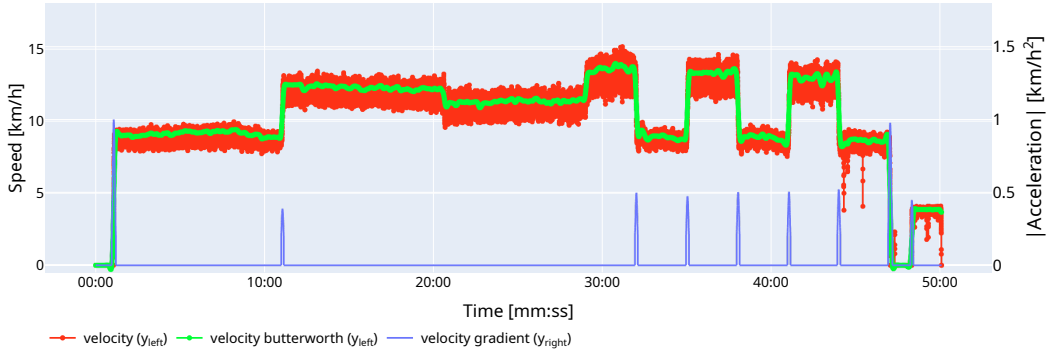
Fig. 9.17.: The mean thermal radiation of the left calf for each participant with different body masks: the labels, the network S with trained stereo data, the network B-B with trained manual data, and the fine-tuned network S-100. Values are smoothed with a Savitzky-Golay filter [152] (window length: 151).

section 4.1.4), the storage bandwidth of the loosely coupled recording application was sufficient and no data loss due to slow I/O operations or slow processing was observed.

The treadmill's speed is not properly recorded by the speed wedge, as indicated by the figure 9.18a of the red data points. The noisy behavior does not represent the correct speed of the treadmill. However, we can reliably find different phases in the data. The phases include standing (0 km/h) and different speeds. The gradient $>= 0.2$ km/h^2 (blue) is taken from the smoothed



(a) Step protocol (T0) from the Incoreloop study.



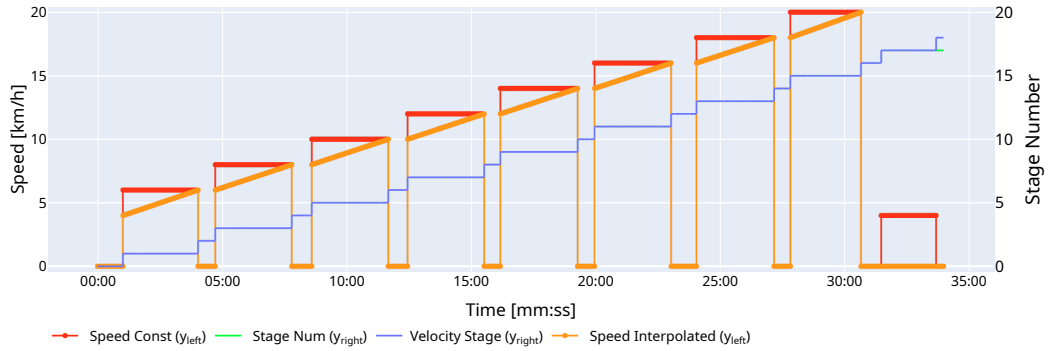
(b) T2 protocol from the Incoreloop study.

Fig. 9.18.: Example velocity processing steps for two different experiments from the Incoreloop study. The unfiltered velocity values are shown (red), the filtered (green), and the clipped gradient (blue).

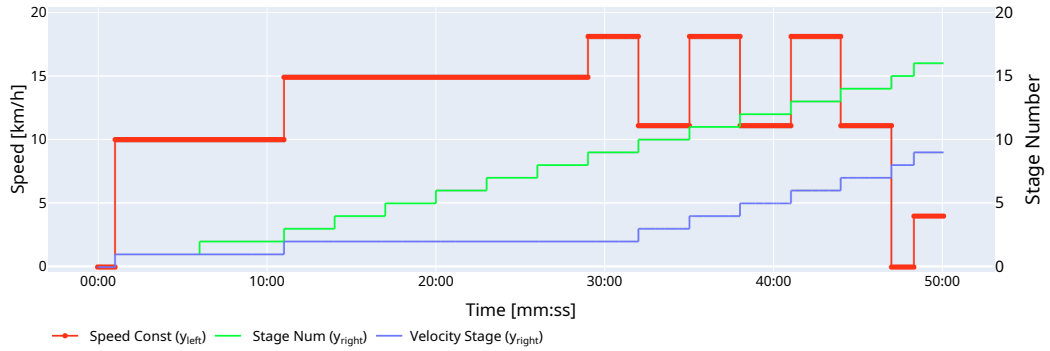
speed data (green). The peaks mark the boundary of a stage. However, in rare cases the sensor data has inconsistencies (e.g. figure 9.18b at ~00:19) that indicate a stage boundary where there shouldn't be one, or small differences in real speed measurements are not detected (~00:29). Changing the gradient cutoff threshold improves one case but worsens the others. Therefore, these cases will be fixed with the protocol definition in the next processing step.

The stages from the manual protocol notes have been successfully matched, as indicated by figure 9.19. A new stage number is given for each pause and run phase. The stages from the protocol are matched to the real stage boundaries of the speed stages if they make sense (time check). Otherwise an assumption is made, as in (b), to have multiple protocol stages within a single speed stage (00:11–00:29), and also to find matching stages where the speed detection goes wrong (00:29).

To select only the straightened and largest calf parts of the leg in the cyclic movement, the peaks in the size evolution are extracted. In figure 9.20 two



(a) Stage protocol (T0) from the Incoreloop study. Light orange represents virtual interpolated velocity between stages. Speed stage and protocol stage are the same except at the end.

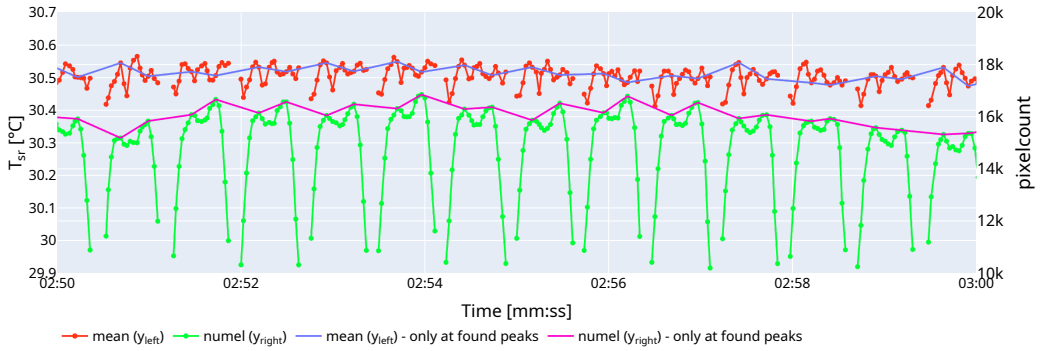


(b) T2 protocol from the Incoreloop study.

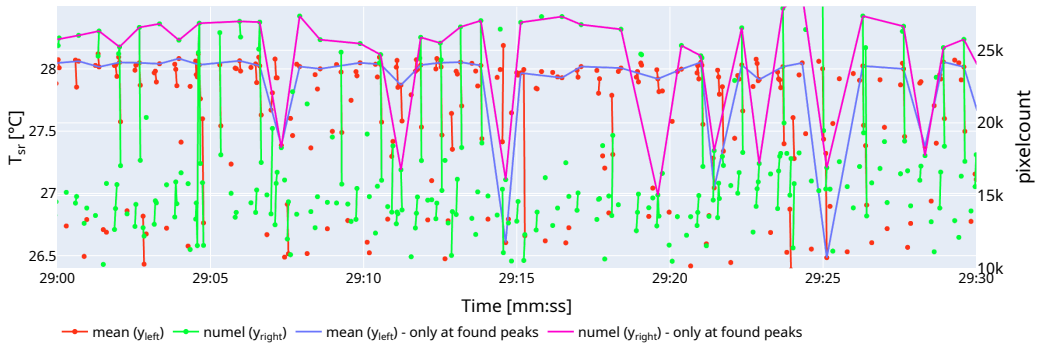
Fig. 9.19.: Example of matched stages for two different experiments from the Incoreloop study. The speed stages (blue) were matched according to the protocol design (green). Red indicates the protocol speed at that stage.

samples of a left calf are shown at the beginning (slow walking speed) and at the end (high running speed) of an experiment. In (a) the peaks of the sizes (green) are found and overlaid as filtered points (purple). However, there are also some intermediate points with small and more local maxima. The segmented points also show the corresponding T_{sr} (blue) as well as all T_{sr} values (red). In the lower example, the later step is more error-prone. Each step consists of fewer valid points. Due to the consistency checks in the segmentation pipeline (see section 6.3), there may be no data where a valid data point is expected. Therefore, peak detection may also need to be adjusted. With faster motion, less data is available due to more occlusions and more motion blur, and the resulting T_{sr} data is more noisy.

The resulting data is smoothed, and figure 9.21 shows an example of T_{sr} and its smoothed variant $\hat{T}_{sr}$. The smoothed curve follows the main characteristics of the data. In low velocity phases, the original data has less variation and the



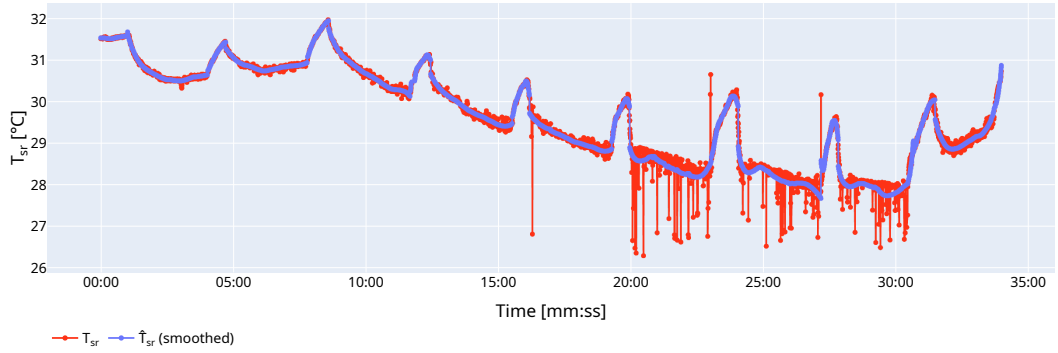
(a) Finding peaks within regular cyclic steps. Some intermediate peaks are also found.



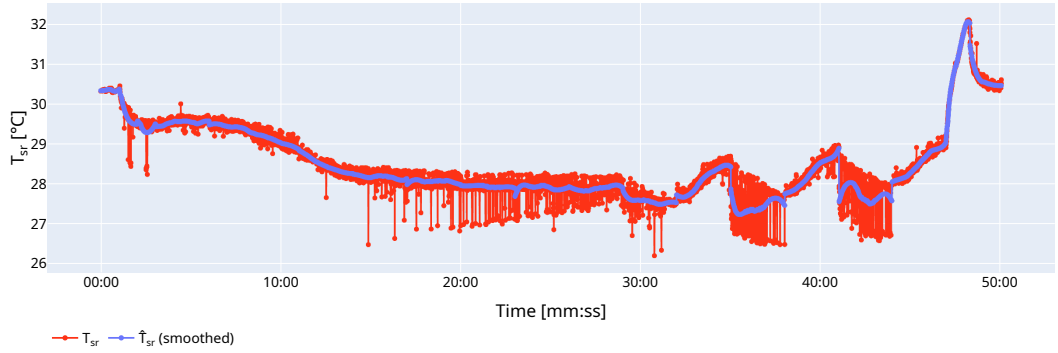
(b) Finding peaks within non-regular short cyclic steps. Not all peaks are found.

Fig. 9.20.: Finding the best image of a step for further analysis. Steps are cyclic and the largest size corresponds to a fully straightened leg, which produces the clearest thermograms. Peaks (purple) are found in the size series (green). The corresponding T_{sr} is shown in red and blue. The graphs show two short snippets of an experiment.

resulting curves fit well. At higher speeds, the previous filtering does not fit the data as well as it did at the beginning. An example is in (a) starting at ~00:20 or in (b) starting at ~00:35. The smoothed curve still have the same trend as before, but will be affected by the outliers and may under- or overestimate the expected values. Each stage is smoothed individually so that data from other stages (with different velocities) does not interfere. This is most important in the first case, where stationary phases occur. Therefore, not perfectly smooth transitions are estimated at the stage boundaries, e.g. (a) at ~00:20. However, the regions around a stage boundary are not reliable at all because of the acceleration and deceleration phases of the treadmill. The appendix figure B.2 shows the smoothed T_{sr} for the left and right calves for each person at the T0 experiment from the Incoreloop study. The curves change differently for each participant. Some have a lower T_{sr} in the first stage than in the second, but continue to decrease in later stages, others decrease the T_{sr} directly from



(a) Example from the Incoreloop T0 protocol.



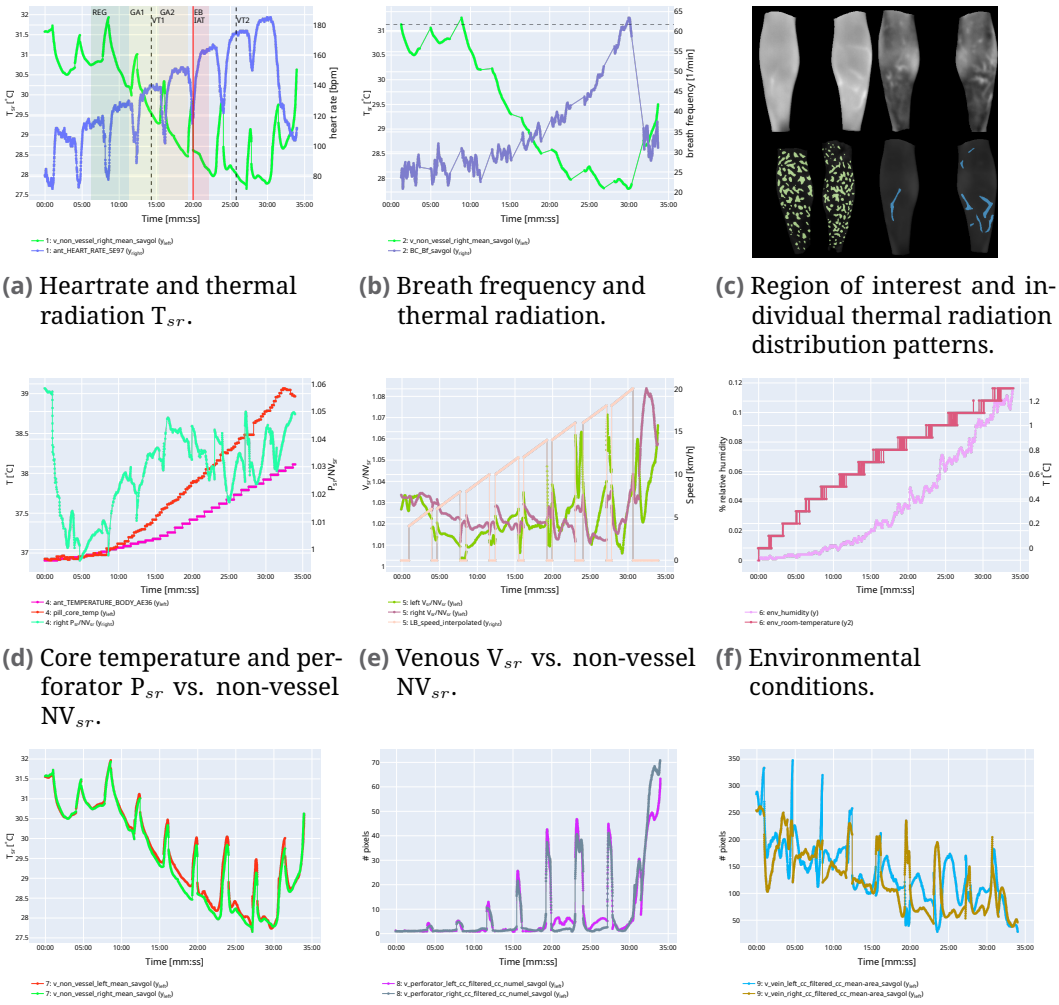
(b) Example from Incoreloop T2 protocol.

Fig. 9.21.: The figures show the smoothing of the noisy data $\hat{T}_{sr}$ and the original data T_{sr} and that the curve still follows the main characteristics.

stage to stage. Some also have a fairly constant T_{sr}. There is no clear behavior of the T_{sr} under the same protocol. In addition, for some people the left and right T_{sr} are similar, while for others they diverge from the beginning or after some time.

The overall analysis is presented in the form of an interactive dashboard. Figure 9.22 gives an impression (see appendix section B.4 for larger plots). Each field represents a different combination of thermal features and sensory data for further interpretation.

Each stage is represented by a single data point for comparison to the manual analysis strategy and to quickly assess key insights. Figure 9.23 shows the extracted ROIs per stage (standing image right after the end of the stage) for the example experiment. However, the reduced data leads to sparse plots, e.g. figure 9.24. The basic curvature of the T_{sr} is still visible. It decreases from stage to stage, but the effects of pauses and how the T_{sr} changes between stages are no longer accessible. A single standing phase (pause) thermogram



(a) Heartrate and thermal radiation T_{sr}.

(b) Breath frequency and thermal radiation.

(c) Region of interest and individual thermal radiation distribution patterns.

(d) Core temperature and perforator P_{sr} vs. non-vessel NV_{sr}.

(e) Venous V_{sr} vs. non-vessel NV_{sr}.

(f) Environmental conditions.

(g) ROI mean T_{sr} left vs. right.

(h) ROI perforator T_{sr} left vs. right.

(i) ROI veins T_{sr} left vs. right.

Fig. 9.22.: Exemplary results dashboard for a single experiment comparing different thermal characteristics with sensor data. In appendix section B.4, the charts are repeated in a larger version.

representing each stage has the advantage for comparability with manual analysis, that there is no motion blur or occlusion and vessel patterns can be seen simultaneously on both sides. In contrast, running phase thermograms contain blurred or occluded areas and are not comparable to manual analysis.

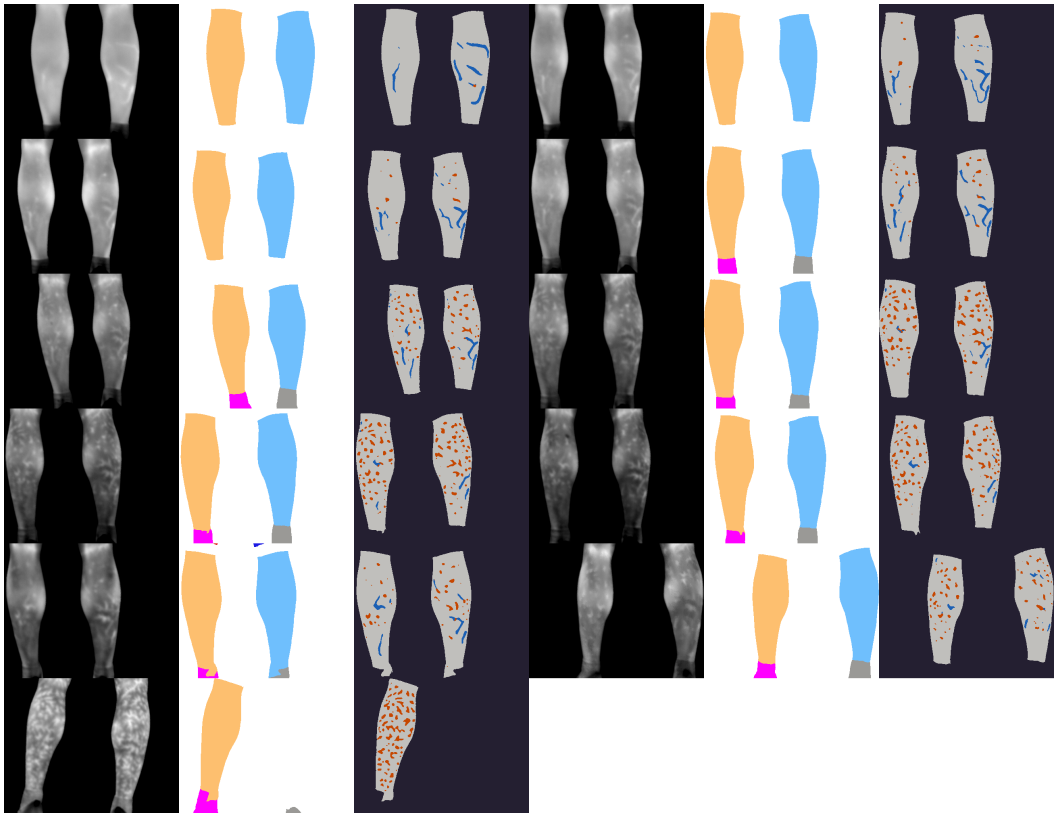
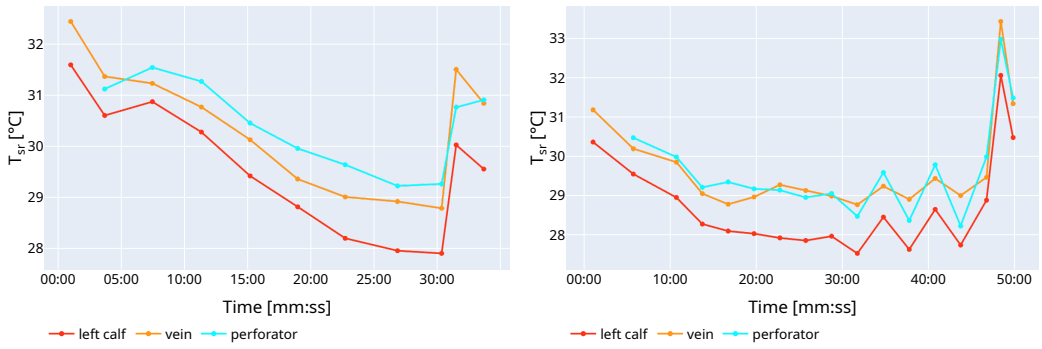


Fig. 9.23.: Image with its predictions after each stage in a standing phase (example from Incoreloop T0 protocol). Stages are from left to right, top to bottom: Pre, 1, 2, 3, 4, 5, 6, 7, max, Rec+1, Rec+3.



(a) T0 experiment.

(b) T2 experiment.

Fig. 9.24.: T_{sr} plot for the ROIs calf, vein, and perforator. One value for each stage image (except pauses).

10 Discussion

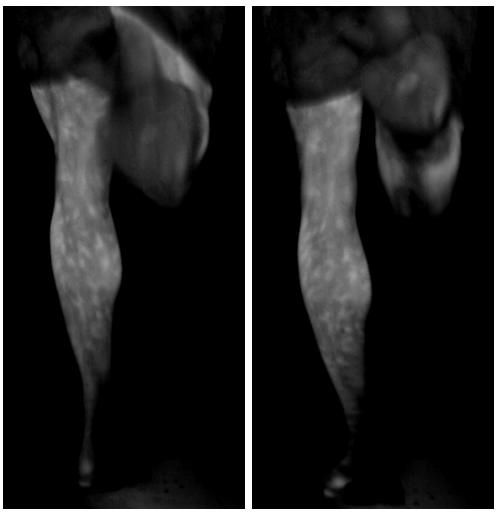
We presented the ThermoNet processing pipeline from data acquisition to image segmentation in two aspects, thermal feature extraction and sensor fusion with other data. In addition, we bootstrap the initial start for new region of interest (ROI) detection by automating the labeling process for neural network training. The results of the presented methods are promising for applications in medicine and sports science. In this chapter, we discuss the individual results and how the whole system supports further investigations and applications in medical fields.

10.1 Thermogram Acquisition

The acquisition of thermograms differs from the acquisition of visible light (VIS) images because thermal cameras have specific characteristics that need to be addressed. In this section we will discuss the two main acquisition-related topics in this thesis: the rolling shutter of the cameras and the radiometric calibration routine.

10.1.1 Rolling Shutter and Integration Time

The two infrared thermography (IRT) cameras, VarioCam HD and VarioCam hr, have the major disadvantage of rolling shutter technology. Each pixel has a fixed integration time of ~8 ms, but the rows of the sensor start their thermal radiation integration sequentially and not all at once (global shutter). In static cases or with slow moving objects this technique is not a problem. In our case of fast running people, we observe severe effects that reduce the overall and local appearance of the thermograms. In figure 10.1, the overlap in the right shoe is clearly visible. In addition, motion blur occurs, further blurring the image.



(a) The right shoe is superimposed on the left leg, which is still visible.

(b) The right shoe blur does not affect other parts of the body.

Fig. 10.1.: Two consecutive thermograms showing the rolling shutter effect. The effect occurs with a rolling shutter when the boot is faster than the rolling image series integration.

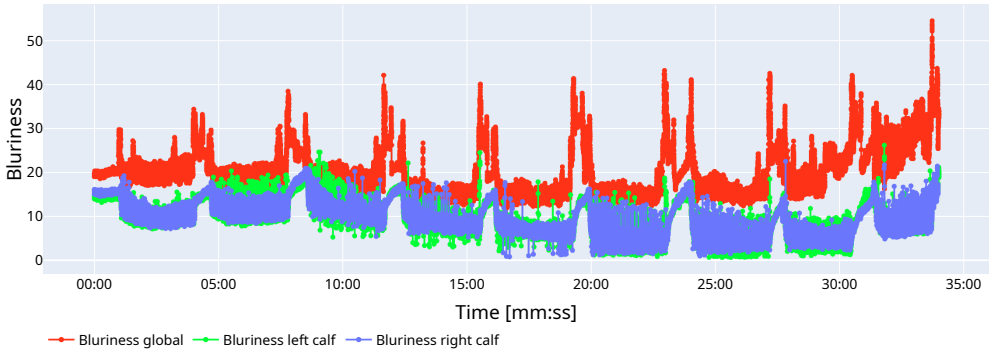
The rolling shutter effects are most visible in the shoes. In the flying phase (see figure 8.5) the impression is less sharp than in the standing phase. This may indicate the influence of the effect in combination with motion blur. Therefore, our algorithm aims to eliminate these images. The approach is described in section 6.3 and 8.2.5. A segmented image is analyzed for several consistency checks. One is that a leg does not have large convexity defects. These defects occur, for example, when a shoe overlaps the leg from the other side. This is also detected when the rolling shutter effect occurs, because the body part network (BPN) does not recognize the shape of the leg and the shoe well in this case. The detection is implicit, not explicit, and cannot be tracked directly. In addition, the time series processing filters the data based on the size of the ROI (either left or right). The assumption is that the largest ROI size is measurable in the standing phase of a step and should therefore be selected. The approach also introduces implicit definitions and does not explicitly analyze motion. To track and improve both implicit assumptions, an approach can directly analyze the step and identify the phases. With more accurate detection of the standing phase, the leg can be selected from nearly the same position within the cycle, leading to more comparable results in thermal statistics. Due to changing speeds and fixed frame rates, an exact match of the same position is not possible. Tracking the gait cycle has additional benefits. Further

analysis of the individual's running behavior can be performed. Running speed can also be estimated, eliminating the need for an external speed sensor or treadmill integration. However, the integration of gait tracking is highly application specific and requires further research and additional specific development resources to adapt available methods to the thermal image domain.

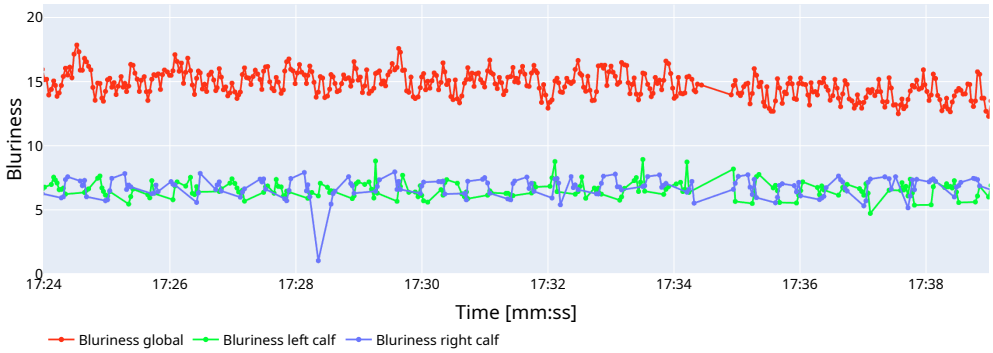
Another approach to mitigate the effects of rolling shutter and low frame rate is to improve the technical specifications of the hardware setup. According to [176], photon detection cameras are able to detect the amount of photons reaching the detector instead of measuring the heating effect of the radiation as in bolometer technology. This allows for higher frame rates as well as custom exposure times in the microsecond range. In addition, the technology allows the implementation of global shutter pixel arrays. A global shutter allows photon detection to start and stop simultaneously for all pixels. Therefore, there are no translation effects caused by pixels moving to other pixels due to different start times. Only exposure matters. However, current devices need to be cooled to low temperatures for operation, which has huge operational and therefore monetary costs. To understand the impact of the rolling shutter effect as well as fixed and high exposure times, future studies should compare the results with different camera types.

Since blur is an issue in thermogram analysis, there is a need for an objective way to further understand the sources of blur and apply methods to improve the results. Therefore, an indicator of the amount of blur in the whole image and in individual ROIs is needed. Pech-Pacheco et al. [127] described the variance of the image's Laplacian as an indicator (high variance is less blurred because more edges were detected, while low variance indicates fewer edges of the more blurred ROI). Figure 10.2 shows the blur results for an exemplary experiment. The red markers describe the overall image blur, which is not applicable for selective ROI analysis. The left and right legs have different blur states, so we take the detected ROIs and calculate a blur value for each. As the plots show, there is a high variation of blur over time, as expected for different phases of the legs, and in the phase shown there is an anti-cyclic behavior like the gait itself. The faster a person moves, the more blur is detected due to the fixed integration time. Since the blur is based on the edges, it is less precise in the early phases, as the inner parts of the ROIs do not have many vessel patterns and often look homogeneous. In later stages, this changes and many patterns are visible and blur detection is more robust. A static threshold cannot be estimated and a dynamic approach for filtering is

necessary. Indicator detection also depends on correct ROI detection, which can fail due to bad images. Blur is not yet integrated into the image selection pipeline, but further research should investigate robust detection of bad blur. In later stages, there is no clear anti-cyclic pattern, so a simple approach of using blur as a filter indicator is generally not possible.



(a) Entire experiment.



(b) Extract of 15 s during a walking phase.

Fig. 10.2.: The graphs show the thermogram blur indicators for the entire image, the left calf, and the right calf of a T0 experiment from the Incoreloop study.

10.1.2 Two-point Radiometric Calibration Target

The two-point calibration has been applied in many studies in this thesis and has improved the stability of the measured thermogram results. It is the first device to calibrate thermal images without knowledge of environmental conditions and object distance to the camera, but the proposed approach does not meet the requirements of full calibration to absolute temperature values.

ArUco Detection The ArUco detection in thermograms detects both markers if they are present in the thermogram. The image must be prepared with high contrast for the marker itself to be detectable by the ArUco algorithms. Therefore, the target temperature is assumed to have a fixed intensity and only pixel areas with values around the target are analyzed. In case of failure, the assumptions are changed and the marker is searched again. The approach is able to detect the markers in our case. However, if the markers are not visible or occluded, the detection fails and a full search is performed, resulting in a very long processing time per image. Thus, the conversion of a sequence can be delayed for a very long time, e.g. if a person accidentally walks through the field of view, as this will occlude the marker for several seconds. To overcome the hand-crafted initial marker detection, the integration of deep neural network (DNN) object detectors such as the YOLO detector family (e.g. YOLOv7 [177]) can be introduced, which presumably work on 16-bit images and have a robust and fast detection without exhaustive iterative search. The ArUco recognition works well and the device construction for the pattern is simple. However, different types of markers or objects have not been discussed. We do not integrate just two plain plates at different temperatures, as they may not be found accurately without manual interaction.

Temperature Calibration The thermistors that provide the temperatures to control the reference plates have a systematic offset error of $\pm$ 2%. The specification of the calibration device defines the systematic offset error as long-term stable in terms of years (Nägele, personal communication, Jan. 22, 2024). Multiple repeated measurements, such as in test-retest study designs, have the same underlying systematic error and are comparable. Therefore, relative surface radiation temperature (T_{sr}) changes are valid and repeatable, but absolute values are unknown, which is sufficient for the studies performed in this thesis.

Although we are only interested in the relative temperature change, it is advantageous to calibrate the measurements to absolute temperatures. The two targets are independently controlled to maintain their temperature and each has a different offset error. Therefore, the true temperature range between the targets is not known. For studies with the same hardware setup, relative comparisons are possible. However, comparing multiple devices in different laboratories would result in different offset errors between the two targets and thus a different temperature range. To obtain true temperature values, the thermistors must be calibrated. According to Bernhard [16, Chapter 7] the

calibration has to be done by comparing the sensor result with a predefined fixed point. Officially standardized are the ITS-90 fixed points (International Temperature Scale of 1990). The authors propose an experiment based on the melting point of water, which does not perfectly match a fixed point, but is practically applicable for calibration with an uncertainty of $\Delta T \leq 5$ mK, which is less than the stability of the PID controlled temperature of the purchased thermistors. In addition, for a highly accurate calibration, the thermistor should be calibrated at its target temperature, which may be more complex to implement.

Following the standard blackbody calibration performed by many researchers in the field of applied thermography on human skin [113], our system improves the accuracy and stability and therefore the overall reliability of the measured thermograms. In the standard procedure, the image acquisition of thermal cameras is also a source of error because the factors that affect the mapping of pixel intensity to temperature over time, such as transmittance, ambient temperature, and humidity, cannot be precisely determined. In addition, a blackbody has a systematic error as well as a stability of its radiated waves. In [58] we presented a study where we applied blackbody calibration by estimating the offset between the measured and defined temperature in each image. We used a reasonable blackbody with an accuracy of $\pm$ 0.5° C @ 100° C and a stability of $\pm$ 0.1° C @ 100° C. However, the target was set to 50° C to match the temperature scale of the camera. The accuracy and stability at this temperature is not given and may vary. Also, the emissivity ε is given by the value 0.95, which is lower than the commonly assumed human skin emissivity of 0.98 (dry and wet skin, see [@25, 34]).

In contrast to the established blackbody calibration, which uses a single reference point from a blackbody device (e.g. [113]), we developed a system with two reference points. To test our method, two experiments can be performed: first, the calibration system must be compared with a blackbody device, and second, with one area of the two-point device for offset correction. The gradients of calibrated images from both methods are compared to the two-point calibration to verify that the relative temperatures between pixels are similar. For similar gradients, a one-point calibration is sufficient, otherwise the two-point target is preferred. With a single reference point, it is not possible to convert pixel intensities to temperatures because the spread and terms of the thermal radiation formula (3.12) are not known. We just record the pixel intensities and cannot perform the comparisons.

Influencing Factors Emissivity describes how the object radiates thermal waves compared to a true blackbody. For true temperature measurements, it is necessary to find the emissivity for the material of the object. In addition to this important factor, transmittance defines how much thermal radiation passes through an object. It is important because the radiation from the object passes through a medium (air), but can be neglected for small distances. Many objects, including the human body, are opaque to thermal radiation. The reflectivity of an object is also important. To our knowledge, human skin has no significant reflectivity [34, 113]. According to radiometric calibration, all these parameters must be considered when building an external device. The device should behave like the target object to ensure correct error compensation. However, the custom two-point calibration target has a black aluminum plate with $\varepsilon = 0.97$, which is 0.01 less than human skin. Although human skin ε is standardized, there could be an influence of human hair or skin composition that could result in different coefficients. Since we are looking at relative changes in radiance over time, we also neglect the error of a small difference in emissivity.

The main thermal camera (VarioCam HD) employed in this work has a sensitivity (thermal resolution, noise equivalent temperature difference (NETD)) of 0.02 K. To overcome single pixel noise, we average many pixels with the same known radiation profile in our calibration marker to get a noiseless and more stable result. Since our calibration device has a stability of $\pm$ 0.01 K, we can assume that small changes in the thermal plate will not be detected or will have a small effect, while the noise effect of the camera will predominate.

Since the calibration device is placed next to the target body, a distance effect can be neglected. The average distance in all experiments is 1.8 m in the StereoThermoLegs study and similar in the other studies. To systematically test the thermal stability at different distances, the distances around the base of 2.1 m are tested. Together with the instantaneous field of view (IFOV) of 0.57 mrad, the spot size of a single pixel is 2.1 m $\cdot\, 0.00057 = 0.001197$ m $= 1.197$ mm. Thus, the captured size that is captured by a thermogram for a distance of 2.1 m, is 1.197 mm $\cdot\, 1024 = 1225.728$ mm wide and 1.197 mm $\cdot\, 768 = 919.926$ mm high. According to Plagenhoef et al. [133], the total leg size is about 90 cm, but the calves are less than 50 cm high. By reducing the distance between the participant and the camera, it would be possible to reduce the area covered by a pixel of the body. Both cameras can be rotated to capture images vertically instead of horizontally, and placed even closer to the participant with a similar field of view and higher local resolution.

However, the disadvantage is that the legs may be out of frame, especially in later phases when the person is running fast, including small flight phases during steps that are higher than the initial positions. In addition, for safety reasons, the camera must be placed with sufficient space behind the treadmill. A hanging solution directly behind the person would not be possible to avoid collisions in case when the persons falls.

The calibration area size in the thermogram is typically about 210 pixels for the upper marker and 180 pixels for the lower marker. The averaging is done on a warped version of this area, which is already an averaging of these pixels. The warped regions are about 4050 pixels in size. It needs to be further investigated whether the warping process affects the result by introducing a two-stage mean instead of a direct mean of the original image. Our experiments have shown that there is a slight difference. To prove the correctness of our approach, we need to use a comparison with a known calibrated temperature source. In the experiment with the warped

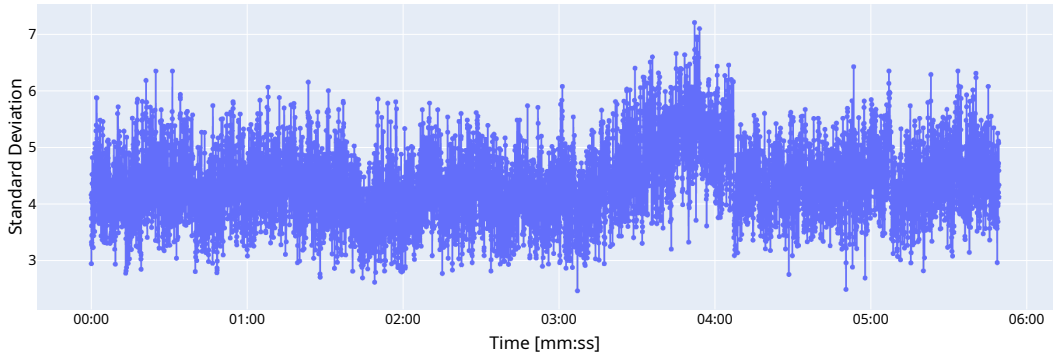


Fig. 10.3.: The standard deviations (SDs) of the lower marker pixel intensities in an example time series. The mean SD is 4.36.

and unwarped markers, the means over all samples are quite similar: the lower marker has the mean calibration value of 2468.47 (unwarped) and 2468.50 (warped). The standard deviations are 3.54 and 4.36 pixels. If the pixel intensity step is assumed to be 10 mK, then the expected SD from the camera specifications would be 40 mK ($\pm$ 20 mK thermal resolution). Both standard deviations are close to this point (about 4 pixels) and therefore valid. In addition, the SD for the upper marker are both a bit above 4 (4.20 and 4.78). For the lower distorted marker figure 10.3 shows an exemplary time series of the SDs. The values are noisy due to image noise. The peak in the second half could be related to changes in humidity or room temperature and therefore more control signals from the PID system. However, the environmental sensors and

access to the PID were not available. Nevertheless, the resulting images are still valid.

In addition to distance effects, we also tested the angular stability of the marker plate with several experiments. The angle of the surface normal to the optical axis should not exceed ~40° in any direction because of the change in emissivity of nonconductors (e.g. [113, 176]). We reproduced the limitation with our own experiment. Therefore, in our studies we place the marker as close to 0° as possible.

The resulting image is an 8-bit thermogram, where intensity 0 corresponds to the lower target temperature and 255 corresponds to the higher target temperature, with each intensity in between scaled linearly. Measuring the average intensity of the marker pixels in these images does not result in 0 or 255 because the original data was given in 16-bit dynamic and the ROI is around the calibration value. However, in the 8-bit image, the out-of-range values are also set to the range boundaries, resulting in a shifted average. However, this phenomenon does not change the measurements significantly because the range was set so that the target pixels (the legs) have an intensity within the range and not outside of it.

Radiometric image calibration employing the two-point device consisting of two ArUco markers has proven to be a valuable tool in applied human thermography. It has been shown to be practical in experimental setup, while reducing the need to acquire additional data about the test area, such as distance or emissivity. The calibration method overcomes the dynamic errors of the thermal camera, which adds a changing error to the measurement (camera drift). However, the quality of the calibration depends on the marker equipment and the reliability of the temperature control system employed. While the systematic error of the thermistors changes slowly (over years), different measurements can be compared with other measurements. Our primary interest is in the repeatability of relative temperature measurements rather than absolute values.

10.2 Annotated Datasets

For the supervised learning of the body part network (BPN) and the vessel network (VN), we annotated 870 thermograms of the backs of runners' legs. The annotation of the 870 images was performed by 6 people with different

backgrounds (medicine, sports science, computer science). The effect of sample images and clear instructions for annotators is investigated and discussed in [141], where the authors find an improvement in performance for domain experts annotating with clear instructions for biomedical image analysis. In particular, the annotation of blood vessels requires an understanding of skin formation and the vascular system. The biological knowledge helps to decide whether a spot belongs to a venous or a perforator pattern. The labeling process for both networks is highly dependent on the annotator, even if they all share the same knowledge. For the VN, it has not yet been defined what the correct size of a vessel is, when it begins to appear on the thermogram, or when it is fully dilated. It is debatable whether labels should more closely reflect the true shape of the underlying structures or just the parts that are clearly visible on the thermogram. The latter may miss connections, while the former may oversegment areas that have more thermoregulatory effects on non-vessel parts. There are other surface radiation patterns visible in the data that are not covered by the VN. These correlate with tendons, other areas such as the hollow of the knee (visible as a horizontal line in thermograms), or area patterns such as the shoulders. Tendon patterns are similar to long vein patterns and therefore counterintuitive for the network to learn. Therefore, they should be included in the network as a separate class. At the moment we do not distinguish between tendons and knee flexion. So they are labeled as non-vessel parts. Although non-vessel parts should be areas with less structure. Introducing these classes can improve the generalization performance of the network and thus increase the accuracy of other classes as well. This may also improve the safety of the other structure classes. The BPN also lacks clearly defined body parts. The separation of shoes, bare skin, and clothing is intuitive. The definition of calves, knees, and thighs is not as clear. The knee area separates the thigh and calf, but it is currently just a small band between the two at the back of the knee. Some annotators define it larger, others do not. In our work [58], we defined calves with an anatomically inspired geometric description. The company ThermoHuman has developed a thermogram analysis software for segmentation into over 100 body regions (whole body front and back) as schematically defined for the lower back in figure 10.4 [@21]. Again, the segmentation is based more on geometric decisions, as in many manual segmentation approaches, and not fully inherited from anatomy.

Annotation strategies to improve the ground truth for BPN and VN, such as involving multiple people annotate the same images and taking a consensus version as ground truth, have not been applied. Also, strategies such as

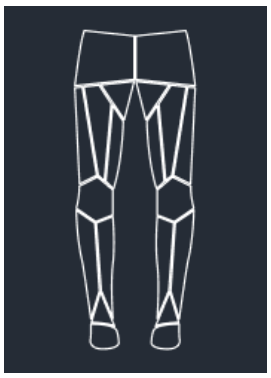


Fig. 10.4.: ROI definition for body parts by ThermoHuman of the human lower back. [@21]

active learning, where a human feedback loop is integrated into the training process, would improve the training results and the quality of the labels. Both require further development and integration of suitable tools. One may also consider the concept of simplified active learning, where the examples are predicted and the annotator optimizes the predictions afterward. The optimized predictions are integrated in a new training run. This approach can mitigate the impact of different annotators, as the network has already combined all strategies internally. However, well-defined classes are required for manual optimization of predictions.

The analysis of the dataset between the training, validation, and test sets shows different class distributions. For the training and validation sets, the class proportions should remain similar, which can be achieved by selecting new images for annotation accordingly. Distribution considerations should also be taken into account for the camera model and for the people's movements.

Thermograms are radiometrically calibrated by two methods: the manufacturer's method (VarioCam hr) and our two-point calibration device (VarioCam HD). The temperature scales are set to different ranges, which changes the visual appearance of the thermograms. The inclusion of multiple cameras and recording settings increases the data distribution. However, if the focus is on an optimized system with a single camera at a fixed resolution, the data should be more reflective of the camera employed. In the current setup, the VarioCam hr is no longer active in favor of the VarioCam HD. Some studies have applied it, but new studies integrate only the VarioCam HD. The dataset consists mostly of images from the VarioCam hr camera. Table 10.1 shows the imbalance of images from both cameras. Therefore, more labels must be generated with the VarioCam HD camera.

Camera / Dataset	VarioCam hr		VarioCam HD		Total
	abs.	rel.	abs.	rel.	abs.
Training	530	0.79	140	0.21	670
Test	140	0.7	60	0.3	200

Tab. 10.1.: Distribution of annotated images with VarioCam hr and VarioCam HD in training and test sets.

We analyze the thermograms not only in the standing phase with 0 km/h. Our approach first fully analyzes the thermograms and calculates the thermal statistics during motion. With the effects of motion blur and rolling shutter, as well as overlapping legs and shows, many situations are introduced that were not present in the standing phase. To handle the moving images, they must be included in the training dataset. In the manual dataset, the proportion of images captured in motion is not representative of a measurement. Assuming the protocol in figure 5.1 21 min (~$^2/_3$) is running and 10 min (~$^1/_3$) standing. However, this ratio is not reflected in the manually annotated dataset, as shown in table 10.2. Annotated images of standing people are more common in the dataset than in a typical experiment. Therefore, images need to be annotated for different speed phases, including different vessel pattern occurrences at each speed, to better match the typical distribution of running and standing images.

Dataset	Person standing		Person moving		Total
	abs.	rel.	abs.	rel.	abs.
Training	279	0.42	391	0.58	670
Test	73	0.365	127	0.635	200

Tab. 10.2.: Distribution of annotated images with people standing and people walking or running in training and test sets.

Finally, the effect of the people in the dataset must be considered. A major influence on skin radiation is body hair. There are no large hairs on the calves in the analyzed ROI, so we do not introduce extra labels or special examples. For skin pigmentation, Charlton et al. [34] indicate that differences have no effect on human skin emissivity. Age, sex, skin type, skin-to-fat ratio, diseases, and other factors can affect the thermoregulatory system. Differences are found within the time series data, but it is not necessary to explicitly label them. One factor is the presence and flow of sweat [136]. The fluid changes the thermal radiation at the affected parts, resulting in possible pattern detections. It is not yet clear how to handle sweat in the dataset. A potential solution is to

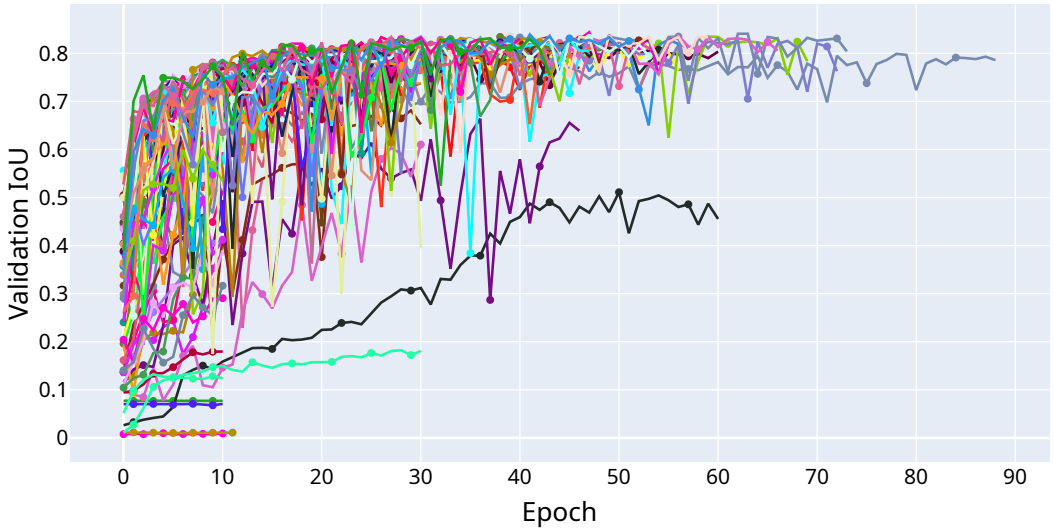
label the sweat areas in order to enable the networks to distinguish them and exclude this data from the thermal statistics.

K-Fold Cross-Validation For the training routine, the manually annotated data was divided into three subsets: training, validation, and test. The selected fold (3, see figure 9.5) has fewer participants in the validation set than other folds. One of them has the most annotated labels (80) compared to the other individuals (10-30). These labels are randomly selected from the whole experiment plus each standing phase. The person is part of the COMMED study with VarioCam hr thermograms only. Many images in the entire dataset are from the same study. Therefore, the data is represented by the selected validation set and maximizes the validation intersection over union (IoU). The cross-validation (CV) was obtained with training models for the VN for a fixed set of hyperparameters. However, further fine tuning would include the CV in the hyperparameter optimization routine. For BPN, the data split is the same. Again, separate optimization would lead to a specialized result, but with the need for more computational resources. Besides the 5-fold CV, other splits can be discussed, such as a leave-one-out CV where one participant is the validation set and all others are the training set. This technique is not applicable to our data distribution. Each person has at least 10 labeled thermograms, mostly from standing phases. Since we also analyze running phases, the validation set must include running images, which is not the case in the leave-one-out CV.

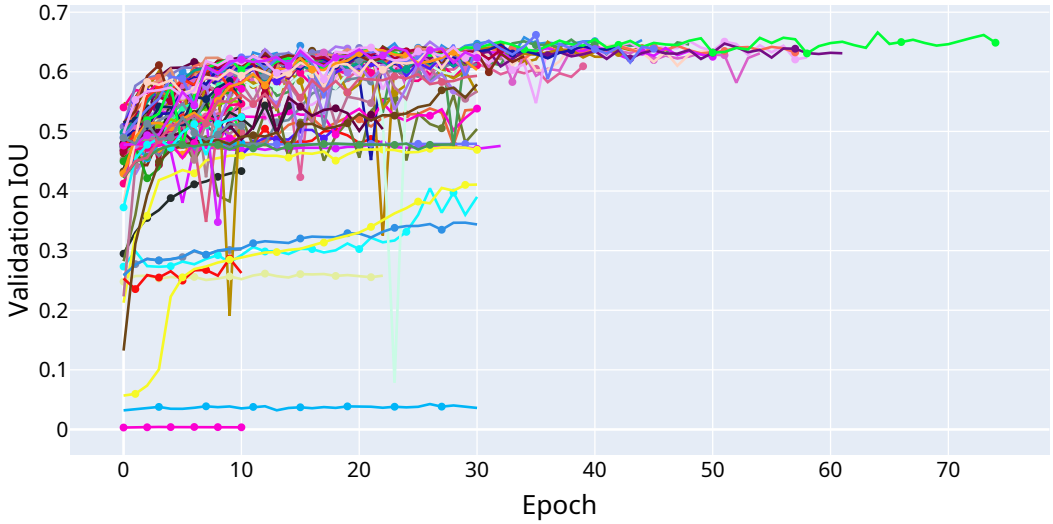
10.3 Thermogram Segmentation

Steps 2 and 3 of the ThermoNet processing pipeline introduce the BPN and VN as semantic segmentation networks for extracting the (left and right) calf, vein, perforator, and non-vessel ROIs. Many hyperparameter combinations show a similar IoU curve over training compared to the others in the corresponding task (see figure 10.5). Although there may be other hyperparameters not considered in this work, such as transformers, that will slightly increase the segmentation performance, the IoU cannot be increased much higher. The ground truth data has several drawbacks, such as several classes are labeled differently, which contradicts the high IoU results for unseen data. There were multiple annotators, all of which have different strategies for finding vessel patterns and also label them with different accuracy and size. This may also

be the case for multiple images labeled by the same annotator on different days. From the small amount of data, a generalization is achieved that does not perfectly reflect all of the ground truth data. Thus, many networks achieve similar performance and do not exceed an upper bound.



(a) Intermediate validation IoU values for the BPN hyperparameter optimization (HPO).



(b) Intermediate validation IoU values for the VN HPO.

Fig. 10.5.: The figures shows the validation IoU values for each trial of the BPN and the VN HPO. Each step represents a training epoch and the reported IoU on the validation data. The different trials are represented by an individual colored graph. The overview shows that many trials have similar learning behavior and do not exceed an upper bound in both tasks.

Loss Functions We have chosen a predefined loss method with the HPO. However, a loss function adapted to the segmentation problem can improve the specific results. The inclusion of the small tubular or snowflake/tree-like structures of veins and perforators can be handled by additional constraints as in the clDice loss. The number of connected components can also be considered in the loss function designs. A prerequisite for the training process is the differentiability of newly introduced terms. Differentiation is required for gradient calculation, which is the basis for weight updating. With the blob loss [88], the connected components are addressed indirectly by focusing on the class imbalances between small components compared to a large background area. This is the case for veins and perforators. The blob loss shown mitigate the effect of small class instances in the image.

BPN Post-Processing The calves segmented in the second step require post-processing for feature extraction. The reason is to remove false predictions and to avoid overlapping areas of both calves or other parts. The BPN predicts the ROIs with an IoU of 0.68 (table 9.3) and the calves with a much higher IoU. Nevertheless, the IoU does not indicate the number of predicted areas. In the body parts there is usually only one area for a class, only in rare cases there are two, but the predictions include multiple classes, including small speckles in other ROIs. The presented post-processing algorithm applies several consistency checks and filters misclassifications. The algorithm parameters were found manually and the filter thresholds may not fit well in all situations. The goal of applying an automated, data-driven segmentation approach is to eliminate manual interaction in the processing pipeline. Therefore, both the loss and performance metrics need to be revised to incorporate the constraints of the maximum number of regions. By limiting the number of connected components, non-pixel-level constructs must be analyzed. Hu et al. [67] presented a novel loss function to preserve the topology as in the ground truth data. Perret and Cousty [131] propose a topology loss based on component trees to address the problem in a differentiable way. Both topology preserving methods would regularize the loss function and help to limit the number of components to the number of ground truth components. Therefore, regularization can improve the misclassification. Nevertheless, the constraint of non-overlapping body parts persists. Consequently, overlapping classes are not distinguished through separate labeling. The introduction of a dedicated class within the dataset could eliminate the need for subsequent post-processing. Alternatively, the incorporation of the manually created

algorithm into the training process could facilitate the appropriate filtering of labels for the purpose of specialized training signals. The manual filtering focuses on valid calves. Therefore, the loss function and the evaluation metric should also only emphasize calf performance. The post-processing of the body ROIs further improves the inference speed of the entire pipeline, since only segmented calves are passed to the VN. If there is no calf, the VN is omitted to avoid unnecessary computation.

Augmentations We defined a fixed set of data augmentations to increase the variability of the training data without labeling more images. The set was estimated by manually designing the training procedure. He et al. [56] review more sophisticated augmentation designs such as label smoothing, sample mixing, or knowledge distillation. The authors note that many augmentations and improvements are available, but they do not necessarily improve semantic segmentation. Therefore, they must be additionally integrated into the search space for the HPO to find appropriate combinations. In addition, radiometric modifications can be integrated into our training process. Many of the manually annotated thermograms are radiometrically calibrated with the two-point calibration target. The original 16-bit information is available, but training is performed on 8-bit calibrated images. Instead of assuming a fixed target temperature range, the scale can be modified slightly. The resulting 8-bit representation changes while the label remains the same. The approach is similar to changing the brightness or contrast of the image, but involving intrinsic radiometric data. However, calibration values are not available for all images. The temperature scale variations should be in a similar range as the target images. The different scales simulate different environmental conditions, such as a warmer background (wall) or different thermal behavior of the object. In addition to the overall change in radiometric properties, another enhancement scheme may include only the target ROIs. Some people have a lower T_{sr} and appear darker in the thermograms than others. Modifying only the target parts also increases the variation in the thermograms and presumably matches more people.

Data normalization is based on mean and SD from a large (unlabeled) dataset representing the image distribution. For BPN, the thermograms were processed with the target temperature scale. But the VN processes only the calves and no background or other classes. Therefore, the mean and the SD are calculated on the masked image with body parts. However, the calculation was done with a hand-crafted segmentation algorithm that does not segment

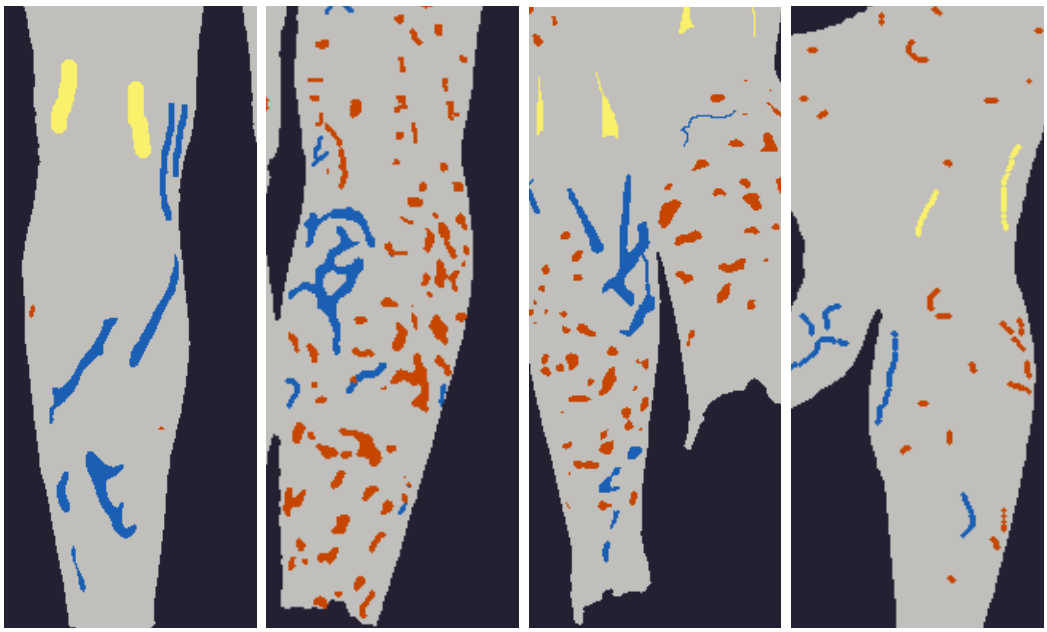
the images well. Nevertheless, the mean and the SD are still dominated by the legs and less disturbed by the surroundings. Further research needs to recalculate the mean and SD for more reliable data with the segmentation results from current BPN networks.

Transfer Learning In the search for a model architecture, we have already included a DeepLabv3+ version pre-trained on ImageNet. The principles of transfer learning, as reviewed by Iman et al. [74], should be explored in more detail in the context of this work and applied to the segmentation tasks. Therefore, pre-training should include pre-training on the automatically annotated dataset as presented in the StereoThermoLegs section. The improved performance when fine-tuned with manual data shows potential, although the labels are not perfectly aligned with the body parts. Other datasets, including existing ones, should be included in the pre-training process to allow the network to focus on general segmentation tasks first and fine-tune with custom data later. The teacher-student principle can be applied as an active learning strategy. A teacher network acts as a basic model that performs many tasks well. The result of the teacher model is optimized by annotators and used for the specific task to train a specialized model. This form of transfer learning is also known as knowledge distillation [64], where a specific piece of knowledge is extracted into a smaller representation, but performs either similarly or better than the general model. There are no exact segmentation datasets for blood vessel segmentation tasks. However, similar tasks exist in other image domains. Retina datasets such as DRIVE [160] or CHASE [50] provide a vessel mask for eyes. In addition, [139] extended DRIVE with vein and artery label differentiation. Angiography data also models vascular structures in both 2D and 3D. However, all datasets containing tubular data, such as blood vessels, have sharp focus and contours. The images are from different domains than thermography. A transfer learning approach may be appropriate to focus the detection weights on these structures before fine-tuning with task-specific data. However, vessel related patterns have a very different appearance, they do not have sharp edges and are not fully connected. There are fewer differences between arteries and veins in the existing datasets. The sharp images allow continuous labeling of long and interconnected tubular structures. However, in our thermographic vessel dataset, the vessel structures are not interconnected as the underlying real structure. In particular, the perforators in the thermal region often appear as small individual patterns rather than long interconnected lanes or as snowflake structures.

When it comes to data augmentation, label generation, pre-training, or knowledge distillation, one can also think of generative learning approaches. Segmentation is a descriptive task. However, by introducing generative networks such as generative adversarial networks (GANs) or auto-encoder approaches, complete datasets from studies can be included and learned in an unsupervised manner. One approach is discussed in [92]. This work investigates different approaches to style transfer of StereoThermoLegs data from the visible to the thermal domain. The goal is to generate labels in the visible domain and also generate a corresponding thermal image. These data can be included in prior training to provide a basis for fine tuning with manual data. The approach only covers body parts where vessel patterns cannot be detected unsupervised. Therefore, this work does not focus on generative methods.

Vessel Patterns Performance Vein and perforator patterns are an interesting and not often studied area of thermograms in medicine and sports science. The patterns occur during exercise and increase greatly after the end of exercise. In several studies this has been observed and analyzed as hyperthermal (hot) spots [115]. However, as mentioned above, these studies are highly manual and lack reproducibility and prior ROI definition. Often, features are considered to form a hot spot and statistics are taken from squared regions (e.g., 5×5 pixels [130]) around these hottest pixels. The definition of hot spots is not the same as our vessel-related definition, it does not distinguish between veins and perforators or other hot regions such as tendons. Our work applies a novel technique in this area by incorporating data-driven deep neural network segmentation. The proposed solution segments the patterns visually well. However, the IoU results for each class (see table 9.4) are very low. This is caused by the instance-aware class imbalance described in [88], where small instances of a class do not have a large impact on the overall IoU metric, while large areas (background and non-vessel) have a larger impact. The second reason for the low metric performance is the imprecise definition and application of pattern labels by different annotators and over multiple years. In some cases, vessels are labeled more like the ideal vessel structure under the skin, while in others they are only annotated if they are visible. Figure 10.6 shows four examples of different label styles, especially different label thicknesses. To improve the results, the labels need to be re-examined. An active learning approach to harmonize the labels should be considered. The indirect approach of multi-label averaging could be applied through the

trained network and the predictions are stored as new labels, which are then reviewed by the annotators. More labels from the latest camera and more situations will also improve performance. However, a high IoU metric is not expected unless it is easy for manual annotators to label images. The network will generalize the labels and produce different results.



(a) Long and thick tubular structures. **(b)** Thick annotation. **(c)** Mixed annotation thickness. **(d)** Thin annotation with same thickness.

Fig. 10.6.: Four different label masks with different types of annotations, including highly adapted shapes, long tubular structures, thick and thin labels. For example, (a) and (d) have different label thicknesses. (a) has adjusted vein labels, but thick tendons. d) has the same thickness for all labels and classes.

According to Farhadpour et al. [47], who compare different macro, micro and weighted aggregation methods of metrics and their impact on classification results, micro-averaging implicitly emphasizes rare classes in unbalanced datasets such as vein or perforators. We employ micro-averaging for IoU in [60]. Although the reported values may be higher, we switch to macro-averaging because we explicitly want to report the mean IoU for individual image instances. Nevertheless, further investigation of training improvements with different aggregation methods can be considered in future work.

Another influence on the IoU calculation that lowers the reported values is the absence of classes. If a class is not present in the label but is predicted, or vice versa, the corresponding class IoU is 0. For the small vessel patterns,

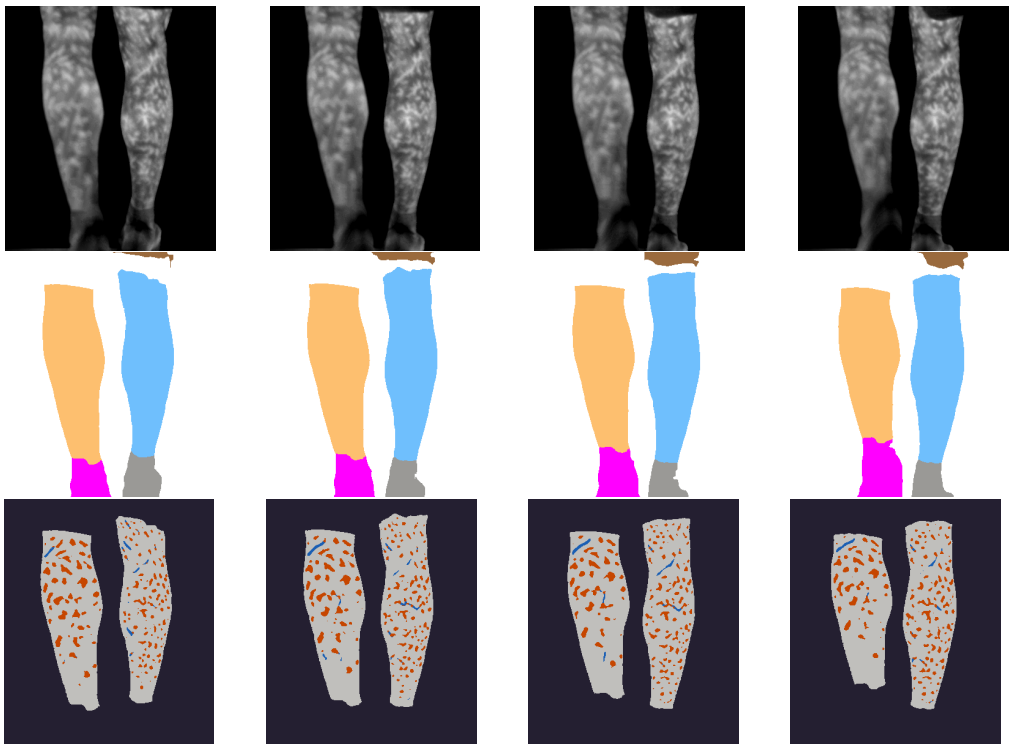
it is often the case that the network has introduced new locations that are missing in the label. Thus, the total IoU is lowered. But again, the network's predictions may reveal more information than the multiple annotators are able to label, since the predictions are more fine-grained than the coarse label shapes. In addition, the IoU definition is based on the intersection of the label and the prediction, as well as their union. Both regions take into account true positive, false negative and false positive predictions. True negatives are not considered. If a class is not present in the label and gets no prediction, then the true negative is correct. However, it is not reflected in IoU and the class still gets a bad score. In future work, it needs to be discussed whether there are better metrics to evaluate the performance of vessel patterns, e.g. introducing a comparison based on skeletonized shapes instead of complete shapes, similar to the definition of the loss function clDice (see section 6.2.4), or considering frame-to-frame performance.

Left or Right The first published approach [60] only included the segmentation of skin areas. There was no distinction between left and right side. However, the sides should be analyzed individually because in the periodic movement of the legs, each of them is in the opposite position and therefore have different thermal properties such as viewing angle. Therefore, the first approach cannot be considered further. The class definitions of the extended BPN consist of left and right classes for shoe, calf, knee, thigh, and clothing. However, a shoe looks similar from left and right perspectives, only the orientations are generally different. Instead of employing semantic classes for same occurrences but with different instances, an instance segmentation approach could be implemented. Instance segmentation allows multiple components of the same class, but each component is separate from the others. Thus, instances do not inherit the side directly, and therefore post-processing is required to find the sides later. The instance approach would allow the analysis pipeline to be extended to multi-person views. One use case for multi-person experiments is live analysis of runners in a race (e.g. [10]). In our limited case of a single person running on a treadmill, we assume that the positions of each leg side are correct (left on left side and right on right side). Without loss of generalization, the semantic segmentation approach fits the problem well.

Additional Training Process Improvements Learning rate scheduling such as cosine annealing with warm restarts promises faster convergence in the train-

ing process [157, 100, 158]. The adaptive optimizer methods employed in this work dynamically change the learning rate based on the optimizer methods, but still take advantage of the performance gains from global learning rate scheduling. Izmailov et al. [75] introduced a method to improve the generalization ability of the models in many cases, called stochastic weight averaging (SWA). The algorithm smoothes the gradient trajectory of the training loss surface, which improves performance on unseen test data and generalization. For the first implementation of the ThermoNet pipeline, learning rate scheduling and SWA were omitted to reduce the complexity of HPO.

Video Segmentation The presented approach for BPN and VN involves automatic segmentation of individual thermograms. While the filtering of the body areas already contains the positions of the ROIs from the previous segmentation, the BPN and the VN only work on individual images. Thermogram pixels are subject to various sources of noise that change with each image, which is expected. In addition, motion blur and rolling shutter effects degrade image quality. As a result, the segmentation varies from frame to frame and is not consistent over time. In figure 10.7, a sequence of four thermograms shows exemplary differences between consecutive thermograms in images without much motion. Predictions of body parts and vessels vary slightly between samples. This is a major limitation of the current work and should be addressed in future work. Changing the segmentation classes in consecutive frames has several effects on the thermal statistics and thus on the time series analysis, since different pixels are considered when extracting thermal features, which introduces an additional type of noise. With the basis of this work as a first processing pipeline, video analysis should address this issue and include frame-by-frame analysis. Zhou et al. [187] discuss in their survey different methods of video segmentation from several studies. An overview of different areas is given, including supervised, semi-supervised, and unsupervised methods. In this context, transformer-based methods, recurrent methods, or optical flow methods are mentioned; others may also be applicable. The approach requires a major architectural change in the training procedure and inference methods. As a result, the robust detection is still applicable to single thermograms, but also treats consecutive thermograms consistently. In the case of the BPN, this will reduce the need for the consistency post-processing described in this work, since important consistency checks are applied inherently. The technique is expected to require more computational resources, so the focus on efficient implementation of



(a) Frame 1. **(b)** Frame 2. **(c)** Frame 3. **(d)** Frame 4.

Fig. 10.7.: Sequence of thermograms with many perforator and different vein patterns. The first row shows the thermogram, the second the body parts, and the third the vessel predictions. Due to inconsistent frame detection, the body parts are different between thermograms (a) and (b) in the upper right calf. In the vessels, the appearance of the veins changes in all images, as do the perforators.

neural networks for training and inference should also be considered. Video segmentation can also be divided into long-term and short-term coherence. For short-term, specialized losses are applicable to image segmentation, e.g. with coherent loss [138]. Rebol and Knöbelreiter [143] implemented a convolutional long-short-term memory (LSTM) network together with temporal consistency regularization and achieved a frame-to-frame performance improvement. The challenge in temporal consistency work is to systematically train and evaluate the approaches due to the lack of labeled video data. The current manual dataset is labeled on single frames, and the thermograms were randomly selected from the studies. Furthermore, the annotators have not been trained for high temporal consistency. With the StereoThermoLegs dataset, the missing video data can be addressed, but only for the BPN data and not for the VN.

Thermal Statistics Many features can be extracted from the estimated ROIs, and thermal statistics can be constructed together with a time domain between successive thermograms. This work presents many features for all classes and some additional features for the vessel patterns. The several hundred features do not allow a simple overview of the thermal characteristics. However, not all features are as informative as others, while some are highly correlated with each other. In further studies, the relevance of each feature must be evaluated to select the most important data. However, the literature does not conclusively discuss which features to select and how [134]. The introduction of connected component analysis of vessel patterns allows a more detailed insight into the data, as not only global features are extracted. The number of individual structures and their size provide insight into the occurrence of a pattern. In step protocol tests, the number and size of perforator patterns increase rapidly after the end of the last stage in the recovery phase. Distinguishing between the total ROI size (in pixels) and the number of perforator instances can be informative. The appearance of multiple instances may indicate a different thermal body response compared to a few growing ROIs. The connected components analysis is also applied to filter out small instances that are less than 8 pixels in size. The small component filtering is done in the thermal feature extraction and has introduced another set of features. However, this filtering would better fit into the segmentation step instead of cluttering the thermal feature data tables. Connected component analysis increases the computational time for inference. Instead of applying semantic segmentation and retrieving connected components as individual instances, a deep neural network can be trained to segment the instances directly. In addition, instance segmentation would be able to distinguish components that are connected, whereas the algorithmic approach treats them as a single object. Although instance segmentation may be less performant than semantic segmentation.

Inference Speed There is a contrast between the camera's frame rate and the achieved computational performance. The camera captures 30 fps, while the average inference speed is only 4 fps. The low speed is due to unoptimized code. Although the networks themselves have been run in a faster inference mode via ONNX, the communication around them is still not optimized. The first step of radiometric calibration has an initial slow detection, which is drastically reduced in subsequent steps by remembering the marker position from the previous frame. The performance can be maintained if the first

marker position is saved and applied for subsequent calibrations without searching for a new marker. However, detection in each frame is necessary to ensure that the marker is always visible, e.g., that no person has blocked the view of the device, resulting in thermograms that cannot be calibrated. For BPN, the image is transferred to the graphics card. The result is then checked for consistency errors on the CPU, while the next step, the VN, is again on the GPU. Moving data to and from the GPU takes time and should be reduced as much as possible, e.g. by reusing the already copied image on the GPU for both BPN and VN instead of making a new copy. Batch processing in model activation is no longer possible due to the stateful consistency checks, which can be optimized by applying the video segmentation approach mentioned above. The performance of thermal feature extraction is also quite high, since it is computed on the CPU, and the decision of which features to analyze later is unclear. A production implementation will take these points into account and will also introduce further network optimization methods, such as quantization (reducing the precision of model weights or activations) or model pruning (reducing weights, neurons, or blocks of low importance).

10.4 Stereo Transformation for Label Generation

With the StereoThermoLegs dataset, we take a different approach to annotating data for the BPN. The applied transformation consists of five steps as described in figure 1.2. The generated dataset is evaluated in a benchmark and compared to the manual dataset. The benchmark shows that the combination of training on the new dataset and fine-tuning on high-quality manually annotated data leads to an overall performance improvement. In addition, fine-tuning with only 50% of the manual data still leads to better results than the network without pre-training.

Stereo Setup The first step is to build an appropriate stereo system. It is important to synchronize the image acquisition of all cameras involved. Since we already discussed how the 8 ms rolling shutter of the thermal camera degrades image quality at 30 fps, the same quality issues occur at the available 15 fps. Even though the frame rate is halved, the IRT camera still measures with an 8 ms shutter. There is no change in the acquisition time, which is suitable for our application, because the automatically labeled thermograms still have the same visual impression as the thermograms acquired in the live

system at 30 fps. The difference is that only half of the images are captured, resulting in less data consumption and less similar images. For the trainable dataset, only 10% of the images are randomly selected to not include highly redundant data from the cyclic run phase. Thus, halving the data acquisition rate is not a limitation of the system for label generation, except for measurements that are analyzed for other reasons, such as medical studies. The available thermograms may have missed the best leg positions for segmentation. Replacement with another color+depth (RGBD) camera should be considered to allow direct RGBD to IRT image transfer and evaluation as described in [144].

Calibration The stereo calibration routine is based on well known and proven principles. The different fields of view result in different sizes of the calibration board in the common visible area. It is necessary to increase the RGBD image size to find the calibration pattern. The low local resolution of the camera at the target distance may result in less accurate calibrations. The calibration is based on the centers found in the circle grid pattern. For the two spectra of the cameras, a special pattern has been constructed to be visible in all domains: visible with black and white and cold metal and warm foam for the thermal domain. The calibration images must be taken quickly to ensure a high contrast of the materials in the thermal domain. Further investigations may find a more suitable material combination with high thermal differences at room temperature without pre-cooling as suggested in [101]. However, it is not necessary for the application of this work as cooling or heating is available. The size of the pattern in the images can also be increased by enlarging the pattern circles within the board. However, results have shown that at the current size, circles with a diameter of 2 cm are suitable for achieving low reprojection errors.

Manual Extrinsic Correction The stereo system is mounted on a frame and placed on a tripod. Both cameras are placed vertically and fixed to the frame with a single screw each to form the fixed stereo rig. Since we found that the rig is not as fixed as it should be, the relative pose between the cameras can change minimally when an external force is applied. If the stereo system is not fixed in the same way as it was calibrated, then the calibration is no longer valid because the relative position and rotation ($[R|t]$) change. Current fixation is based on the $1/4''$ camera mount, but both camera cases provide multipoint fixation. With a multipoint mount on the frame, the cameras are more static

than with a single mount, thus preventing relative pose changes. However, the multipoint approach was not available during the StereoThermoLegs study. A small relative pose corruption error was detected, but the stereo system could not be recalibrated. Therefore, the extrinsics had to be reconstructed from previously acquired thermograms. Due to the lack of automatic feature matching in the two different image domains, manual point correspondences were found, which were not as accurate as required. Although the epipolar lines of the sample images were improved, the complete images of the study were not. The offset between the transformed image and the original image increased. The worse result is caused by the low quality correspondences due to fewer features, especially in the thermal domain. The different image sizes of both domains lead to inaccurate locations of found corners, and the available thermal features are fuzzy due to the blurred contours of the objects. As a result, only a few areas with promising matching points were found, mostly in the outer areas of the images and not well distributed in the image. Nevertheless, the application of the original extrinsics has achieved a smaller offset than the recovered extrinsics, and the post-processing methods reduce the effect of the misalignment. Therefore, the proposed method is not affected by the calibration error, but an accurate calibration would further improve the results.

Label Generation in RGBD The class definition problem (unclear body regions and vessel definitions) is not solved by the stereo transformation approach. It is still present, but now shifted to the RGBD domain. In the images, skeleton pose estimation can be applied without additional development. An additional skin detector allows the creation of a segmentation mask that matches our label definition. Much more research has been done in the visible light spectrum for human body analysis and can be applied with less effort. Foundation models for segmentation such as Segment Anything [85] allow for quick and easy propagation of labels. However, foundation models provide a task-unspecific solution and need to be adapted for the specific use case. In our presented approach, we do not consider these large and basic models because there are already specific models for our task. Nevertheless, further research in label formation should be applied to improve the overall performance. In particular, propagating the approach to new ROI foundation models will be the first choice for initial label generation. Although our approach is suitable for a proof of concept, the label generation will improve again when the body regions are well defined as described above.

RGBD to IRT Conversion The stereo approach is based on transforming texture information from the visual image domain of known depth to the thermal domain. The process relies on accurate depth information. The time-of-flight (ToF) camera provides real-time depth values. However, they are not as accurate as needed. Many pixels are missing or incorrectly detected. Therefore, a smoothing algorithm improves the depth mask. Stereo transformation without smoothing results in more artifacts due to incorrect depth or missing data. However, even the smoothed mask is not perfect, there are still larger areas where the depth mask had larger areas of bad results. To improve the overall quality of the transformation, the first step is to improve the depth mask. The Azure Kinect camera provides a pixel-wise match between the visual and depth masks, although the two cameras are not perfectly aligned and also form a stereo system. The vendor alignment is not evaluated in this work. ToF and VIS images have different resolutions and fields of view, calibrating and integrating a global stereo system eliminates hidden states and unknown calibration routines. It is not verified whether the depth and visual image match perfectly. The Azure Kinect also has built-in smoothing and optimization techniques that can improve the quality of the depth map provided; these should also be considered for performance improvements.

The systematic lack of information near one side of the legs is inherent to the stereo approach. The two cameras have different optical centers and different fields of view. Object points are seen from different angles and mapped onto different numbers of pixels. The objects are not flat, and with the different viewing angle, one camera is not able to receive the same information as the other. Therefore, the transformation process cannot recover all the pixels in the destination area from the source area. This phenomenon occurs only on one side. Filling in this unknown data is part of the post-process. Several synchronized RGBD cameras around the IRT camera can improve the overall depth map. Multiple views from different angles provide more accurate information than imputing information through post-processing steps. The combination also reduces artifacts in the depth map. However, the integration of multiple cameras increases the technical overhead. The proposed transformation system does not have perfect transformation properties due to different viewing angles, but the errors can be minimized by post-processing, which is cheaper than integrating more cameras.

Label Refinement in IRT The post-processing step removes minor transformation artifacts and prepares the labels for the watershed algorithm application.

Since the flood-filling algorithm takes the transformed labels as true values, they are not modified, so additional steps have been introduced, such as reducing label sizes to ensure non-overlapping class regions and non-wrong label associations. However, the post-processing steps are manually optimized for the current task of leg segmentation. As with the label generation step, the refinement step can be improved by introducing the application of foundation models for segmentation in the IRT domain. A combination of transformed labels and segmentation masks would provide a suitable initial mask for the watershed algorithm. However, applying additional foundation models increases the processing time for each image.

For pre- and post-processing, an additional frame-to-frame approach can also improve the consistency and thus the training results. In the approach presented here, labels are estimated from a single image pair of RGBD and IRT data. Image consistency is not enforced and consistency checks are not provided. An introduction of prior information in both steps is necessary to ensure a consistent label result, as discussed above for video segmentation. Providing a consistent video dataset is a crucial part for supervised training of video segmentation models. The stereo transformation approach is a viable method for time-consistent label generation if the pre- and post-processing steps are conducted in a manner that accounts for frame consistency. The current proposed dataset was randomly selected from the entire dataset, in video segmentation all images would have been affected.

Dataset The StereoThermoLegs dataset first proposes a dataset of healthy participants standing, walking and running on a treadmill. The publication of the dataset [6], including automatically generated labels from corresponding visual images and thermograms, allows researchers to perform analysis on moving targets in running exercise protocols with thermograms. This is a major improvement over current research methods that do not involve moving people. The BPN is integrated to analyze the images with a deep neural network. However, supervised training of such a model requires a huge annotated dataset. Since the cyclic step introduces a huge amount of redundancy, only 10% of the images were randomly selected from the entire study. To improve the dataset distribution, additional task-specific constraints can be considered, such as combining body part classes and vessel classes to include similar amounts of both major vessel types for each body part at different step phases. In the BPN, a constraint prevents the predicted calves from intersecting with each other or with other ROIs. However, even in these

cases, the current models perform worse when image acquisition effects (such as rolling shutter or motion blur) occur. The inclusion of such patterns makes the proposed dataset more balanced. Overall, the dataset is a novel work that for the first time allows other research groups to integrate high-quality thermograms with corresponding labels and visual images. Nevertheless, the dataset formation needs to be revised to address inherent problems of the data such as overlapping parts and others. Generating new datasets for other ROIs will require reconsidering the constraints and applying appropriate methods to meet the requirements. Finally, a quality metric must be established to evaluate the performance of the dataset itself. A representative set of thermograms must be manually labeled and compared to the automatic labels, e.g. with the IoU. However, in this work, quality assessment was not possible due to the expensive and time-consuming annotation process.

Deep Neural Network Benchmark The stereo dataset is a valuable starter dataset for pre-training and fine-tuning a model with a small amount of high-quality, manually annotated data. We have shown that the fine-tuning approach still achieves similar performance when fine-tuned with only half of the manual data instead of all of it. With only 10% fine-tuning, there was no improvement over the network with only the manual data. Further investigation to estimate the minimum amount of manual data is needed in future work. Nevertheless, the application of the presented approach will enhance future BPN development with new ROIs. In addition, improving the stereo dataset either with better initial labels or refining the resulting labels to better match the manual annotations will further improve the result. Therefore, the label definition and collection need to be further refined. The general training procedure has been proven on larger datasets and is also applicable to transfer learning. By introducing the improvements to the training procedure described above, this application will also benefit. In addition, HPO would find a best fitting hyperparameter set for the new dataset, which is not necessarily the same as for the BPN trained with manual data. For simplicity, we have skipped this part. A direct comparison of the stereo-trained network (without fine tuning) with the manually trained network shows a worse performance for the stereo-trained network. The reason for this is that the manual test set is more diverse because it contains images from multiple cameras and a higher variation of people performing tests, rather than a single camera capturing stereo images in a single study. The stereo-trained networks and the manual network were also compared for

their performance in assessing thermal features in time series analysis. It was shown that the stereo approaches are similar to the labels, which is expected. However, the ROIs extracted from the manual approach shows slightly higher T_{sr}, but also adds more noise and is less accurate to the labels in the later stages. Fine-tuned networks still preserve the curves of the labels over time, but also have a higher T_{sr}. Thus, both advantages are combined to produce a reliable prediction at high speeds.

Application to other ROIs The proposed approach has been suggested as a possible way to collect a baseline dataset to train an initial model. Therefore, the pre- and post-processing steps are heavily focused on the runner's back view, new methods with new classes need to be established to apply the approach to new ROIs.

The stereo system is able to generate labels for the BPN. However, the most challenging part is to find labels for the VN. This is not possible with the proposed approach. Labels are generated in the visual domain and transformed to the thermal domain. However, in the visual domain, the vessel patterns are not visible and therefore cannot be detected with common models. For the VN labels, a completely different approach has to be found to automate the annotation process or at least to further support manual annotation.

Tempski [165] analyzed the transformation capabilities of the approach for thermal face recognition in a bachelor thesis. Therefore, a custom dataset (StereoThermoFace) of eight people cycling on an ergometer while capturing their face was performed with the hardware system presented in this work. The cameras are oriented horizontally to the face of the cyclist, who was asked to look into the camera while pedaling. The protocol starts with a rest of 30 s and ends with a rest of 3 min. In between, the exercise phase starts at 40 W and increases by 40 W every two minutes until the participant is exhausted. If the rate of perceived exertion (RPE) was above 17, or at the user's request, the trial will end. However, the protocol was adapted dynamically to allow the participant more time to move and to obtain more pronounced thermoregulatory responses. Due to the small number of participants, we choose only one person as a test set, while the others are used for training and validation. A total of 111,558 images were acquired. Tempski applied different face recognition methods to the RGBD images and transformed facial landmarks to the thermal domain by the concepts explained in this work. In the thermal domain, face detectors already exist, but they do not perform well

on faces after external physical stress (exercise test). Therefore, new models have been trained and compared. The approach differs in pre- and post-processing as the labels are points instead of a dense mask. The study shows that the system can be applied to new ROIs with less effort than manually labeling a whole new dataset.

The stereo system approach can also be applied to generative models. In the bachelor thesis of Lang [92], the corresponding image pairs are applied to evaluate different models with the goal of generating or transforming a given VIS image into a corresponding thermal image without knowing the real thermogram. The work operates on the StereoThermoLegs data. The work considers unsupervised learning techniques for style transfer from the visual to the thermal domain. With this approach, label generation and transformation should be approached in a different way. Although the training data consists of stereo pairs, the application aims to generate thermal images and corresponding data involving only visual images. Lang compared several recent works for thermal image generation. However, a benchmark for the performance of BPN or VN is still needed.

10.5 Time Series Analysis and Sensor Fusion

The final processing step in the ThermoNet pipeline consists of time series analysis and sensor fusion. In (sports) medicine, not only thermography is applied for measurements, also other sensors for heart rate and breath-by-breath analysis, as well as log data like treadmill speed or environmental data. Sensor fusion allows combined analysis of different multi-sensor characteristics.

The basic time system is the microsecond timestamp. Microsecond precision is machine dependent and not reliable across devices. Since the frequency of all sensors is below the IRT frequency, we find the closest time in the sensor data for each IRT image and take those values. Because of the large frequency differences, we assume that there is no loss of sensor data. However, this is not guaranteed. In addition, when the IRT camera undertakes a nonuniform calibration (NUC) no images are taken and therefore the sensor data in this time range is lost. However, the physiological adaptation occurs over many seconds and is consistent afterward, so a small loss of data does not limit the insights of the data over time. For statistical analysis, the time series

data often needs to have the same amount of data for different features, so including data with a missing part would not be beneficial for these methods. The assumptions we make are applied to the entire IRT data. However, prior to data merging, the images are already filtered to include only thermograms with suitable information from the legs (straight and not occluded). The inclusion of IRT data depends on the step cycle speed and the step execution (more or less overlap depends on how the runner moves his legs). Thus, the IRT data has no fixed frequency and is lower than the 30 fps captured by the camera. A naive approach to overcome the problem of missing IRT time points is to add virtual time points to match the original frame rate. In statistical analysis, it is necessary to decide whether to include these data or not. Another approach, which requires further investigation, is to resample the frequencies to a common one. The IRT data should be sampled with at least twice the frequency of the fastest step cycle to properly include both steps. The step frequency is up to ~10 steps/s, but usually lower. Thus, the filtered IRT data is still more present than other sensors. However, stabilizing the frequency improves the analysis strategy. IRT filtering is performed by peak estimation of the sizes of the detected ROIs. As shown in the results (figure 9.20), not all peaks are found in the data due to different step frequencies and segmentation artifacts. The filtering process can be improved by considering step frequency to evaluate valid data based on a modeled motion. However, if there is no suitable ROI (e.g., due to overlapping legs), there is still no data. Future research is needed to determine whether imputation methods could adequately fill the gap.

Physiological data is protocol dependent. An external sensor was introduced to measure speed, but it is unreliable and requires several steps of denoising and correction in a hand-crafted algorithm to correctly recover the exercise protocol. Nevertheless, a more robust approach is to incorporate protocol definition and control into a connected system instead of manually controlling the treadmill speed. With BlueCherry this is already possible and has been applied in bicycle ergometer studies. An alternative and treadmill independent approach introduces gait tracking and speed estimation from the steps of the runner. This would reduce dependencies and is also applicable to IRT-only measurements. Speed estimation from visual images needs to be transferred to the thermal domain.

Thermal statistics and sensor fusion introduce a large number of features at a single time point. In particular, the different ROIs for each side of the body in the thermograms and the large number of different features are difficult to interpret as a human. The current focus has been on basic features such

as mean T_{sr}, size in pixels, or number of individual patterns. However, a systematic approach to select the most informative features is still missing. Further studies should introduce automatic correlations and comparisons to reduce the number of features.

Some data features are noisy due to several effects. One type of feature noise is introduced by incorrect ROI detection, resulting in false thermal features. A small amount of noise is also introduced by camera noise and other acquisition effects. Sweat, hair, and environmental changes also introduce uncertainty into thermal feature extraction. Addressing the false predictions for ROI segmentation must be addressed in the design and training of the machine learning model, as discussed above. Camera noise is inherent in the data, and smoothing can help eliminate it to some extent without changing the true value. It depends on the camera model and the acquisition conditions. In our work, a smoothing algorithm is included. However, the window length is fixed at 151 data points. At full frame rate, 5 s of the measurement are considered. In the filtered data, a larger time span is considered. To minimize the effects between stages and different velocities, smoothing is applied only within a stage. As an alternative to window length limited smoothers, Eilers [46] presents the Whittaker-Eilers smoother, which has advantages over Savitzky-Golay in that it does not consider a fixed window length and performs better at data boundaries. However, more research is needed on the optimal smoothing method with the trade-off between smoothing the curve and preserving real data changes.

A reduced version of the full experiment analysis is to analyze only a single point per stage. This approach mimics the manual analysis strategy and is useful for comparisons with other research groups performing manual analysis. Per-stage analysis also facilitates comparisons between individuals with many participants. However, it remains to be shown whether the full data or the reduced data is necessary to gain insight into the physiological data. The premise of this work is to collect all data to broaden the data base rather than relying on a few data points as in manual analysis. The step walking protocol includes a pause after each stage where the participant stands. To evaluate the data for a stage, the last 30 data points of a stage are averaged. A representative and comparable thermogram is selected from the next standing phase. This strategy provides a good insight into the characteristics of the phase, as the body state has been adapted to the phase speed in the last period of a phase. The T1-T3 protocols of the Incoreloop study are different. There is no pause between the stages. Therefore, no standing image can be

extracted. However, the thermal statistics of the last data points of a stage are still available. Although the thermograms are selected in the standing phase, there may still be invalid images. Since the predictions could fail (even in the standing phase), a thermogram may be selected that is not well segmented. If the segmentation cannot be improved, the standing image selection must check for conditions to select an image with both legs detected.

In [60] we compared deep neural network segmentation with a manual thresholding approach. Since the manual approach is not capable of analyzing thousands of thermograms, the comparison was made on the manually selected images of 12 participants from the three studies. The manual images are selected during the standing phase of the incremental exercise test between phases and analyzed with both the manual and the automatic skin-only BPN strategies (see table 9.2). The analysis methods were statistically compared and showed comparable results in ROI size and thus T_{sr}. The presented approach was not yet fully capable of the features proposed in this work. In particular, it lacks the ability to discriminate between the left and right side and between different parts of the leg. Nevertheless, we have demonstrated the superior analysis pipeline of our automatic approach over a manual one, due to the faster and more repeatable evaluation and the introduction of the full analysis of all captured thermograms.

Overall, time series analysis involves several steps, including merging, selecting, interpolating, and smoothing data. Each step has advantages and disadvantages. Many hyperparameters and hand-crafted algorithm design decisions must be made to obtain a final consolidated and merged data table. Finally, a clean data table from the ThermoNet pipeline is provided for further analysis and interpretation by domain experts in sports science and medicine.

10.6 Exercise Physiology Application

The ThermoNet processing pipeline aims to provide a reliable data source for the concepts of exercise radiomics. The non-invasive nature of applied thermography allows the analysis of individuals without interfering with their behavior. The treadmill exercise testing for physical capacity assessment, previously performed with breath-by-breath analysis (spiroergometry), heart rate assessment, and other data, is expanded to include the continuous

thermal surface radiation of the calves. The huge amount of data allows for more detailed investigations. In [59] we applied the pipeline for running and cycling. Comparisons between oxygen intake VO_2 and surface radiation temperature (T_{sr}) (see section 2.4) as well as between heart rate and T_{sr} show the relationship of calves thermograms to the gold standard metrics. Interestingly, running and cycling T_{sr} relate differently to maximum heart rate and $VO_{2,peak}$, allowing an understanding of the underlying neural activation associated with thermoregulation. The study demonstrated that T_{sr} provides meaningful insights to acute exercise response. If only the calves are considered, the promising separation between vein and perforator areas left behind. Nevertheless, further targeted studies need to consider these structures for a deeper understanding of the acute response of their thermoregulatory system and to infer physical adaptations. With the specialized dashboard, we provide a convenient starting point for new studies to assess experimental data for an individual participant. Current studies have not yet analyzed a test-retest reliability of thermal responses with exercises and over multiple experiments to assess long-term adaptations. In the Incoreloop study the comparison will be applied with the new pipeline for the protocols T1–T3. However, new confirmatory studies with larger sample sizes to validate the physical capacity are necessary to validate correlations between current measurements and our new system.

The ThermoNet pipeline is adaptable to new study designs with new background environments, including outdoor, different hardware setup, new body ROIs, and different movements. The scenario and hardware require extensions to the dataset, but the processing remains largely similar. Adaptation to new body parts has already been extensively discussed with automatic label generation. In addition, the BPN is already capable of direct transfer to new views, such as the cycling analysis in [59] with remarkable results. The most challenging part is the introduction of new movements that are complex and compound to study physical adaptation during sports, work and other activities. The analysis must be robust even when the activity does not involve continuous visibility in the camera's field of view (FOV), such as for cyclists crossing the FOV. In particular, the vessel analysis of these movements needs to be investigated for the significance to the physiological constitution.

In the following, the application of the segmentation methods to new, unseen cycling data is discussed in more detail. In addition, the results of a simultaneous IRT analysis with three cameras of different body parts of a cyclist are evaluated. However, the ThermoNet pipeline was not yet applicable and a

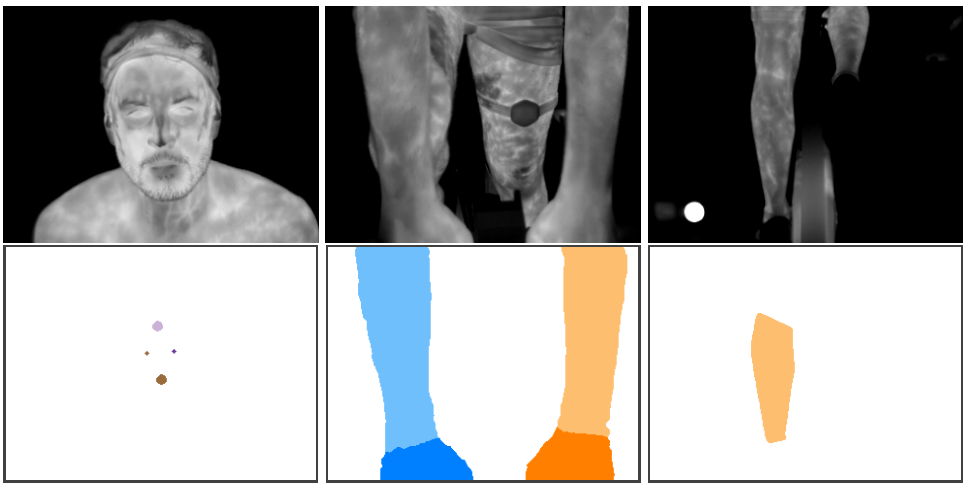
semi-automatic analysis approach is presented. Finally, a detailed analysis of the dashboard results is given.

Running and Cycling Differences Although the ThermoNet pipeline is optimized and evaluated on the calves of runners on a treadmill, it is already applicable to the same view of cyclists on a cycle ergometer. Figure 10.8 provides two examples of a cycling experiment with good and bad calves segmentation. It was not analyzed where the calf detection problem occurs in BPN or in post-processing. In [59] ten participants perform both a running and a cycling exercise test. The developed solution segmented the calves into left and right side. The processing pipeline only considered running in the training, but performed well on the cycling thermograms. The preliminary visual analysis indicates that the results are promising. A systematic evaluation is not yet possible because no labeled dataset has been created, neither manually nor automatically. In both evaluations (running and cycling), the calf was found to have a comparable ROI due to the reduced influence of body hair, less occlusion with other body parts, similar size in the images and ease of assessment. The study reveals intraindividual differences in T_{sr} development between the two modalities, which may be related to the different training protocols. The running protocol includes a standing phase while cycling does not.

(a) Thermogram. **(b)** Filtered BPN prediction. **(c)** Thermogram. **(d)** Filtered BPN prediction. **(e)** VN prediction.

Fig. 10.8.: BPN and VN sometimes perform poorly (a–b, no VN mask because no calf is detected) and sometimes perform well (c–e) in cycling segmentation, even though the networks were not trained on cycling situations. The example is taken from a cycling experiment.

Multiple ROIs Comparison In a separate study [58] we investigated together with the Research Group in Sports Biomechanics (GIBD), Department of Physical Education and Sports, University of Valencia, Spain, the effect of exercise



(a) Face. **(b)** Forearms. **(c)** Left calf with blackbody.

Fig. 10.9.: Example images from the study [58]. Face, forearms and calves (only the left is shown) are acquired simultaneously. Segmentation masks are manually obtained with a hand-crafted optimization algorithm. In (c), the blackbody is visible as a white circle in the lower left corner.

on thermal skin responses in different ROIs. Three identical cameras simultaneously record the calves, forearms and face of a participant as shown in figure 10.9. The cameras were not capable of continuous recording, so we only take pictures at the end of each phase. The participant cycles on a bicycle ergometer and must alternately extend and hold the legs for the duration of the recording before continuing to pedal. The load protocol had a pyramidal shape from low to medium to high and back to medium and low. The study was conducted with different cameras and environmental settings than in this work. Therefore, the image distribution is different and our methods could not be applied to calf detection. Bootstrapping a segmentation network for all ROIs was not possible due to the lack of a stereo system. Therefore, an automatic-assisted manual ROI extraction for the body parts was implemented. Radiometric calibration was performed at least by blackbody shift correction for the images containing the calves. A blackbody device was visible in each image and the difference between measured and target temperature was used to correct the temperature values. The calibration method is commonly applied, but in our case it does not provide a calibration for all three cameras because two of them do not capture the same reference object. In contrast to the developed two-point calibration device, the single-layer calibration in combination with the temperature values reported by the cameras is a different approach than ours. A comparison between the

two systems needs to be made for accuracy and validity. The study shows that different ROIs T_{sr} have different correlations with external load and internal thermoregulation. The calves T_{sr} are inversely related to load and RPE (temperature decreases when load and RPE increase and increases again when load and RPE decrease). In contrast, the inner canthus of the eye and the forehead are more related to body core temperature and exercise duration than to the currently applied load. In summary, the ROIs have a different time evolution in their T_{sr} and thus relate to different physiological and thermoregulatory processes of the body.

Dashboard Insights Presenting a complete analysis in a dashboard (see figure 9.22) provides several insights into the individual's thermoregulatory and physical responses compared to gold standard methods. However, the results have not been fully analyzed and more experience is needed to provide statistical information regarding reliable thermography results. In the example dashboard, the first field (a) shows the relationship between the increasing heart rate in the step walking protocol and the decreasing T_{sr} of a calf. The effect of a pause on both legs is also clearly visible. It is not yet confirmed to what extent the temperature difference between the standing and walking phases is related to the blur (motion and rolling shutter) and to what extent the effect of convective cooling by wind is included in the decrease. However, the maximum T_{sr} in the resting phase also decreases, which is consistent with other studies that only take single images in the standing phase. Thus, the convective effect does not limit the possible insights. The representation of T_{sr} is not directly related to training zones and lactate or ventilatory thresholds (VTs). The second field (b) compares T_{sr} with the breath-by-breath analysis, which also shows an anti-proportional behavior, i.e. respiratory frequency increases while T_{sr} decreases. The third field (c) provides an easy assessment of the maximum thermal vessel patterns throughout the measurement and allows the examiner to easily detect asymmetries manually. The fourth field (d) contains two different methods of core temperature and the relationship between the detected perforator temperature P_{sr} and the non-vessel area temperature NV_{sr}. As the core temperature increases, P_{sr} increases more than NV_{sr}, but a clear result is not possible, especially at the end where many perforators are present. One explanation is that the detection of perforators is not consistent over time as discussed above, the detected perforators change and so do the temperatures. In previous fields, only one side was analyzed to compare the results with other measurements.

However, the left and right sides must be compared to detect asymmetries. In the fifth field (e) the focus is on the left and right side of the vessel temperature V_{sr} against the non-vessel temperature. Additionally, the virtual speed of the protocol shows the current phase of the experiment. It can be seen that the vessel/non-vessel T_{sr} is similar on both sides for this person. Differences may indicate pathophysiological problems. The following sixth field (f) shows the environmental properties for temperature and humidity. These are not yet included in the analysis. However, they are implicitly included in the image calibration because the thermogram calibration is performed with each image, so if different environmental settings affect the thermograms, they will be taken into account. The effect of different environmental conditions on the thermoregulation was not analyzed in the studies conducted. The next three fields also compare the left and right sides. First, the seventh field (g) compares the temperature of non-vessel areas NV_{sr} to see differences in areas without underlying direct blood vessel related patterns. Additionally, the eighth field (h) shows the number of perforator pixels in the legs. These increase especially towards the end. While the size of the venous pattern decreases with increasing external load (i). These graphs are related to the opposing effects of thermoregulation and adaptation of the body to external activity.

10.7 Further Applications

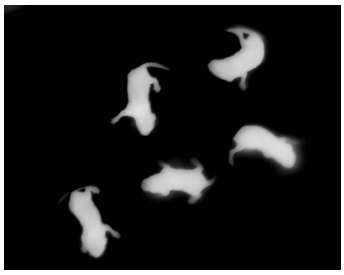
In addition to the field of exercise physiology, further applications can benefit from in-depth thermal analysis. Two promising applications of thermal analysis are the detection of vascular diseases and pharmacological drug studies.

Vascular Diseases Up to now, physiological processes in the body of healthy people have been studied. However, the study of unhealthy individuals and diseases is another potential area for thermography and is already widely applied in medicine [81]. Some of the annotation labels in our dataset come from diseased patients, but a full analysis with our proposed method has not yet been performed. A promising additional use case is the application and adaptation of the system to peripheral artery disease (PAD). Patients with PAD chronically suffer from impaired blood flow to the extremities, especially

the legs, leading to serious problems such as impaired wound healing, nutrition. Symptoms progress gradually from asymptomatic to limited ability to walk with pain, chronic pain at rest, non-healing open wounds, and dying body parts that require amputation. Many studies discuss the differences between the sides to visualize asymmetries in muscles and other parts to detect injuries [134]. Since the studies cannot distinguish between blood vessel patterns or analyze the dynamic movement during exercise, our system is superior when applied to PAD. In PAD diagnosis, a patient has to walk as long as possible until pain occurs on a treadmill, which is about after 200 m [1]. Recently, Ilo et al. [73] showed that IRT can be successfully used to support PAD diagnosis. However, no dynamic measurement is performed during the walking test. Further research is needed to show whether blood vessel discrimination can visualize impaired blood flow and thus improve the current state of PAD diagnosis.

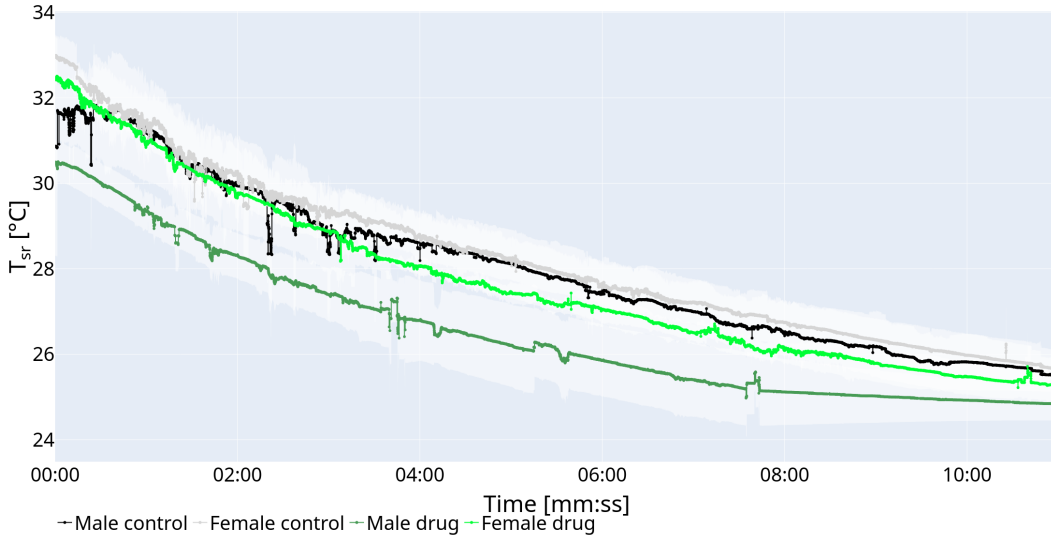
Pharmacology Thermoregulation is not unique to human physiology. The analysis of animals in veterinary medicine and pharmacological research promises additional information. In equine medicine thermography helps to detect injuries [134]. In pharmacological studies, the effect of drugs can be analyzed in a new dimension.

In collaboration with the Department of Pharmacology, University Medical Center Mainz, and the Department of Sports Medicine, Prevention and Rehabilitation, Johannes Gutenberg University Mainz, a small pilot study was conducted on mother mice and their 29 newborn offspring. The aim was to test whether there is an effect on the newborns when the mother has received a drug before pregnancy. All mice were divided into four groups: males and females, and for each, whether the mother had received the drug or not. The newborns were captured with IRT and T_{sr} was extracted from the whole body of the newborn mice over several minutes. The T_{sr} was obtained by averaging over the whole body of each subject. A simple thresholding algorithm separates the mouse from the background (figure 10.10a–b). The T_{sr} of the mice from the four groups are averaged per group and the time series are compared (c). A different thermal response compared to the other groups has been shown for male mice (dark green curve) when the mother animal had the drug treatment. The results are currently being prepared for publication.



(a) Thermogram 23–39° C.

(b) Threshold to segment each mouse.



(c) Time series of averaged T_{sr} for each group. Male group with drug has different T_{sr} compared to the other three groups.

Fig. 10.10.: Example images of an experiment with 5 mice and the time series of the averaged T_{sr} results for all measurements.

11 Conclusion

Applied infrared thermography is a growing field in medical and sports sciences for non-invasive assessment of thermal properties of the human body. Current applications lack reproducibility, easy implementation in new environments, and analysis of the person in motion. In this work, we have developed the automated processing pipeline “ThermoNet” which addresses the major problems with current approaches and successfully prototyped it in studies to demonstrate the potential of the exercise radiomics approach. It is now possible to analyze thermal adaptation in different parts of the body and extract thermal features for comparison with other established measurements. Several studies have shown a relationship between physiological processes and surface thermal radiation patterns. In particular, the extraction of thermal features, especially vascular structures, during exercise is novel and allows non-invasive analysis of human thermoregulation. Our complete analysis strategy is superior to previous manual analysis due to the amount of data processed and less need for standardization. The extraction of vascular patterns provides a new way to assess physiological capacity during activity. Furthermore, the newly developed context-free radiometric calibration allows comparison between multiple experiments. The new deep neural networks for body part and vessel segmentation is based on our manual dataset. In addition, the stereo bootstrapping method for new datasets significantly reduces the amount of manual data required. For vessel data, however, only the manual approach is available. The labels in the manual data are still highly variable, resulting in low segmentation scores. Nevertheless, the thermal features are promising. The time series data for an entire experiment are optimized and merged with other sensory data. Further developments include the integration of frame-to-frame consistent pattern recognition, the expansion of high-quality vascular datasets, and the automatic expansion of body parts to new body regions. To gain new insights into thermorgulation and physical adaptation, the developed processing pipeline needs to be confirmed in new appropriate studies that benefit from the newly available information, especially vascular analysis. Besides, complex and compound movements and new measurement environments can be included.

Bibliography

[1] V. Aboyans, J.-B. Ricco, M.-L. E. L. Bartelink, M. Björck, M. Brodmann, T. Cohnert, J.-P. Collet, M. Czerny, et al. “2017 ESC Guidelines on the Diagnosis and Treatment of Peripheral Arterial Diseases, in collaboration with the European Society for Vascular Surgery (ESVS)”. In: *European Heart Journal* 39.9 (Mar. 2018), pp. 763–816. DOI: 10.1093/eurheartj/ehx095 (cit. on p. 194).

[2] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. “Optuna: A Next-generation Hyperparameter Optimization Framework”. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, NY, USA: ACM, July 2019, pp. 2623–2631. DOI: 10.1145/3292500.3330701 (cit. on p. 61).

[3] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. C. Raff, K. Roberts, P. Walter, J. H. Wilson, T. Hunt, and B. Alberts. *Blood Vessels and Endothelial Cells*. 4th. New York, NY: Garland Science, 2008 (cit. on p. 9).

[4] K. Ammer and F. Ring. *The Thermal Human Body*. Jenny Stanford Publishing, May 2019. DOI: 10.1201/9780429019982 (cit. on p. 11).

[5] D. Andrés López, B. Hillen, M. Nägele, P. Simon, and E. Schömer. *StereoThermoLegs Dataset*. Zenodo, Apr. 2024. DOI: 10.5281/zenodo.8289870 (cit. on pp. 8, 56).

[6] D. Andrés López, B. Hillen, M. Nägele, P. Simon, and E. Schömer. “StereoThermoLegs: label propagation with multimodal stereo cameras for automated annotation of posterior legs during running at different velocities”. In: *Journal of Thermal Analysis and Calorimetry* (June 2024). DOI: 10.1007/s10973-024-13343-w (cit. on pp. 7, 8, 56, 99, 100, 104, 106, 107, 142, 144, 182).

[7] D. Andrés López, B. Hillen, M. Nägele, P. Simon, and E. Schömer. “ThermoNet: advanced deep neural network-based thermogram processing pipeline for automatic time series analysis of specific skin areas in moving legs”. In: *Journal of Thermal Analysis and Calorimetry* JTACC-V4 2023 (Sept. 2024). DOI: 10.1007/s10973-024-13625-3 (cit. on pp. 7, 59, 61, 62, 85, 87, 88, 130–133, 230).

[8] J. Ansel, E. Yang, H. He, N. Gimelshein, A. Jain, M. Voznesensky, B. Bao, P. Bell, et al. “PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation”. In: *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*. ASPLOS ’24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 929–947. DOI: 10.1145/3620665.3640366 (cit. on p. 61).

[9] E. Arens and H. Zhang. “The skin’s role in human thermoregulation and comfort”. In: *Thermal and Moisture Transport in Fibrous Materials*. Elsevier, 2006, pp. 560–602. DOI: 10.1533/9781845692261.3.560 (cit. on p. 11).

[10] P. E. Aylwin, S. Racinais, S. Bermon, A. Lloyd, S. Hodder, and G. Havenith. “The use of infrared thermography for the dynamic measurement of skin temperature of moving athletes during competition; methodological issues”. In: *Physiological Measurement* 42.8 (Aug. 2021), p. 084004. DOI: 10.1088/1361-6579/ac1872 (cit. on p. 174).

[11] A. N. Azhar and M. L. Khodra. “Fine-tuning Pretrained Multilingual BERT Model for Indonesian Aspect-based Sentiment Analysis”. In: *2020 7th International Conference on Advanced Informatics: Concepts, Theory and Applications, ICAICTA 2020* (2020), pp. 2980–2988. DOI: 10.1109/ICAICTA49861.2020.9428882 (cit. on p. 75).

[12] L. B. Baker. “Physiology of sweat gland function: The roles of sweating and sweat composition in human health”. In: *Temperature* 6.3 (2019), pp. 211–259. DOI: 10.1080/23328940.2019.1632145 (cit. on p. 20).

[13] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl. “Algorithms for hyper-parameter optimization”. In: *Advances in Neural Information Processing Systems*. Ed. by J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger. Curran Associates, Inc., 2011 (cit. on p. 82).

[14] J. Bergstra, D. Yamins, and D. Cox. “Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures”. In: *Proceedings of the 30th International Conference on Machine Learning*. Ed. by S. Dasgupta and D. McAllester. Vol. 28. Proceedings of Machine Learning Research 1. Atlanta, Georgia, USA: PMLR, 2013, pp. 115–123 (cit. on p. 82).

[15] M. Berman, A. Rannen, and T. Matthew. “The Lovasz-Softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks”. In: *Cvpr* (2018), pp. 4413–4421 (cit. on p. 77).

[16] F. Bernhard, ed. *Handbuch der Technischen Temperaturmessung*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. DOI: 10.1007/978-3-642-24506-0 (cit. on pp. 27, 159).

[17] J. Bertels, T. Eelbode, M. Berman, D. Vandermeulen, F. Maes, R. Bisschops, and M. B. Blaschko. “Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory and Practice”. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Ed. by D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan. Vol. 11765 LNCS. Cham: Springer, 2019, pp. 92–100. DOI: 10.1007/978-3-030-32245-8_11. eprint: 1911.01685 (cit. on p. 76).

[18] R. Beschi, X. Feng, S. Melillo, L. Parisi, and L. Postiglione. *Stereo camera system calibration: the need of two sets of parameters*. Jan. 2021. arXiv: 2101.05725 (cit. on p. 26).

[19] M. K. Bhowmik, K. Das, and D. Bhattacharjee. “Temperature profile guided segmentation for detection of early subclinical inflammation in arthritis knee joints from thermal images”. In: *Infrared Physics and Technology* 99.January (2019), pp. 102–112. DOI: 10.1016/j.infrared.2019.04.011 (cit. on p. 60).

[20] B. Bischl, M. Binder, M. Lang, T. Pielok, J. Richter, S. Coors, J. Thomas, T. Ullmann, M. Becker, A.-L. Boulesteix, D. Deng, and M. Lindauer. “Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges”. In: *WIREs Data Mining and Knowledge Discovery* 13.2 (Mar. 2023), e1484. DOI: 10.1002/widm.1484 (cit. on p. 82).

[21] S. Bogomilsky, O. Hoffer, G. Shalmon, and M. Scheinowitz. “Preliminary study of thermal density distribution and entropy analysis during cycling exercise stress test using infrared thermography”. In: *Scientific Reports* (2022), pp. 1–7. DOI: 10.1038/s41598-022-18233-5 (cit. on p. 34).

[22] F. Bolelli, S. Allegretti, L. Baraldi, and C. Grana. “Spaghetti Labeling: Directed Acyclic Graphs for Block-Based Connected Components Labeling”. In: *IEEE Transactions on Image Processing* 29.1 (2020), pp. 1999–2012. DOI: 10.1109/TIP.2019.2946979 (cit. on pp. 84, 89).

[23] G. Borg. “Borg’s perceived exertion and pain scales.” In: *Human Kinetics* July 1998 (1998), p. 111 (cit. on p. 16).

[24] G. A. V. Borg. “Perceived exertion: A note on “history” and methods”. In: *Medicine and Science in Sports* 5.2 (1973), pp. 90–93. DOI: 10.1249/00005768-197300520-00017 (cit. on p. 16).

[25] G. Bradski. “The OpenCV Library”. In: *Dr. Dobb’s Journal of Software Tools* (2000) (cit. on pp. 24, 25, 61, 68, 92, 94, 97, 102).

[26] A. Breheret. *Pixel Annotation Tool*. 2017 (cit. on pp. 66, 67).

[27] A. Buades, B. Coll, and J.-m. Morel. “Non-Local Means Denoising”. In: *Image Processing On Line* 1 (Sept. 2011), pp. 208–212. DOI: 10.5201/ipol.2011.bcm_nlm (cit. on p. 96).

[28] S. Bultmann, J. Quenzel, and S. Behnke. “Real-time multi-modal semantic fusion on unmanned aerial vehicles with label propagation for cross-domain adaptation”. In: *Robotics and Autonomous Systems* 159 (Jan. 2023), p. 104286. DOI: 10.1016/j.robot.2022.104286. eprint: arXiv:2210.09739v1 (cit. on p. 92).

[29] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin. “Albumentations: Fast and Flexible Image Augmentations”. In: *Information* 11.2 (2020), pp. 1–20. DOI: 10.3390/info11020125 (cit. on p. 61).

[30] S. Butterworth. “On the Theory of Filter Amplifiers”. In: *Experimental Wireless & The Wireless Engineer* 7.6 (1930), pp. 536–541 (cit. on p. 115).

[31] E. Cabrera, L. Ortiz, B. Silva, E. Clua, and L. Gonçalves. “A Versatile Method for Depth Data Error Estimation in RGB-D Sensors”. In: *Sensors* 18.9 (Sept. 2018), p. 3122. DOI: 10.3390/s18093122 (cit. on pp. 21, 22).

[32] J. Canny. “A Computational Approach to Edge Detection”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8.6 (Nov. 1986), pp. 679–698. DOI: 10.1109/TPAMI.1986.4767851 (cit. on p. 107).

[33] N. Charkoudian. “Mechanisms and modifiers of reflex induced cutaneous vasodilation and vasoconstriction in humans”. In: *Journal of Applied Physiology* 109.4 (Oct. 2010), pp. 1221–1228. DOI: 10.1152/japplphysiol.00298.2010 (cit. on p. 13).

[34] M. Charlton, S. A. Stanley, Z. Whitman, V. Wenn, T. J. Coats, M. Sims, and J. P. Thompson. “The effect of constitutive pigmentation on the measured emissivity of human skin”. In: *PLOS ONE* 15.11 (Nov. 2020). Ed. by F. Li, e0241843. DOI: 10.1371/journal.pone.0241843 (cit. on pp. 160, 161, 166).

[35] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou. *TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation*. Feb. 2021 (cit. on p. 73).

[36] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation”. In: *Computer Vision – ECCV 2018*. Ed. by V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss. Vol. 11211 LNCS. Cham: Springer, 2018, pp. 833–851. DOI: 10.1007/978-3-030-01234-2_49. eprint: 1802.02611 (cit. on pp. 74, 99).

[37] S. N. Cheuvront and R. W. Kenefick. “CORP: Improving the status quo for measuring whole body sweat losses”. In: *Journal of Applied Physiology* 123.3 (2017), pp. 632–636. DOI: 10.1152/japplphysiol.00433.2017 (cit. on p. 20).

[38] I. Cruz-Vega, D. Hernandez-Contreras, H. Peregrina-Barreto, J. d. J. Rangel-Magdaleno, and J. M. Ramirez-Cortes. “Deep learning classification for diabetic foot thermograms”. In: *Sensors (Switzerland)* 20.6 (2020), pp. 1–22. DOI: 10.3390/s20061762 (cit. on p. 60).

[39] CVAT.ai Corporation. *Computer Vision Annotation Tool (CVAT)*. 2024. DOI: 10.5281/zenodo.10783399 (cit. on p. 65).

[40] H. A. Daanen, V. Kohlen, and L. P. Teunissen. “Heat flux systems for body core temperature assessment during exercise”. In: *Journal of Thermal Biology* 112.January (2023), p. 103480. DOI: 10.1016/j.jtherbio.2023.103480 (cit. on p. 50).

[41] K. Das, M. K. Bhowmik, and D. Prasad Mukherjee. “Segmentation of Knee Thermograms for Detecting Inflammation”. In: *2019 IEEE International Conference on Image Processing (ICIP)*. Taipei, Taiwan: IEEE, Sept. 2019, pp. 1550–1554. DOI: 10.1109/ICIP.2019.8803094 (cit. on p. 60).

[42] T. DeVries and G. W. Taylor. “Improved Regularization of Convolutional Neural Networks with Cutout”. In: (Aug. 2017). arXiv: 1708.04552 (cit. on p. 72).

[43] A. Dias, C. Bras, A. Martins, J. Almeida, and E. Silva. “Thermographic and visible spectrum camera calibration for marine robotic target detection”. In: *OCEANS 2013 MTS/IEEE - San Diego: An Ocean in Common*. 2013. DOI: 10.23919/OCEANS.2013.6741230 (cit. on p. 94).

[44] H.-H. Dickhuth, M. Huonker, T. Münzel, H. Drexler, A. Berg, and J. Keul. “Individual Anaerobic Threshold for Evaluation of Competitive Athletes and Patients with Left Ventricular Dysfunction”. In: *Advances in Ergometry*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1991, pp. 173–179. DOI: 10.1007/978-3-642-76442-4_26 (cit. on p. 18).

[45] H.-H. Dickhuth, L. Yin, A. Niess, K. Röcker, F. Mayer, H.-C. Heitkamp, and T. Horstmann. “Ventilatory, Lactate-Derived and Catecholamine Thresholds During Incremental Treadmill Running: Relationship and Reproducibility”. In: *International Journal of Sports Medicine* 20.02 (Feb. 1999), pp. 122–127. DOI: 10.1055/s-2007-971105 (cit. on p. 18).

[46] P. H. C. Eilers. “A Perfect Smoother”. In: *Analytical Chemistry* 75.14 (July 2003), pp. 3631–3636. DOI: 10.1021/ac034173t (cit. on p. 187).

[47] S. Farhadpour, T. A. Warner, and A. E. Maxwell. “Selecting and Interpreting Multiclass Loss and Accuracy Assessment Metrics for Classifications with Class Imbalance: Guidance and Best Practices”. In: *Remote Sensing* 16.3 (Jan. 2024), p. 533. DOI: 10.3390/rs16030533 (cit. on p. 173).

[48] O. Faude, W. Kindermann, and T. Meyer. “Lactate threshold concepts: how valid are they?” In: *Sports medicine (Auckland, N.Z.)* 39.6 (2009), pp. 469–90. DOI: 10.2165/00007256-200939060-00003 (cit. on p. 17).

[49] P. Fechteler and P. Eisert. “Adaptive color classification for structured light systems”. In: *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, June 2008, pp. 1–7. DOI: 10.1109/CVPRW.2008.4563048 (cit. on p. 34).

[50] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman. “An Ensemble Classification-Based Approach Applied to Retinal Blood Vessel Segmentation”. In: *IEEE Transactions on Biomedical Engineering* 59.9 (Sept. 2012), pp. 2538–2548. DOI: 10.1109/TBME.2012.2205687 (cit. on pp. 53, 171).

[51] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez. “Automatic generation and detection of highly reliable fiducial markers under occlusion”. In: *Pattern Recognition* 47.6 (June 2014), pp. 2280–2292. DOI: 10.1016/j.patcog.2014.01.005 (cit. on pp. 45, 46, 48).

[52] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, et al. “Array programming with NumPy”. In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2 (cit. on p. 61).

[53] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Second. New York, NY, USA: Cambridge University Press, Mar. 2004. DOI: 10.1017/CBO9780511811685 (cit. on pp. 25–27, 34, 93, 102).

[54] R. I. Hartley. “Theory and practice of projective rectification”. In: *International Journal of Computer Vision* 35.2 (1999), pp. 115–127. DOI: 10.1023/A:1008115206617 (cit. on p. 102).

[55] T. Hastie, R. Tibshirani, and J. Friedman. “The Elements of Statistical Learning - Data Mining, Inference, and Prediction”. In: *Springer Series in Statistics* 27.2 (2009), p. 745 (cit. on p. 81).

[56] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li. “Bag of tricks for image classification with convolutional neural networks”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2019-June (2019), pp. 558–567. DOI: 10.1109/CVPR.2019.00065. eprint: 1812.01187 (cit. on p. 170).

[57] Y. He, B. Deng, H. Wang, L. Cheng, K. Zhou, S. Cai, and F. Ciampa. “Infrared machine vision and infrared thermography with deep learning: A review”. In: *Infrared Physics and Technology* 116.April (2021). DOI: 10.1016/j.infrared.2021.103754 (cit. on p. 60).

[58] B. Hillen, D. Andrés López, J. M. Marzano-Felisatti, J. L. Sanchez-Jimenez, R. M. Cibrián Ortiz de Anda, M. Nägele, M. R. Salvador-Palmer, P. Pérez-Soriano, E. Schömer, P. Simon, and J. I. Priego-Quesada. “Acute physiological responses to a pyramidal exercise protocol and the associations with skin temperature variation in different body areas”. In: *Journal of Thermal Biology* 115.May (July 2023), p. 103605. DOI: 10.1016/j.jtherbio.2023.103605 (cit. on pp. 7, 8, 160, 164, 190, 191).

[59] B. Hillen, D. Andrés López, D. Pfirrmann, E. W. Neuberger, K. Mertinat, M. Nägele, E. Schömer, and P. Simon. “An exploratory, intra- and interindividual comparison of the deep neural network automatically measured calf surface radiation temperature during cardiopulmonary running and cycling exercise testing: A preliminary study”. In: *Journal of Thermal Biology* 113 (Apr. 2023), p. 103498. DOI: 10.1016/j.jtherbio.2023.103498 (cit. on pp. 7, 55, 59, 61, 130, 189, 190).

[60] B. Hillen, D. Andrés López, E. Schömer, M. Nagele, and P. Simon. “Towards Exercise Radiomics: Deep Neural Network-Based Automatic Analysis of Thermal Images Captured During Exercise”. In: *IEEE Journal of Biomedical and Health Informatics* 26.9 (Sept. 2022), pp. 4530–4540. DOI: 10.1109/JBHI.2022.3186530 (cit. on pp. 7, 59, 61–64, 74, 130–132, 136–138, 173, 174, 188, 230).

[61] B. Hillen, D. Pfirrmann, M. Nägele, and P. Simon. “Infrared Thermography in Exercise Physiology: The Dawning of Exercise Radiomics”. In: *Sports Medicine* 50.2 (2020), pp. 263–282. DOI: 10.1007/s40279-019-01210-w (cit. on pp. 4, 5, 13, 15, 16, 33, 53, 63, 129).

[62] B. Hillen, P. Simon, S. Schlotter, O. Nitsche, V. Bähner, K. Poplawska, and D. Pfirrmann. “Feasibility and implementation of a personalized, web-based exercise intervention for people with cystic fibrosis for 1 year”. In: *BMC Sports Science, Medicine and Rehabilitation* 13.1 (Dec. 2021), p. 95. DOI: 10.1186/s13102-021-00323-y (cit. on p. 54).

[63] G. Hinton, N. Srivastava, and K. Swersky. “Lecture 6e - rmsprop: Divide the gradient by a running average of its recent magnitude”. In: *COURSERA: Neural networks for machine learning*. 2012, pp. 26–31 (cit. on p. 80).

[64] G. Hinton, O. Vinyals, and J. Dean. *Distilling the Knowledge in a Neural Network*. 2015. arXiv: 1503.02531 (cit. on p. 171).

[65] W. Hollmann and J. P. Prinz. "Ergospirometry and its History". In: *Sports Medicine* 23.2 (Feb. 1997), pp. 93–105. DOI: 10.2165/00007256-199723020-00003 (cit. on p. 18).

[66] R. Horaud, M. Hansard, G. Evangelidis, and C. Ménier. "An overview of depth cameras and range scanners based on time-of-flight technologies". In: *Machine Vision and Applications* 27.7 (Oct. 2016), pp. 1005–1020. DOI: 10.1007/s00138-016-0784-4. eprint: 2012.06772 (cit. on p. 34).

[67] X. Hu, L. Fuxin, D. Samaras, and C. Chen. "Topology-preserving deep image segmentation". In: *Advances in Neural Information Processing Systems* 32.NeurIPS (2019), pp. 1–12. arXiv: 1906.05404 (cit. on p. 169).

[68] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. "Densely connected convolutional networks". In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* 2017-January (2017), pp. 2261–2269. DOI: 10.1109/CVPR.2017.243. eprint: 1608.06993 (cit. on p. 73).

[69] H. Hymczak, A. Golab, K. Mendrala, D. Plicner, T. Darocha, P. Podsiadlo, D. Hudziak, R. Gocol, and S. Kosiński. "Core temperature measurement—principles of correct measurement, problems, and complications". In: *International Journal of Environmental Research and Public Health* 18.20 (2021). DOI: 10.3390/ijerph182010606 (cit. on pp. 12, 19).

[70] N. F. A. Ibrahim, N. Sabani, S. Johari, A. A. Manaf, A. A. Wahab, Z. Zakaria, and A. M. Noor. "A Comprehensive Review of the Recent Developments in Wearable Sweat-Sensing Devices". In: *Sensors* 22.19 (2022). DOI: 10.3390/s22197670 (cit. on p. 20).

[71] T. Igarashi, K. Nishino, and S. K. Nayar. "The appearance of human skin: A survey". In: *Foundations and Trends in Computer Graphics and Vision* 3.1 (2007), pp. 1–95. DOI: 10.1561/0600000013 (cit. on p. 14).

[72] T. L. Igarashi, T. L. Fernandes, A. J. Hernandez, C. E. Keutenedjian Mady, and C. Albuquerque. "Behavior of skin temperature during incremental cycling and running indoor exercises". In: *Heliyon* 8.10 (Oct. 2022), e10889. DOI: 10.1016/j.heliyon.2022.e10889 (cit. on p. 15).

[73] A. Ilo, P. Romsi, and J. Mäkelä. "Infrared Thermography as a Diagnostic Tool for Peripheral Artery Disease". In: *Advances in Skin & Wound Care* 33.9 (Sept. 2020), pp. 482–488. DOI: 10.1097/01.ASW.0000694156.62834.8b (cit. on p. 194).

[74] M. Iman, H. R. Arabnia, and K. Rasheed. "A Review of Deep Transfer Learning and Recent Advancements". In: *Technologies* 11.2 (Mar. 2023), p. 40. DOI: 10.3390/technologies11020040. eprint: 2201.09679 (cit. on p. 171).

[75] P. Izmailov, D. Podoprikhin, T. Garipov, D. Vetrov, and A. G. Wilson. “Averaging weights leads to wider optima and better generalization”. In: *34th Conference on Uncertainty in Artificial Intelligence 2018, UAI 2018* 2 (2018), pp. 876–885. arXiv: 1803.05407 (cit. on p. 175).

[76] K. Jamieson and A. Talwalkar. “Non-stochastic Best Arm Identification and Hyperparameter Optimization”. In: *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, AISTATS 2016* (Feb. 2015), pp. 240–248. arXiv: 1502.07943 (cit. on p. 82).

[77] S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio. “The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation”. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* 2017-July (2017), pp. 1175–1183. DOI: 10.1109/CVPRW.2017.156. eprint: 1611.09326 (cit. on p. 73).

[78] R. E. Jurdi, C. Petitjean, P. Honeine, V. Cheplygina, V. C. Nl, and F. Abdallah. “A Surprisingly Effective Perimeter-based Loss for Medical Image Segmentation”. In: *Proceedings of Machine Learning Research* 143 (2021), pp. 158–167 (cit. on p. 78).

[79] G. P. Kenny. “Human thermoregulation: separating thermal and nonthermal effects on heat loss”. In: *Frontiers in Bioscience* 15.1 (2010), p. 259. DOI: 10.2741/3620 (cit. on p. 11).

[80] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. Ben Ayed. “Boundary loss for highly unbalanced segmentation”. In: *Medical Image Analysis* 67 (2021), pp. 1–21. DOI: 10.1016/j.media.2020.101851. eprint: 1812.07032 (cit. on p. 78).

[81] D. Kesztyüs, S. Brucher, C. Wilson, and T. Kesztyüs. “Use of Infrared Thermography in Medical Diagnosis, Screening, and Disease Monitoring: A Scoping Review”. In: *Medicina* 59.12 (Dec. 2023), p. 2139. DOI: 10.3390/medicina59122139 (cit. on p. 193).

[82] T. L. B. Khanh, D.-P. Dao, N.-H. Ho, H.-J. Yang, E.-T. Baek, G. Lee, S.-H. Kim, and S. B. Yoo. “Enhancing U-Net with Spatial-Channel Attention Gate for Abnormal Tissue Segmentation in Medical Imaging”. In: *Applied Sciences* 10.17 (Aug. 2020), p. 5729. DOI: 10.3390/app10175729 (cit. on p. 73).

[83] D. P. Kingma and J. L. Ba. “Adam: A method for stochastic optimization”. In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (2015), pp. 1–15. arXiv: 1412.6980 (cit. on p. 80).

[84] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollar. “Panoptic Segmentation”. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 2019-June. IEEE, June 2019, pp. 9396–9405. DOI: 10.1109/CVPR.2019.00963. eprint: 1801.00868 (cit. on pp. 64, 65).

[85] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick. *Segment Anything*. Apr. 2023. arXiv: 2304.02643 (cit. on p. 180).

[86] V. V. Kniaz, V. A. Knyaz, J. Hladůvka, W. G. Kropatsch, and V. Mizginov. "ThermalGAN: Multimodal Color-to-Thermal Image Translation for Person Re-identification in Multispectral Dataset". In: *Computer Vision – ECCV 2018 Workshops*. Springer International Publishing, 2019, pp. 606–624. DOI: 10.1007/978-3-030-11024-6_46 (cit. on pp. 53, 60).

[87] V. A. Knyaz and P. V. Moshkantsev. "JOINT GEOMETRIC CALIBRATION OF COLOR AND THERMAL CAMERAS FOR SYNCHRONIZED MULTIMODAL DATASET CREATING". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-2/W18 (Nov. 2019), pp. 79–84. DOI: 10.5194/isprs-archives-XLII-2-W18-79-2019 (cit. on p. 92).

[88] F. Kofler, S. Shit, I. Ezhov, L. Fidon, I. Horvath, R. Al-Maskari, H. B. Li, H. Bhatia, et al. "blob loss: Instance Imbalance Aware Loss Functions for Semantic Segmentation". In: *2023 International Conference on Information Processing in Medical Imaging*. Ed. by A. Frangi, M. de Bruijne, D. Wassermann, and N. Navab. Cham: Springer Nature Switzerland, 2023, pp. 755–767. DOI: 10.1007/978-3-031-34048-2_58 (cit. on pp. 169, 172).

[89] M. Kopaczka, R. Kolk, and D. Merhof. "A fully annotated thermal face database and its application for thermal facial expression recognition". In: *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. Houston, TX, USA: IEEE, May 2018, pp. 1–6. DOI: 10.1109/I2MTC.2018.8409768 (cit. on p. 53).

[90] A. S. Kornilov and I. V. Safonov. "An overview of watershed algorithm implementations in open source libraries". In: *Journal of Imaging* 4.10 (2018). DOI: 10.3390/jimaging4100123 (cit. on pp. 66, 99).

[91] B. M. Lane and E. P. Whitenton. *Calibration and Measurement Procedures for a High Magnification Thermal Camera*. Tech. rep. Gaithersburg, MD: National Institute of Standards and Technology, Jan. 2016. DOI: 10.6028/NIST.IR.8098 (cit. on p. 44).

[92] D. Lang. "Generative Bildmodelle zur Synthese von Thermogrammen aus optischen Aufnahmen". Bachelor thesis. Johannes Gutenberg University Mainz, 2023 (cit. on pp. 172, 185).

[93] R. J. Lee, S. Sivakumar, and K. H. Lim. "Review on remote heart rate measurements using photoplethysmography". In: *Multimedia Tools and Applications* (2023). DOI: 10.1007/s11042-023-16794-9 (cit. on p. 17).

[94] L. Li, K. Jamieson, A. Rostamizadeh, E. Gonina, M. Hardt, B. Recht, and A. Talwalkar. *A System for Massively Parallel Hyperparameter Tuning*. Oct. 2018. arXiv: 1810.05934 (cit. on p. 82).

[95] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar. "Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization". In: *Journal of Machine Learning Research* 18 (Mar. 2016), pp. 1–52. arXiv: 1603.06560 (cit. on p. 82).

[96] X. Li, Y. Yang, B. Zhang, X. Lin, X. Fu, Y. An, Y. Zou, J.-X. Wang, Z. Wang, and T. Yu. "Lactate metabolism in human health and disease". In: *Signal Transduction and Targeted Therapy* 7.1 (Sept. 2022), p. 305. DOI: 10.1038/s41392-022-01151-3 (cit. on p. 17).

[97] D. Lin, P. Westfeld, and H. G. Maas. "Shutter-less temperature-dependent correction for uncooled thermal camera under fast changing FPA temperature". In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 42.1W1 (2017), pp. 619–625. DOI: 10.5194/isprs-archives-XLII-1-W1-619-2017 (cit. on p. 45).

[98] J. Long, E. Shelhamer, and T. Darrell. "Fully convolutional networks for semantic segmentation". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA: IEEE, 2015, pp. 3431–3440. DOI: 10.1109/CVPR.2015.7298965 (cit. on p. 99).

[99] I. Loshchilov and F. Hutter. "Decoupled Weight Decay Regularization". In: *International Conference on Learning Representations*. Nov. 2019 (cit. on p. 80).

[100] I. Loshchilov and F. Hutter. "SGDR: Stochastic gradient descent with warm restarts". In: *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings* (2017), pp. 1–16. arXiv: 1608.03983 (cit. on p. 175).

[101] T. Luhmann, J. Piechel, and T. Roelfs. "Geometric calibration of thermographic cameras". In: *Remote Sensing and Digital Image Processing* 17.1 (2013), pp. 27–42. DOI: 10.1007/978-94-007-6639-6_2 (cit. on pp. 94, 179).

[102] Y. Luo and Z. Luo. "Infrared and Visible Image Fusion: Methods, Datasets, Applications, and Prospects". In: *Applied Sciences* 13.19 (Sept. 2023), p. 10891. DOI: 10.3390/app131910891 (cit. on p. 92).

[103] J. Ma. *Segmentation Loss Odyssey*. May 2020. arXiv: 2005.13449 (cit. on p. 75).

[104] Á. S. Machado, J. I. Priego-Quesada, I. Jimenez-Perez, M. Gil-Calvo, F. P. Carpes, and P. Perez-Soriano. "Influence of infrared camera model and evaluator reproducibility in the assessment of skin temperature responses to physical exercise". In: *Journal of Thermal Biology* 98.March (2021). DOI: 10.1016/j.jtherbio.2021.102913 (cit. on p. 5).

[105] G. Machin, R. Simpson, and M. Broussely. "Calibration and validation of thermal imagers". In: *Quantitative InfraRed Thermography Journal* 6.2 (Dec. 2009), pp. 133–147. DOI: 10.3166/qirt.6.133-147 (cit. on p. 44).

[106] B. A. MacRae, S. Annaheim, C. M. Spengler, and R. M. Rossi. "Skin temperature measurement using contact thermometry: A systematic review of setup variables and their effects on measured values". In: *Frontiers in Physiology* 9.JAN (2018), pp. 1–24. DOI: 10.3389/fphys.2018.00029 (cit. on p. 20).

[107] C. Magalhaes, J. Mendes, and R. Vardasca. "Meta-Analysis and Systematic Review of the Application of Machine Learning Classifiers in Biomedical Applications of Infrared Thermography". In: *Applied Sciences* 11.2 (Jan. 2021), p. 842. DOI: 10.3390/app11020842 (cit. on p. 33).

[108] C. Magalhaes, J. M. R. Tavares, J. Mendes, and R. Vardasca. “Comparison of machine learning strategies for infrared thermography of skin cancer”. In: *Biomedical Signal Processing and Control* 69.August (2021). DOI: 10.1016/j.bspc.2021.102872 (cit. on p. 60).

[109] D. Maji, S. Nagori, M. Mathew, and D. Poddar. “YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss”. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. New Orleans, LA, USA: IEEE, June 2022, pp. 2636–2645. DOI: 10.1109/CVPRW56347.2022.00297 (cit. on p. 99).

[110] P. Malhotra, S. Gupta, D. Koundal, A. Zaguia, and W. Enbeyle. “Deep Neural Networks for Medical Image Segmentation”. In: *Journal of Healthcare Engineering* 2022 (Mar. 2022). Ed. by C. Chakraborty, pp. 1–15. DOI: 10.1155/2022/9580991 (cit. on p. 68).

[111] M. C. E. Manuel, S.-P. Lin, W.-H. Lu, and P. T. Lin. “Errors in Thermographic Camera Measurement Caused by Known Heat Sources and Depth Based Correction”. In: *International Journal of Automation and Smart Technology* 6.1 (Mar. 2016), pp. 5–12. DOI: 10.5875/ausmt.v6i1.1003 (cit. on p. 92).

[112] R. Mazaheri, C. Schmied, D. Niederseer, and M. Guazzi. “Cardiopulmonary Exercise Test Parameters in Athletic Population: A Review”. In: *Journal of Clinical Medicine* 10.21 (Oct. 2021), p. 5073. DOI: 10.3390/jcm10215073 (cit. on p. 18).

[113] S. Mazdeyasna, P. Ghassemi, and Q. Wang. “Best Practices for Body Temperature Measurement with Infrared Thermography: External Factors Affecting Accuracy”. In: *Sensors* 23.18 (Sept. 2023), p. 8011. DOI: 10.3390/s23188011 (cit. on pp. 45, 160, 161, 163).

[114] W. McKinney. “Data Structures for Statistical Computing in Python”. In: *Proceedings of the 9th Python in Science Conference*. Ed. by S. van der Walt and J. Millman. 2010, pp. 56–61. DOI: 10.25080/Majora-92bf1922-00a (cit. on p. 122).

[115] A. Merla, P. A. Mattei, L. Di Donato, and G. L. Romani. “Thermal imaging of cutaneous temperature modifications in runners during graded exercise”. In: *Annals of Biomedical Engineering* 38.1 (2010), pp. 158–163. DOI: 10.1007/s10439-009-9809-8 (cit. on p. 172).

[116] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos. “Image Segmentation Using Deep Learning: A Survey”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.7 (2022), pp. 3523–3542. DOI: 10.1109/TPAMI.2021.3059968. eprint: 2001.05566 (cit. on p. 68).

[117] A. Mumuni and F. Mumuni. “Data augmentation: A comprehensive survey of modern approaches”. In: *Array* 16.August (Dec. 2022), p. 100258. DOI: 10.1016/j.array.2022.100258 (cit. on p. 71).

[118] F. Niklaus, C. Vieider, and H. Jakobsen. “MEMS-based uncooled infrared bolometer arrays: a review”. In: *MEMS/MOEMS Technologies and Applications III* 6836.March (2007), p. 68360D. DOI: 10.1117/12.755128 (cit. on p. 32).

[119] D. Nister. “An efficient solution to the five-point relative pose problem”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26.6 (June 2004), pp. 756–770. DOI: 10.1109/TPAMI.2004.17 (cit. on p. 98).

[120] P. W. Nugent, J. A. Shaw, and N. J. Pust. “Radiometric calibration of infrared imagers using an internal shutter as an equivalent external blackbody”. In: *Optical Engineering* 53.12 (2014), p. 123106. DOI: 10.1117/1.oe.53.12.123106 (cit. on p. 44).

[121] D. T. Ochmann, K. F. A. Philippi, P. Zeier, M. Sandner, B. Hillen, E. W. I. Neuberger, I. Ruiz de Azua, K. Lieb, M. Wessa, B. Lutz, P. Simon, and A. Brahmer. “Association of Innate and Acquired Aerobic Capacity With Resilience in Healthy Adults: Protocol for a Randomized Controlled Trial of an 8-Week Web-Based Physical Exercise Intervention”. In: *JMIR Research Protocols* 10.11 (Nov. 2021), e29712. DOI: 10.2196/29712 (cit. on p. 55).

[122] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert. “Attention U-Net: Learning where to look for the pancreas”. In: *1st Conference on Medical Imaging with Deep Learning*. Amsterdam, Netherlands, 2018. arXiv: 1804.03999 (cit. on pp. 73, 74).

[123] N. Otsu. “A Threshold Selection Method from Gray-Level Histograms”. In: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (Jan. 1979), pp. 62–66. DOI: 10.1109/TSMC.1979.4310076 (cit. on p. 106).

[124] J. C. Paetzold, I. Ezhov, G. Tetteh, A. Ertürk, and B. Menze. “clDice-a Novel Connectivity-Preserving Loss Function for Vessel Segmentation”. In: NeurIPS (2019) (cit. on p. 77).

[125] A. Parashar, R. Rishi, A. Parashar, and I. Rida. “Medical imaging in rheumatoid arthritis: A review on deep learning approach”. In: *Open Life Sciences* 18.1 (2023), pp. 15–17. DOI: 10.1515/biol-2022-0611 (cit. on p. 60).

[126] V. Pavez, G. Hermosilla, M. Silva, and G. Farias. “Advanced Deep Learning Techniques for High-Quality Synthetic Thermal Image Generation”. In: *Mathematics* 11.21 (Oct. 2023), p. 4446. DOI: 10.3390/math11214446 (cit. on p. 61).

[127] J. L. Pech-Pacheco, G. Cristöbal, J. Chamorro-Martínez, and J. Fernândez-Valdivia. “Diatom autofocusing in brightfield microscopy: A comparative study”. In: *Proceedings - International Conference on Pattern Recognition* 15.3 (2000), pp. 314–317. DOI: 10.1109/icpr.2000.903548 (cit. on pp. 89, 157).

[128] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on p. 61).

[129] J. D. Périard, T. M. H. Eijsvogels, and H. A. M. Daanen. “Exercise under heat stress: thermoregulation, hydration, performance implications, and mitigation strategies”. In: *Physiological Reviews* 101.4 (Oct. 2021), pp. 1873–1979. DOI: 10.1152/physrev.00038.2020 (cit. on pp. 11–13).

[130] D. Perpetuini, D. Formenti, D. Cardone, C. Filippini, and A. Merla. “Regions of interest selection and thermal imaging data analysis in sports and exercise science: a narrative review”. In: *Physiological Measurement* 42.8 (Aug. 2021), 08TR01. DOI: 10.1088/1361-6579/ac0fbd (cit. on pp. 4, 172).

[131] B. Perret and J. Cousty. “Component Tree Loss Function: Definition and Optimization”. In: *DGMM 2022: Discrete Geometry and Mathematical Morphology*. Strasbourg, France: Springer, Cham, 2022, pp. 248–260. DOI: 10.1007/978-3-031-19897-7_20 (cit. on p. 169).

[132] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld. “Adaptive histogram equalization and its variations”. In: *Computer Vision, Graphics, and Image Processing* 39.3 (Sept. 1987), pp. 355–368. DOI: 10.1016/S0734-189X(87)80186-X (cit. on p. 67).

[133] S. Plagenhoef, F. G. Evans, and T. Abdelnour. “Anatomical Data for Analyzing Human Motion”. In: *Research Quarterly for Exercise and Sport* 54.2 (June 1983), pp. 169–178. DOI: 10.1080/02701367.1983.10605290 (cit. on p. 161).

[134] J. I. Priego-Quesada. *Application of Infrared Thermography in Sports Science*. Ed. by J. I. Priego Quesada. Biological and Medical Physics, Biomedical Engineering. Cham: Springer International Publishing, 2017. DOI: 10.1007/978-3-319-47410-6 (cit. on pp. 4, 5, 11, 12, 33, 177, 194).

[135] J. I. Priego-Quesada. “New Advances in Human Thermophysiology”. In: *Life* 12.8 (2022), pp. 11–13. DOI: 10.3390/life12081261 (cit. on p. 4).

[136] J. I. Priego-Quesada, A. S. Machado, M. Gil-Calvo, I. Jimenez-Perez, R. M. Cibrian Ortiz de Anda, R. Salvador Palmer, and P. Perez-Soriano. “A methodology to assess the effect of sweat on infrared thermography data after running: Preliminary study”. In: *Infrared Physics & Technology* 109 (Sept. 2020), p. 103382. DOI: 10.1016/j.infrared.2020.103382 (cit. on p. 166).

[137] M. Pugsley and R. Tabrizchi. “The vascular system”. In: *Journal of Pharmacological and Toxicological Methods* 44.2 (Sept. 2000), pp. 333–340. DOI: 10.1016/S1056-8719(00)00125-8 (cit. on pp. 9, 10).

[138] M. Qian, Y. Fu, X. Tan, Y. Li, J. Qi, H. Lu, S. Wen, and E. Ding. *Coherent Loss: A Generic Framework for Stable Video Segmentation*. 2020. arXiv: 2010.13085 (cit. on p. 176).

[139] T. A. Qureshi, M. Habib, A. Hunter, and B. Al-diri. “A Manually-Labeled , Artery / Vein Classified Benchmark for the DRIVE Dataset Sunderland Eye Infirmary , UK”. In: *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems* (2013), pp. 485–488 (cit. on pp. 53, 171).

[140] S. Racinais, G. Havenith, P. Aylwin, M. Ihsan, L. Taylor, P. E. Adami, M.-C. Adamuz, M. Alhammoud, et al. “Association between thermal responses, medical events, performance, heat acclimation and health status in male and female elite athletes during the 2019 Doha World Athletics Championships”. In: *British Journal of Sports Medicine* 56.8 (Apr. 2022), pp. 439–445. DOI: 10.1136/bjsports-2021-104569 (cit. on p. 4).

[141] T. Rädsch, A. Reinke, V. Weru, M. D. Tizabi, N. Schreck, A. E. Kavur, B. Pekdemir, T. Roß, A. Kopp-Schneider, and L. Maier-Hein. “Labelling instructions matter in biomedical image analysis”. In: *Nature Machine Intelligence* 5.3 (Mar. 2023), pp. 273–283. DOI: 10.1038/s42256-023-00625-5. eprint: 2207.09899 (cit. on p. 164).

[142] J. Rangel, S. Soldan, and A. Kroll. “3D Thermal Imaging: Fusion of Thermography and Depth Cameras”. In: *Proceedings of the 2014 International Conference on Quantitative InfraRed Thermography*. Bordeaux, France: QIRT Council, 2014. DOI: 10.21611/qirt.2014.035 (cit. on p. 92).

[143] M. Rebol and P. Knöbelreiter. “Frame-To-Frame Consistent Semantic Segmentation”. In: *Joint Austrian Computer Vision And Robotics Workshop (ACVRW)*. Apr. 2020, pp. 79–86. DOI: 10.3217/978-3-85125-752-6-18. eprint: 2008.00948 (cit. on p. 176).

[144] J. Richter, C. Wiede, S. Kaden, M. Weigert, and G. Hirtz. “Skin Temperature Measurement based on Human Skeleton Extraction and Infra-red Thermography - An Application of Sensor Fusion Methods in the Field of Physical Training”. In: *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. Vol. 6. Visigrapp. Porto, Portugal: SCITEPRESS - Science and Technology Publications, 2017, pp. 59–66. DOI: 10.5220/0006095100590066 (cit. on pp. 93, 179).

[145] E. F. J. Ring and K. Ammer. “The technique of infrared imaging in medicine*”. In: *Infrared Imaging*. 2053-2563. Bristol, UK: IOP Publishing, Sept. 2015, pp. 1–10. DOI: 10.1088/978-0-7503-1143-4ch1 (cit. on p. 44).

[146] O. Ronneberger, P. Fischer, and T. Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi. Cham: Springer International Publishing, 2015, pp. 234–241 (cit. on p. 73).

[147] C. Rother, V. Kolmogorov, and A. Blake. “"GrabCut" - Interactive foreground extraction using iterated graph cuts”. In: *ACM Transactions on Graphics* 23.3 (2004), pp. 309–314. DOI: 10.1145/1015706.1015720 (cit. on p. 68).

[148] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. “LabelMe: A Database and Web-Based Tool for Image Annotation”. In: *International Journal of Computer Vision* 77.1-3 (May 2008), pp. 157–173. DOI: 10.1007/s11263-007-0090-8 (cit. on p. 65).

[149] M. Saint-Cyr, C. Wong, M. Schaverien, A. Mojallal, and R. J. Rohrich. “The Perforasome Theory: Vascular Anatomy and Clinical Implications”. In: *Plastic and Reconstructive Surgery* 124.5 (Nov. 2009), pp. 1529–1544. DOI: 10.1097/PRS.0b013e3181b98a6c (cit. on p. 9).

[150] S. S. M. Salehi, D. Erdogmus, and A. Gholipour. “Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 10541 LNCS. 2017, pp. 379–387. DOI: 10.1007/978-3-319-67389-9_44. eprint: 1706.05721 (cit. on p. 77).

[151] S. Sammito and I. Böckelmann. “Möglichkeiten und Einschränkungen der Herzfrequenzmessung und der Analyse der Herzfrequenzvariabilität mittels mobiler Messgeräte: Eine systematische Literaturübersicht”. In: *Herzschrittmachertherapie und Elektrophysiologie* 27.1 (2016), pp. 38–45. DOI: 10.1007/s00399-016-0419-5 (cit. on p. 17).

[152] A. Savitzky and M. J. E. Golay. “Smoothing and Differentiation of Data by Simplified Least Squares Procedures.” In: *Analytical Chemistry* 36.8 (July 1964), pp. 1627–1639. DOI: 10.1021/ac60214a047 (cit. on pp. 120, 147).

[153] Y. Sheng and L. Zhu. “The crosstalk between autonomic nervous system and blood vessels.” In: *International journal of physiology, pathophysiology and pharmacology* 10.1 (2018), pp. 17–28 (cit. on pp. 10, 11).

[154] S. Shit, J. C. Paetzold, A. Sekuboyina, I. Ezhov, A. Unger, A. Zhylka, J. P. W. Pluim, U. Bauer, and B. H. Menze. “clDice-a Novel Topology-Preserving Loss Function for Tubular Structure Segmentation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 16560–16569 (cit. on p. 77).

[155] P. Simard, D. Steinkraus, and J. Platt. “Best practices for convolutional neural networks applied to visual document analysis”. In: *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.* Vol. 1. Icdar. IEEE Comput. Soc, 2003, pp. 958–963. DOI: 10.1109/ICDAR.2003.1227801 (cit. on p. 72).

[156] K. Skala, T. Lipić, I. Sović, and I. Grubišić. “Dynamic thermal models for human body dissipation”. In: *Periodicum Biologorum* 117.1 (2015), pp. 167–176 (cit. on p. 93).

[157] L. N. Smith. “Cyclical learning rates for training neural networks”. In: *Proceedings - 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017* April (2017), pp. 464–472. DOI: 10.1109/WACV.2017.58. eprint: 1506.01186 (cit. on p. 175).

[158] L. N. Smith and N. Topin. “Super-convergence: very fast training of neural networks using large learning rates”. In: *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*. Ed. by T. Pham. SPIE, May 2019, p. 36. DOI: 10.1117/12.2520589. eprint: 1708.07120 (cit. on p. 175).

[159] I. Sobel and G. Feldman. *An Isotropic 3x3 Image Gradient Operator*. 1968. DOI: 10.13140/RG.2.1.1912.4965 (cit. on p. 68).

[160] J. J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, B. van Ginneken, and B. van Ginneken. "Ridge based vessel segmentation in color images of the retina". In: *IEEE Transactions on Medical Imaging* 23.4 (2004), pp. 501–509 (cit. on pp. 53, 171).

[161] M. Švantner, V. Lang, J. Skála, T. Kohlschütter, M. Honner, L. Muzika, and E. Kosová. "Statistical Study on Human Temperature Measurement by Infrared Thermography". In: *Sensors* 22.21 (Nov. 2022), p. 8395. DOI: 10.3390/s22218395 (cit. on p. 45).

[162] A. Szurko, T. Kasprzyk-Kucewicz, A. Cholewka, M. Kazior, K. Sieroń, A. Stanek, and T. Morawiec. "Thermovision as a Tool for Athletes to Verify the Symmetry of Work of Individual Muscle Segments". In: *International Journal of Environmental Research and Public Health* 19.14 (2022). DOI: 10.3390/ijerph19148490 (cit. on p. 34).

[163] G. Tanda. "The use of infrared thermography to detect the skin temperature response to physical activity". In: *Journal of Physics: Conference Series* 655.1 (2015). DOI: 10.1088/1742-6596/655/1/012062 (cit. on p. 4).

[164] T. pandas development team. *pandas-dev/pandas: Pandas*. 2023. DOI: 10.5281/zenodo.8092754 (cit. on p. 122).

[165] R. L. S. Tempski. "Gesichtserkennung unter sportlicher Belastung im Wärmebildbereich". Bachelor thesis. Johannes Gutenberg University Mainz, 2023 (cit. on p. 184).

[166] M. Thiriet. *Anatomy and Physiology of the Circulatory and Ventilatory*. 2013, p. 140 (cit. on p. 9).

[167] J. Thomas, J. Crompton, and K. Koppenhoefer. "Multiphysics Analysis of Infra Red Bolometer". In: (2015), pp. 1–12 (cit. on p. 32).

[168] Y. Tian, Y. Zhang, and H. Zhang. "Recent Advances in Stochastic Gradient Descent in Deep Learning". In: *Mathematics* 11.3 (Jan. 2023), p. 682. DOI: 10.3390/math11030682 (cit. on pp. 79, 80).

[169] K. Tomita and M. Y. L. Chew. "A Review of Infrared Thermography for Delamination Detection on Infrastructures and Buildings". In: *Sensors* 22.2 (Jan. 2022), p. 423. DOI: 10.3390/s22020423 (cit. on p. 28).

[170] R. Tsai. "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses". In: *IEEE Journal on Robotics and Automation* 3.4 (Aug. 1987), pp. 323–344. DOI: 10.1109/JRA.1987.1087109 (cit. on p. 25).

[171] M. Unger, M. Markfort, D. Halama, and C. Chalopin. "Automatic detection of perforator vessels using infrared thermography in reconstructive surgery". In: *International Journal of Computer Assisted Radiology and Surgery* 14.3 (Mar. 2019), pp. 501–507. DOI: 10.1007/s11548-018-1892-6 (cit. on p. 60).

[172] R. Usamentiaga, P. Venegas, J. Guerediaga, L. Vega, J. Molleda, and F. G. Bulnes. “Infrared thermography for temperature measurement and non-destructive testing”. In: *Sensors (Switzerland)* 14.7 (2014), pp. 12305–12348. DOI: 10.3390/s140712305 (cit. on pp. 27–30, 33, 49).

[173] S. Vidas, R. Lakemond, S. Denman, C. Fookes, S. Sridharan, and T. Wark. “A mask-based approach for the geometric calibration of thermal-infrared cameras”. In: *IEEE Transactions on Instrumentation and Measurement* 61.6 (2012), pp. 1625–1635. DOI: 10.1109/TIM.2012.2182851 (cit. on p. 94).

[174] E. Villa and N. Arteaga-marrero. “Performance Assessment of Low-Cost Thermal”. In: *Sensors* 20.1321 (2020), pp. 1–16 (cit. on pp. 32, 60).

[175] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, et al. “SciPy 1.0: fundamental algorithms for scientific computing in Python”. In: *Nature Methods* 17.3 (Mar. 2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2 (cit. on p. 122).

[176] M. Vollmer and K.-P. Möllmann. *Infrared Thermal Imaging - Fundamentals, Research and Applications*. Second. Wiley-VCH, 2017 (cit. on pp. 4, 5, 27, 28, 30, 31, 33, 157, 163).

[177] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao. “YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors”. In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. July. Vancouver, BC, Canada: IEEE, June 2023, pp. 7464–7475. DOI: 10.1109/CVPR52729.2023.00721 (cit. on pp. 99, 159).

[178] M. Westoby, J. Brasington, N. Glasser, M. Hambrey, and J. Reynolds. “‘Structure-from-Motion’ photogrammetry: A low-cost, effective tool for geoscience applications”. In: *Geomorphology* 179 (Dec. 2012), pp. 300–314. DOI: 10.1016/j.geomorph.2012.08.021 (cit. on p. 34).

[179] N. Wingenfeld. “Thermal Image Segmentation”. Bachelor thesis. Johannes Gutenberg University Mainz, 2019 (cit. on p. 73).

[180] T. Wunsch and H. Schwameder. “Biomechanik des Laufens und Laufanalyse”. In: *Bewegung, Training, Leistung und Gesundheit* (2021), pp. 1–20. DOI: 10.1007/978-3-662-53386-4_11-1 (cit. on p. 118).

[181] H. Yousef, M. Alhajj, and S. Sharma. *Anatomy, Skin (Integument), Epidermis*. 2024 (cit. on p. 14).

[182] H. Yu and J. Sun. “Sweat detection theory and fluid driven methods: A review”. In: *Nanotechnology and Precision Engineering* 3.3 (2020), pp. 126–140. DOI: 10.1016/j.npe.2020.08.003 (cit. on p. 20).

[183] L. Yu, Y. Guo, H. Zhu, M. Luo, P. Han, and X. Ji. “Low-Cost Microbolometer Type Infrared Detectors”. In: *Micromachines* 11.9 (Aug. 2020), p. 800. DOI: 10.3390/mi11090800 (cit. on pp. 32, 60).

[184] J. Zhang, Y. Zhang, and X. Xu. “Pyramid U-Net for Retinal Vessel Segmentation”. In: *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2021, pp. 1125–1129. DOI: 10.1109/ICASSP39728.2021.9414164 (cit. on pp. 73, 74).

[185] Z. Zhang. “A flexible new technique for camera calibration”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.11 (2000), pp. 1330–1334. DOI: 10.1109/34.888718 (cit. on pp. 24, 25, 97).

[186] S. Zhao, Y. Wang, Z. Yang, and D. Cai. “Region mutual information loss for semantic segmentation”. In: *Advances in Neural Information Processing Systems* 32.1 (2019), pp. 1–11. arXiv: 1910.12037 (cit. on p. 78).

[187] T. Zhou, F. Porikli, D. J. Crandall, L. V. Gool, and W. Wang. “A Survey on Deep Learning Technique for Video Segmentation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022), pp. 1–20. DOI: 10.1109/TPAMI.2022.3225573. eprint: 2107.01153 (cit. on p. 175).

[188] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang. “Unet++: A nested u-net architecture for medical image segmentation”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 11045 LNCS. Springer Verlag, 2018, pp. 3–11. DOI: 10.1007/978-3-030-00889-5_1 (cit. on p. 73).

[189] J. Zhuang, T. Tang, Y. Ding, S. Tatikonda, N. Dvornek, X. Papademetris, and J. S. Duncan. “AdaBelief Optimizer: Adapting Stepsizes by the Belief in Observed Gradients”. In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle and M. Ranzato and R. Hadsell and M.F. Balcan and H. Lin. Vol. 33. NeurIPS. virtual: Curran Associates, Inc., 2020, pp. 18795–18806. arXiv: 2010.07468 (cit. on p. 80).

Webpages

[@1] ANT Wireless by Garmin Canada Inc. *ANT / ANT+ Defined - THIS IS ANT*. https://www.thisisant.com/developer/ant-plus/ant-antplus-defined/ (visited on May 23, 2024) (cit. on p. 40).

[@2] S. Bhowmick. *Imaging Sensor Technology And The Latest Developments*. 2022. https://www.electronicsforu.com/buyers-guides/imaging-sensor-technology-latest-developments (visited on Apr. 14, 2024) (cit. on pp. 32, 33).

[@3] BodyCAP. *e-Celsius Performance*. https://www.bodycap-medical.com/e-celsius-performance-body-temperature-monitoring-telemetric-pill/ (visited on Mar. 21, 2024) (cit. on p. 52).

[@4] Cleveland Clinic. *Heart Rate Zones Explained*. 2023. https://health.clevelandclinic.org/exercise-heart-rate-zones-explained (visited on Mar. 11, 2024) (cit. on p. 19).

[@5] Y. Collet. *LZ4 - Extremely fast compression*. 2011. https://lz4.org/ (visited on Jan. 29, 2024) (cit. on pp. 43, 110).

[@6] CORE. *Scientific Data Validating CORE's Accuracy*. https://corebodytemp.com/pages/scientific-data-validating-cores-accuracy (visited on Feb. 26, 2024) (cit. on p. 50).

[@7] Cosinuss GmbH. *°One - cosinuss°*. https://www.cosinuss.com/de/produkte/im-ohr-sensoren/one/ (visited on Mar. 21, 2024) (cit. on p. 50).

[@8] W. Falcon and T. P. L. Team. *PyTorch Lightning*. 2023. DOI: 10.5281/zenodo.8435466. https://doi.org/10.5281/zenodo.8435466 (visited on Nov. 1, 2023) (cit. on p. 61).

[@9] Geratherm Respiratory GmbH. *Ergostik - Geratherm Respiratory: innovative medical products*. https://www.geratherm-respiratory.com/product-groups/cpet/ergostik/ (visited on Jan. 28, 2024) (cit. on pp. 51, 110).

[@10] mesics GmbH. *LC Lactat – mesics GmbH*. https://www.mesics.de/lc-lactat/ (visited on Jan. 29, 2024) (cit. on p. 52).

[@11] Google LLC. *Protocol Buffers Documentation*. https://protobuf.dev/ (visited on Jan. 28, 2024) (cit. on p. 110).

[@12] Guangzhou Aosong Electronic Co. *AM2302 SIP Packaged Temperature and Humidity Sensor-Sensor-Temperature and Humidity-Guangzhou Aosong Electronic Co., Ltd.* http://www.aosong.com/en/products-22.html (visited on Mar. 21, 2024) (cit. on p. 51).

[@13] InfraTec GmbH. *Infrared cameras for use in electronics | InfraTec GmbH*. https://www.infratec.eu/thermography/industries-applications/electronics-electrical/ (visited on Apr. 19, 2024) (cit. on p. 28).

[@14] Laird Thermal Systems Inc. *TC-XX-PR-59 Temperature Controller | The World Leader in Thermal Management Solutions*. https://lairdthermal.com/de/products/product-temperature-controllers/tc-xx-pr-59-temperature-controller (visited on Jan. 20, 2024) (cit. on p. 126).

[@15] LEMO SA. *LEMO Connectors | LEMO Connectors and cables*. https://www.lemo.com/lemo-connectors (visited on Mar. 24, 2024) (cit. on p. 41).

[@16] B. E. Moore and J. J. Corso. *FiftyOne*. 2020. https://github.com/voxel51/fiftyone (visited on Mar. 7, 2024) (cit. on p. 65).

[@17] MSO Meßtechnik und Ortung GmbH. *Speed Wedge MKII Radar - MSO-Technik*. https://www.mso-technik.de/mso-produkte/geschwindigkeitsmessung/speed-wedge-radar.html (visited on Mar. 21, 2024) (cit. on p. 51).

[@18] OpenCV. *cv::SimpleBlobDetector Class Reference*. 2023. https://docs.opencv.org/4.8.0/d0/d7a/classcv_1_1SimpleBlobDetector.html (visited on Jan. 6, 2024) (cit. on pp. 25, 96).

[@19] OpenCV. *OpenCV: Camera Calibration and 3D Reconstruction*. 2023. https://docs.opencv.org/4.9.0/d9/d0c/group__calib3d.html (visited on Mar. 3, 2024) (cit. on pp. 92, 96).

[@20] Oracle. *Oracle VM VirtualBox*. http://www.virtualbox.org/ (visited on Jan. 28, 2024) (cit. on p. 112).

[@21] Pema Thermo Group SL. *ThermoHuman | Software for automatic human thermography analysis*. https://thermohuman.com/software/ (visited on Mar. 27, 2024) (cit. on pp. 4, 164, 165).

[@22] Plotly Technologies Inc. *Plotly: Low-Code Data App Development*. https://plotly.com/ (visited on Mar. 30, 2024) (cit. on p. 122).

[@23] Polar Electro. *Polar H10 | Polar UK*. https://www.polar.com/uk-en/products/accessories/polar_h10_heart_rate_sensor (visited on Mar. 21, 2021) (cit. on p. 50).

[@24] Scikit-learn. *Visualizing cross-validation behavior in scikit-learn — scikit-learn 1.3.2 documentation*. https://scikit-learn.org/stable/auto_examples/model_selection/plot_cv_indices.html (visited on Mar. 12, 2024) (cit. on p. 61).

[@25] Shenzen Everbrest Machinery Industry co. LTD (Brand: CEM). *BX-350/500/BXC-15*. https://www.cem-instruments.com/en/product-id-884 (visited on Jan. 23, 2024) (cit. on p. 160).

[@26] L. Smith. *WD My Book Duo 20TB Review*. 2017. https://www.storagereview.com/review/wd-my-book-duo-20tb-review (visited on Mar. 24, 2024) (cit. on p. 52).

[@27] Teledyne FLIR. *Teledyne FLIR ADAS Dataset*. 2018. https://www.flir.com/oem/adas/adas-dataset-form/ (visited on Jan. 6, 2024) (cit. on pp. 28, 53).

[@28] The Linux Foundation. *ONNX | Home*. https://onnx.ai/ (visited on Mar. 13, 2024) (cit. on p. 83).

[@29] K. Wada. *Labelme: Image Polygonal Annotation with Python*. 2021. DOI: 10.5281/zenodo.5711226. https://github.com/labelmeai/labelme (visited on Mar. 7, 2024) (cit. on p. 65).

List of Figures

List of Tables

List of Algorithms

List of Acronyms

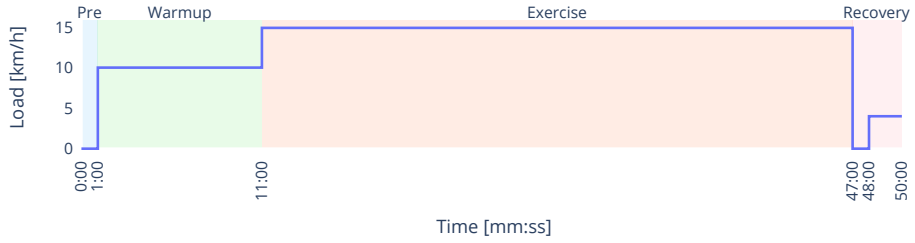
ANS autonomic nervous system
BPN body part network
CC connected component
CE cross entropy loss
CPET cardiopulmonary exercise testing
CV cross-validation
DNN deep neural network
FOV field of view
FPA focal plane array
GAN generative adversarial network
HPO hyperparameter optimization
IAT individual anaerobic threshold
IFOV instantaneous field of view
IRT infrared thermography
IoU intersection over union
MLSS maximum lactate steady state
NETD noise equivalent temperature difference
NUC nonuniform calibration
PAD peripheral artery disease
PAT PixelAnnotationTool
RGBD color+depth
ROI region of interest
RPE rate of perceived exertion
SGD stochastic gradient descent
SD standard deviation
ToF time-of-flight
T$_{sr}$ surface radiation temperature
VIS visible light
VN vessel network
VT ventilatory threshold

Appendix

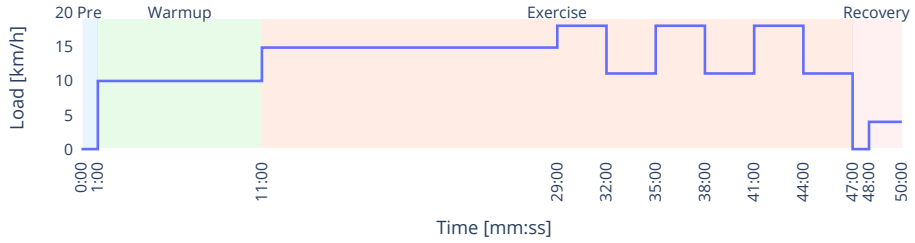
Definitions and Fields

A

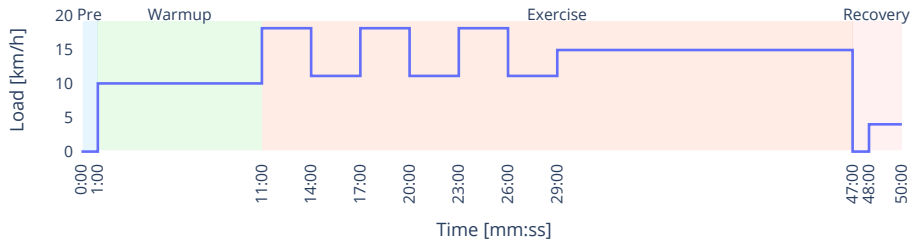
A.1 Incoreloop Exercise Protocols



(a) Incoreloop protocol T1: steady load.



(b) Incoreloop protocol T2: steady and alternating load.



(c) Incoreloop protocol T3: alternating and steady load.

Fig. A.1.: Incoreloop test protocols for long runs without pauses. Speeds are adjusted to the participant's performance in T0, a standard step walking protocol (figure 5.1, chapter 5)

A.2 Segmentation Class Definitions

Name	Side	ID	Color (RGB)	Color (RGB)		Weight
Background	-	10	255,255,255	#FFFFFF		1.376876
Clothes	left	9	106,61,154	#6A3D9A		67.173098
Clothes	right	109	154,106,61	#9A6A3D		68.967985
Knee	left	7	255,127,0	#FF7F00		117.829826
Knee	right	107	0,127,255	#007FFF		115.135736
Leg lower	left	6	253,191,111	#FDBF6F		13.114468
Leg lower	right	106	111,191,253	#6FBFFD		13.144193
Leg upper	left	5	227,26,28	#E31A1C		50.231477
Leg upper	right	105	28,26,227	#1C1AE3		54.054786
Other body	-	12	36,31,49	#241F31		20825.231104
Shoes	left	13	255,0,255	#FF00FF		62.361923
Shoes	right	113	154,153,150	#9A9996		49.170006

Tab. A.1.: Classes and colors for body part segmentation masks based on [7] (see section 6.1). In addition, body class weights for weighted loss functions to handle class imbalances in the dataset are given (see section 6.2.4).

Name	Color (RGB)	Color (RGB)	
Background	255,255,255	#FFFFFF	
Clothes	130,84,108	#82546C	
Skin	93,213,93	#5dd55d	

Tab. A.2.: Classes and colors for body part segmentation masks for body part network (BPN) defined in [60].

Name	ID	Color (RGB)	Color (RGB)		Weight
Background	11	36,31,49	#241f31		1.255996
Non-vessel	4	192,191,188	#C0BFBC		5.213695
Perforator	2	198,70,0	#C64600		243.946706
Tendon	3	249,240,107	#F9F06B		-
Vein	1	26,95,180	#1A5FB4		126.301348

Tab. A.3.: Classes and colors for blood vessel segmentation masks based on [60] (see section 6.1). In addition, class weights for weighted loss functions to handle class imbalances in the dataset are given (see section 6.2.4).

A.3 Data Fields

Value	Unit/Type	Source	Comment
user_id	string		pseudonymized user ID
test_id	string		experiment test id from study design
timestamp	ISO 8601 timestamp	Acquisition system	Global timestamp in UTC with microseconds precision
time relative	ISO 8601 timestamp	Acquisition system, BlueCherry Observer	Relative time to the begin of the experiment
BlueCherry state	categorical	BlueCherry Observer	Phase ∈ [Ruhe, Warmup, Last, Erholung]
Zeit	s	BlueCherry	Relative time from the begin of the experiment
Velocity	km/h	BlueCherry	Protocol: Treadmill speed
Incline	%	BlueCherry	Protocol: Treadmill inclination
VO_2	l/min	BlueCherry	oxygen intake
VO_2/kg	ml/min/kg	BlueCherry	normalized oxygen intake
VCO_2	l/min	BlueCherry	carbon dioxide exhale
RER	-	BlueCherry	respiratory exchange ratio: $\frac{VO_2}{VCO_2}$
Bf	1/min	BlueCherry	breathing frequency
VE	l	BlueCherry	ventilation
VE/VO_2	l	BlueCherry	oxygen breath equivalent
VE/VCO_2	%	BlueCherry	CO_2 breath equivalent
Height	cm	BlueCherry	Participant's body height
Weight	kg	BlueCherry	Participant's body weight
Age	years	BlueCherry	Participant's age
$VO_{2,\max}$	-	BlueCherry	$VO_{2,\max}$
VT_1	time	BlueCherry	time point of VT_1
VT_2	time	BlueCherry	time point of VT_2
Zeit no dup	timestamp	BlueCherry	relative timestamp with no duplicates
Stage		Protocol	
Speed const	km/h	Protocol	
Speed virtual	km/h	calculated	Linear increase within a stage
Training zone	categorical	calculated	
Pause	categorical	calculated	pause, moving, beginning of moving, end of moving
IAT (speed)	km/h	Lactate	The speed at which the individual anaerobic threshold (IAT) occurs (relative to the virtual speed)
Heart rate	Hz	Polar H10	
Heart rate	Hz	Cosinuss° One	Two sensors
Core temperature	° C	Cosinuss° One	Two sensors

Value	Unit/Type	Source	Comment
Core temperature	° C	CORE	Three sensors
Skin temperature	° C	CORE	Three sensors
Core temperature	° C	Pill	
Room temperature	° C	Environmental sensor	
Room humidity	relative in %	Environmental sensor	

Tab. A.4.: Subset of all merged sensor data fields. It does not include thermal features, some specific BlueCherry fields, and other sensor data not used in this thesis.

B Results

B.1 Hyperparameter Optimization

B.1.1 Body Part Network

#	State	Val IoU	Batch Size	Loss	Learning Rate	Model Architecture	Optimizer	Best Epoch
0	c	0.8128	8	dice	0.0001	attention-unet	adabelief	40
1	c	0.7077	8	dice	0.001	deeplabv3plus-imagenet	adabelief	4
2	p		8	dice-perimeter	7.4e-05	attention-unet-supervision-pyramid	adamw	
3	p	0.0771	4	boundary	0.003657	unet	adabelief	
4	c	0.7953	4	tversky-02-08	0.000166	deeplabv3plus-imagenet	adamw	36
5	p	0.3862	4	tversky-08-02	0.012088	attention-unet-supervision	rmsprop	
6	c	0.5111	4	soft-dice-cldice	6.7e-05	unet	adam	50
7	p	0.5167	4	dicefocal	0.024542	attention-unet-supervision-pyramid	adam	
8	c	0.6639	1	dice-perimeter	0.007008	attention-unet-supervision	adabelief	36
9	p	0.6372	2	tversky-08-02	3.8e-05	unet	rmsprop	
10	p	0.0707	6	wdice	1.1e-05	attention-unet	sgd	
11	p	0.6450	8	tversky-02-08	0.000359	attention-unet	adamw	
12	p		1	boundary-wdice	0.091289	deeplabv3plus	adamw	
13	p	0.0092	6	boundary-dice	0.000247	deeplabv3plus-imagenet	sgd	
14	p	0.5136	2	dicece	0.000292	tiramisu103	adamw	
15	c	0.7854	8	tversky-02-08	0.000811	deeplabv3plus-imagenet	adabelief	23

#	State	Val IoU	Batch Size	Loss	Learning Rate	Model Architecture	Optimizer	Best Epoch
16	p	0.6405	4	ce-weighted	0.000155	attention-unet	adamw	
17	p		8	boundary-wdice-05	1.9e-05	tiramisu103	adabelief	
18	c	0.7990	2	focal	9.3e-05	deeplabv3plus	adam	78
19	p	0.6523	2	focal	3.8e-05	deeplabv3plus	adam	
20	p	0.4108	2	ce	1e-05	deeplabv3plus	adam	
21	c	0.8074	2	dice	0.000125	attention-unet	adam	50
22	c	0.7867	2	dice	9.2e-05	attention-unet	adam	25
23	p	0.6102	2	rmi	0.000425	attention-unet	adam	
24	p	0.7375	2	focal	0.000141	deeplabv3plus	adam	
25	p	0.6690	2	dice	3.8e-05	attention-unet	adam	
26	p	0.1797	1	rmi-dice	0.002153	attention-unet	sgd	
27	c	0.8179	6	dice	0.000464	deeplabv3plus	rmsprop	33
28	p	0.6590	6	dice	0.000652	attention-unet	rmsprop	
29	p		6	dice	0.001444	attention-unet-supervision-pyramid	rmsprop	
30	p		6	tanimoto	0.000602	attention-unet-supervision	rmsprop	
31	c	0.8258	8	dice	0.000238	deeplabv3plus	adabelief	39
32	c	0.7879	8	dice	0.000233	deeplabv3plus	adabelief	12
33	p	0.6146	8	lovasz	0.000442	deeplabv3plus	adabelief	
34	c	0.8041	8	dice	0.001253	deeplabv3plus	adabelief	37
35	p		8	boundary	0.000159	tiramisu103	adabelief	
36	p		8	dice	0.000847	attention-unet-supervision-pyramid	rmsprop	
37	c	0.8127	6	dice	0.00024	unet	adabelief	46
38	p	0.7374	6	soft-dice-cldice	0.000275	unet	adabelief	
39	p	0.6182	6	dicefocal	6.2e-05	unet	adabelief	
40	p	0.3904	6	wdice	0.000482	unet	adabelief	
41	p	0.7761	6	dice	0.000113	unet	adabelief	
42	p		8	dice	0.000212	attention-unet-supervision	rmsprop	
43	p	0.4122	1	dice-perimeter	0.000175	attention-unet	adabelief	
44	p	0.7827	6	dice	6.8e-05	deeplabv3plus-imagenet	adabelief	
45	p	0.7586	4	boundary-wdice	0.000389	deeplabv3plus	rmsprop	
46	c	0.0115	8	dicece	0.000296	unet	sgd	1
47	p	0.7683	6	boundary-dice	0.000134	attention-unet	adabelief	

#	State	Val IoU	Batch Size	Loss	Learning Rate	Model Architecture	Optimizer	Best Epoch
48	p		8	ce-weighted	0.000196	attention-unet-supervision-pyramid	adabelief	
49	p	0.7665	4	tversky-08-02	5.4e-05	deeplabv3plus	adamw	
50	p	0.1240	1	rmi-dice	9.7e-05	tiramisu103	sgd	
51	p	0.3945	8	dice	0.001472	deeplabv3plus	adabelief	
52	c	0.8348	8	dice	0.000603	deeplabv3plus	adabelief	59
53	c	0.8279	8	dice	0.000345	deeplabv3plus	adabelief	45
54	c	0.8268	8	boundary-wdice-05	0.000675	deeplabv3plus	adabelief	62
55	p	0.7659	8	ce	0.000673	deeplabv3plus	adabelief	
56	c	0.8083	8	boundary-wdice-05	0.000913	deeplabv3plus	adabelief	30
57	c	0.8290	8	boundary-wdice-05	0.000327	deeplabv3plus	adabelief	47
58	c	0.8216	8	boundary-wdice-05	0.000513	deeplabv3plus	adabelief	44
59	c	0.8142	8	boundary-wdice-05	0.000307	deeplabv3plus	adabelief	30
60	c	0.8032	8	boundary-wdice-05	0.000594	deeplabv3plus	adabelief	18
61	c	0.8013	8	boundary-wdice-05	0.000373	deeplabv3plus	adabelief	17
62	c	0.8209	8	boundary-wdice-05	0.000511	deeplabv3plus	adabelief	45
63	c	0.8072	8	boundary-wdice-05	0.000826	deeplabv3plus	adabelief	30
64	c	0.8158	8	boundary-wdice-05	0.000521	deeplabv3plus	adabelief	34
65	c	0.8177	8	boundary-wdice-05	0.000337	deeplabv3plus	adabelief	38
66	c	0.7823	8	boundary-wdice-05	0.001068	deeplabv3plus	adabelief	15
67	c	0.8265	8	rmi	0.000658	deeplabv3plus	adabelief	28
68	**c**	**0.8439**	**8**	**rmi**	**0.000686**	**deeplabv3plus**	**adamw**	**47**
69	p	0.8066	8	rmi	0.000707	deeplabv3plus-imagenet	adamw	
70	p	0.7275	8	rmi	0.001784	deeplabv3plus	adamw	
71	c	0.8343	8	rmi	0.001032	deeplabv3plus	adamw	38
72	c	0.8408	8	rmi	0.000927	deeplabv3plus	adamw	63
73	c	0.8337	8	rmi	0.001077	deeplabv3plus	adamw	43
74	c	0.8349	8	rmi	0.001014	deeplabv3plus	adamw	30
75	p	0.8063	8	rmi	0.001166	deeplabv3plus	adamw	
76	p		8	rmi	0.00241	attention-unet-supervision	adamw	

#	State	Val IoU	Batch Size	Loss	Learning Rate	Model Architecture	Optimizer	Best Epoch
77	c	0.8314	8	rmi	0.000884	deeplabv3plus	adamw	30
79	c	0.8366	8	rmi	0.001008	deeplabv3plus	adamw	56
80	p	0.7297	8	rmi	0.000995	deeplabv3plus	adamw	
81	p		4	rmi	0.00261	tiramisu103	adamw	
82	p	0.7996	8	rmi	0.001353	deeplabv3plus	adamw	
83	p	0.8078	8	rmi	0.001025	deeplabv3plus	adamw	
84	p	0.7223	8	rmi	0.001533	deeplabv3plus	adamw	
85	p		1	tanimoto	0.000828	deeplabv3plus	adamw	
86	p	0.7802	8	tversky-02-08	0.001881	deeplabv3plus	adamw	
87	p		8	lovasz	0.001156	attention-unet-supervision-pyramid	adamw	
88	p	0.7963	8	rmi	0.003432	deeplabv3plus-imagenet	adamw	
89	c	0.8369	8	rmi	0.000849	deeplabv3plus	adamw	49
90	p		8	rmi	0.000853	attention-unet-supervision	adamw	
91	p	0.2906	8	boundary	0.001659	deeplabv3plus	adamw	
92	c	0.8285	8	rmi	0.00106	deeplabv3plus	adamw	37
93	p	0.6769	8	dicefocal	0.001388	deeplabv3plus	adamw	
94	p	0.7866	8	soft-dice-cldice	0.000452	deeplabv3plus	adamw	
95	c	0.8318	8	rmi	0.000786	deeplabv3plus	adamw	28
96	p	0.6487	8	rmi	0.000802	deeplabv3plus	adamw	
97	c	0.8210	4	rmi	0.00059	deeplabv3plus	adamw	29
98	p	0.5961	2	dice-perimeter	0.001256	deeplabv3plus	adamw	
99	p	0.3165	1	dicece	0.002009	tiramisu103	adamw	
100	p	0.5743	8	wdice	0.000741	deeplabv3plus	adamw	
101	p		8	boundary-dice	0.000962	attention-unet-supervision-pyramid	adamw	
102	c	0.8316	8	rmi	0.000592	deeplabv3plus	adamw	34
103	c	0.8365	8	rmi	0.001253	deeplabv3plus	adamw	43
104	c	0.8327	8	rmi	0.000593	deeplabv3plus	adamw	30
105	p	0.1807	8	focal	0.00128	deeplabv3plus	sgd	
106	p	0.7306	8	tversky-08-02	0.00159	deeplabv3plus	adamw	
107	p	0.5348	8	ce-weighted	0.000407	deeplabv3plus	adamw	
108	p	0.6357	8	rmi	0.00074	deeplabv3plus	adamw	
109	p	0.8072	8	rmi	0.001128	deeplabv3plus	adamw	
110	p	0.7173	8	boundary-wdice	0.000574	deeplabv3plus-imagenet	adamw	
111	p	0.7207	8	ce	0.000443	deeplabv3plus	adam	

#	State	Val IoU	Batch Size	Loss	Learning Rate	Model Architecture	Optimizer	Best Epoch
112	c	0.8392	8	rmi	0.000619	deeplabv3plus	adamw	41
113	p	0.8112	8	rmi	0.000728	deeplabv3plus	adamw	
114	c	0.8323	8	rmi-dice	0.000911	deeplabv3plus	adamw	33

Tab. B.1.: Hyperparameter search results for BPN. States: c=completed, p=pruned. The validation intersection over union (IoU) for completed studies is the best IoU, for pruned studies it is either the last IoU or none if the study is pruned due to constraints such as an invalid model batch size combination.

B.1.2 Vessel Network

Number	State	Val IoU	Batch Size	Loss	Learning Rate	Model	Optimizer	Best Epoch
0	c	0.6579	8	dice	0.0001	attention-unet	adabelief	34
1	c	0.616	8	dice	0.001	deeplabv3plus-imagenet	adabelief	29
2	p	0.4797	1	lovasz	0.084284	attention-unet-supervision	adabelief	
3	p	0.0	8	wdice	0.000134	attention-unet-supervision-pyramid	adam	
4	c	0.5045	4	rmi	0.061903	attention-unet	rmsprop	12
5	c	0.4787	1	dice-perimeter	0.000538	unet	rmsprop	22
6	p	0.4334	4	focal	0.000108	attention-unet-supervision-pyramid	sgd	
7	c	0.6503	2	soft-dice-cldice	0.000266	attention-unet-supervision-pyramid	adabelief	36
8	p	0.0	4	dicece	0.000145	tiramisu103	adam	
9	p	0.5453	1	dicece	9.6e-05	tiramisu103	adam	
10	p	0.537	6	tanimoto	1.1e-05	attention-unet	adamw	
11	c	0.5211	2	soft-dice-cldice	0.004764	deeplabv3plus	adabelief	8
12	p	0.4996	2	ce	1.4e-05	attention-unet	adabelief	
13	p	0.0	8	rmi-dice	0.002291	attention-unet-supervision-pyramid	adabelief	
14	p	0.0036	2	tversky-02-08	3.6e-05	deeplabv3plus-imagenet	sgd	
15	p	0.6108	6	dice	0.000398	deeplabv3plus	adamw	
16	c	0.641	2	boundary-wdice-05	3.3e-05	unet	adabelief	22

Number	State	Val IoU	Batch Size	Loss	Learning Rate	Model	Optimizer	Best Epoch
17	p	0.0	8	boundary-wdice	0.000367	attention-unet-supervision	adabelief	
18	p	0.5041	8	ce-weighted	0.003079	attention-unet	adabelief	
19	p	0.3019	2	boundary-dice	3.8e-05	attention-unet-supervision-pyramid	sgd	
20	p	0.0	6	boundary	0.000216	attention-unet-supervision-pyramid	adamw	
21	p	0.6284	2	boundary-wdice-05	4.2e-05	unet	adabelief	
22	p	0.6215	2	boundary-wdice-05	4.7e-05	unet	adabelief	
23	c	0.628	2	tversky-08-02	2e-05	unet	adabelief	14
24	p	0.6177	2	soft-dice-cldice	6.9e-05	unet	adabelief	
25	c	0.6405	8	dicefocal	2.2e-05	attention-unet	rmsprop	44
26	p	0.6195	2	boundary-wdice-05	0.000235	attention-unet-supervision	adabelief	
27	p	0.54	2	soft-dice-cldice	8.1e-05	deeplabv3plus	adabelief	
28	p	0.0	8	dice	2.3e-05	tiramisu103	adabelief	
29	p	0.5854	8	dice	0.000691	deeplabv3plus-imagenet	adabelief	
30	p	0.0	1	tversky-08-02	0.00022	deeplabv3plus-imagenet	adamw	
31	p	0.5765	8	dicefocal	1.8e-05	attention-unet	rmsprop	
32	p	0.4984	8	dicefocal	1e-05	attention-unet	rmsprop	
33	p	0.4792	8	lovasz	6.2e-05	attention-unet	rmsprop	
34	c	0.6462	8	dicefocal	2.9e-05	attention-unet	rmsprop	40
35	p	0.5383	4	rmi	3e-05	attention-unet-supervision-pyramid	adam	
36	p	0.3901	8	wdice	5.4e-05	unet	rmsprop	
37	p	0.4691	1	focal	0.000103	attention-unet	sgd	
38	p	0.5033	4	dice-perimeter	7.3e-05	attention-unet-supervision-pyramid	rmsprop	
39	c	0.6539	6	tanimoto	0.000141	attention-unet	adam	38
40	c	0.6518	6	tanimoto	0.000139	attention-unet	adam	48
41	p	0.498	6	tanimoto	0.000137	attention-unet	adam	
42	p	0.5155	6	tanimoto	0.000181	attention-unet	adam	
43	p	0.6132	6	tanimoto	0.000123	attention-unet	adam	
44	p	0.4804	6	tanimoto	0.000309	attention-unet	adam	

Number	State	Val IoU	Batch Size	Loss	Learning Rate	Model	Optimizer	Best Epoch
45	p	0.5647	6	rmi-dice	0.000151	attention-unet	adam	
46	p	0.0	6	ce	9.8e-05	tiramisu103	adam	
47	p	0.6188	6	tversky-02-08	0.000517	attention-unet	adam	
48	c	0.6206	1	dicece	0.000291	attention-unet-supervision	adam	11
49	p	0.5564	6	soft-dice-cldice	0.000935	attention-unet	rmsprop	
50	c	0.2615	4	boundary-wdice	0.000156	attention-unet-supervision-pyramid	sgd	12
51	p	0.5491	2	boundary-dice	5.2e-05	deeplabv3plus	adabelief	
52	p	0.5672	8	ce-weighted	3.5e-05	unet	adabelief	
53	c	0.64	2	dice	8.4e-05	attention-unet	adamw	20
54	c	0.6436	2	tanimoto	2.9e-05	attention-unet-supervision-pyramid	adabelief	47
55	p	0.0	6	tanimoto	1.5e-05	attention-unet-supervision-pyramid	adabelief	
56	p	0.0	8	boundary	0.000106	attention-unet-supervision-pyramid	adam	
57	p	0.626	2	tanimoto	2.9e-05	attention-unet-supervision-pyramid	adabelief	
58	p	0.3439	2	tanimoto	4.3e-05	attention-unet-supervision-pyramid	sgd	
59	p	0.0	6	lovasz	6.2e-05	tiramisu103	adabelief	
60	p	0.5509	8	dicefocal	0.000187	deeplabv3plus-imagenet	adam	
61	p	0.5752	2	boundary-wdice-05	2.7e-05	unet	adabelief	
62	p	0.2624	2	wdice	1.6e-05	attention-unet-supervision	adabelief	
63	c	0.6427	2	soft-dice-cldice	4.1e-05	attention-unet	adabelief	40
64	c	0.5527	2	rmi	7.5e-05	attention-unet	adabelief	4
65	c	0.6446	2	soft-dice-cldice	4.7e-05	attention-unet	adabelief	32
66	c	0.6474	1	soft-dice-cldice	0.000121	attention-unet	adamw	51

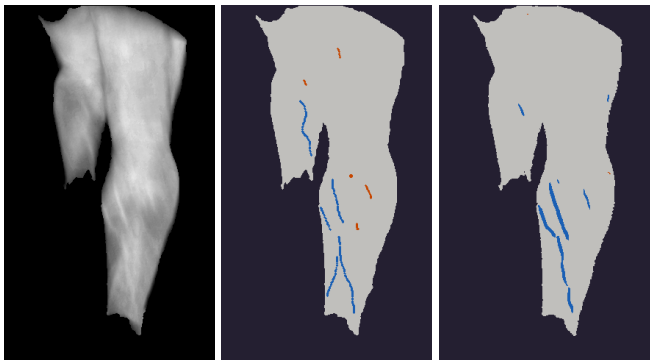
Number	State	Val IoU	Batch Size	Loss	Learning Rate	Model	Optimizer	Best Epoch
67	p	0.627	1	soft-dice-cldice	0.000139	attention-unet	adamw	
68	p	0.6243	1	soft-dice-cldice	0.000225	attention-unet	adamw	
69	p	0.6277	1	soft-dice-cldice	0.000115	attention-unet	adamw	
70	p	0.4857	1	dice-perimeter	0.000411	attention-unet	adamw	
71	p	0.5463	8	dice	5.4e-05	deeplabv3plus	rmsprop	
72	p	0.5097	2	focal	8.5e-05	attention-unet	adabelief	
73	p	0.6304	4	soft-dice-cldice	2.3e-05	attention-unet-supervision-pyramid	adabelief	
74	p	0.63	8	dicefocal	4.7e-05	attention-unet	adamw	
75	c	0.6329	2	tversky-08-02	6.7e-05	attention-unet	rmsprop	24
76	p	0.0	1	tanimoto	9.8e-05	deeplabv3plus-imagenet	adabelief	
77	c	0.6182	6	rmi-dice	0.00018	attention-unet	adam	18
78	p	0.0	8	dicece	3.3e-05	attention-unet-supervision-pyramid	adabelief	
79	c	0.5762	2	ce	0.000122	attention-unet	rmsprop	12
80	p	0.0	6	boundary-wdice	6.1e-05	tiramisu103	sgd	
81	c	0.6454	2	soft-dice-cldice	3.6e-05	attention-unet	adabelief	32
83	p	0.6285	2	soft-dice-cldice	4.1e-05	attention-unet	adabelief	
84	p	0.579	2	soft-dice-cldice	2.6e-05	attention-unet	adabelief	
85	p	0.5477	2	ce-weighted	3.5e-05	attention-unet	adabelief	
86	c	0.6398	2	boundary-dice	1.8e-05	attention-unet-supervision	adam	22
87	p	0.5545	8	tversky-02-08	7.9e-05	attention-unet	adabelief	
88	**c**	**0.666**	**4**	**tanimoto**	**5.2e-05**	**attention-unet**	**adabelief**	**64**
89	c	0.6619	4	soft-dice-cldice	0.000278	attention-unet	adamw	35
90	p	0.6282	4	dice	0.000253	attention-unet	adamw	
91	p	0.5714	4	tanimoto	0.000169	attention-unet	adamw	
92	p	0.5242	4	soft-dice-cldice	0.000131	attention-unet	adamw	

Number	State	Val IoU	Batch Size	Loss	Learning Rate	Model	Optimizer	Best Epoch
93	p	0.4108	4	soft-dice-cldice	9.6e-05	attention-unet	adamw	
94	p	0.6031	4	soft-dice-cldice	5.3e-05	attention-unet	adam	
95	c	0.4796	4	boundary	0.000292	attention-unet	rmsprop	18
96	p	0.6324	4	dicefocal	0.000196	attention-unet	adabelief	
97	p	0.5932	6	soft-dice-cldice	0.000153	deeplabv3plus	adam	
98	p	0.6234	4	tanimoto	7.4e-05	attention-unet	adamw	
99	p	0.5094	1	rmi	0.000247	attention-unet	adabelief	
100	p	0.0361	6	wdice	0.000114	attention-unet	sgd	

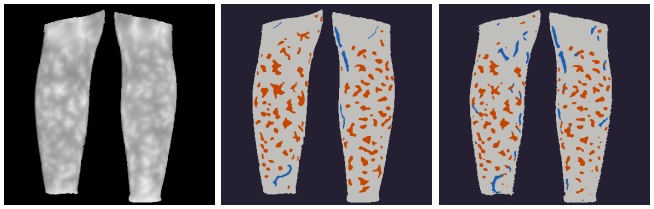
Tab. B.2.: Hyperparameter search results for vessel network (VN). States: c=completed, p=pruned. The validation IoU for completed studies is the best IoU, for pruned studies it is either the last IoU or none if the study is pruned due to constraints such as an invalid model batch size combination.

B.2 Applied Segmentation

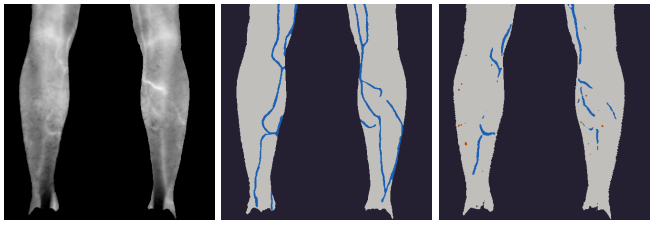
B.2.1 Vessel Network Predictions



(a) Thermogram. (b) Label. (c) Network V-B prediction.



(d) Thermogram. (e) Label. (f) Network V-B prediction.



(g) Thermogram. (h) Label. (i) Network V-B prediction.

Fig. B.1.: Three additional example results for the VN from the network V-B (see section 9.3.2) compared to the ground truth and the original image.
IoU values for (c): mean: 0.5434, background: 0.9980, vein: 0.2202, perforator: 0, non-vessel: 0.9554.
IoU values for (f): mean: 0.6794, background: 0.9968, vein: 0.2308, perforator: 0.5839, non-vessel: 0.9061.
IoU values for (i): mean: 0.5381, background: 0.9917, vein: 0.2614, perforator: 0, non-vessel: 0.8992.

B.2.2 Thermal Features Examples

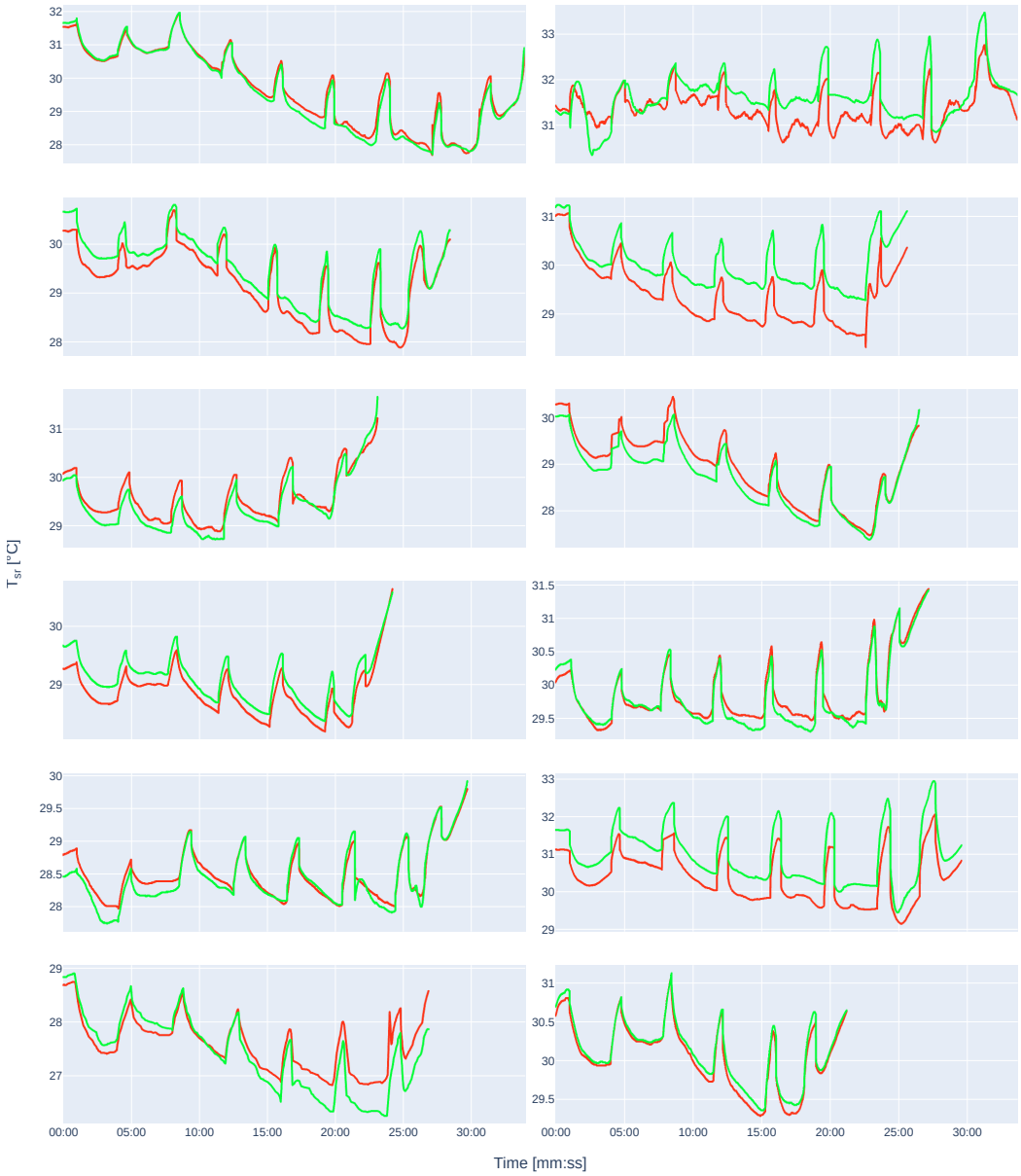


Fig. B.2.: Incoreloop overview of the smoothed T_{sr} of the left (red) and right (green) calves at the T0 experiment (see section 9.5). Each plot represents a different participant.

B.3 Stereo Label Transformation

B.3.1 Calibration

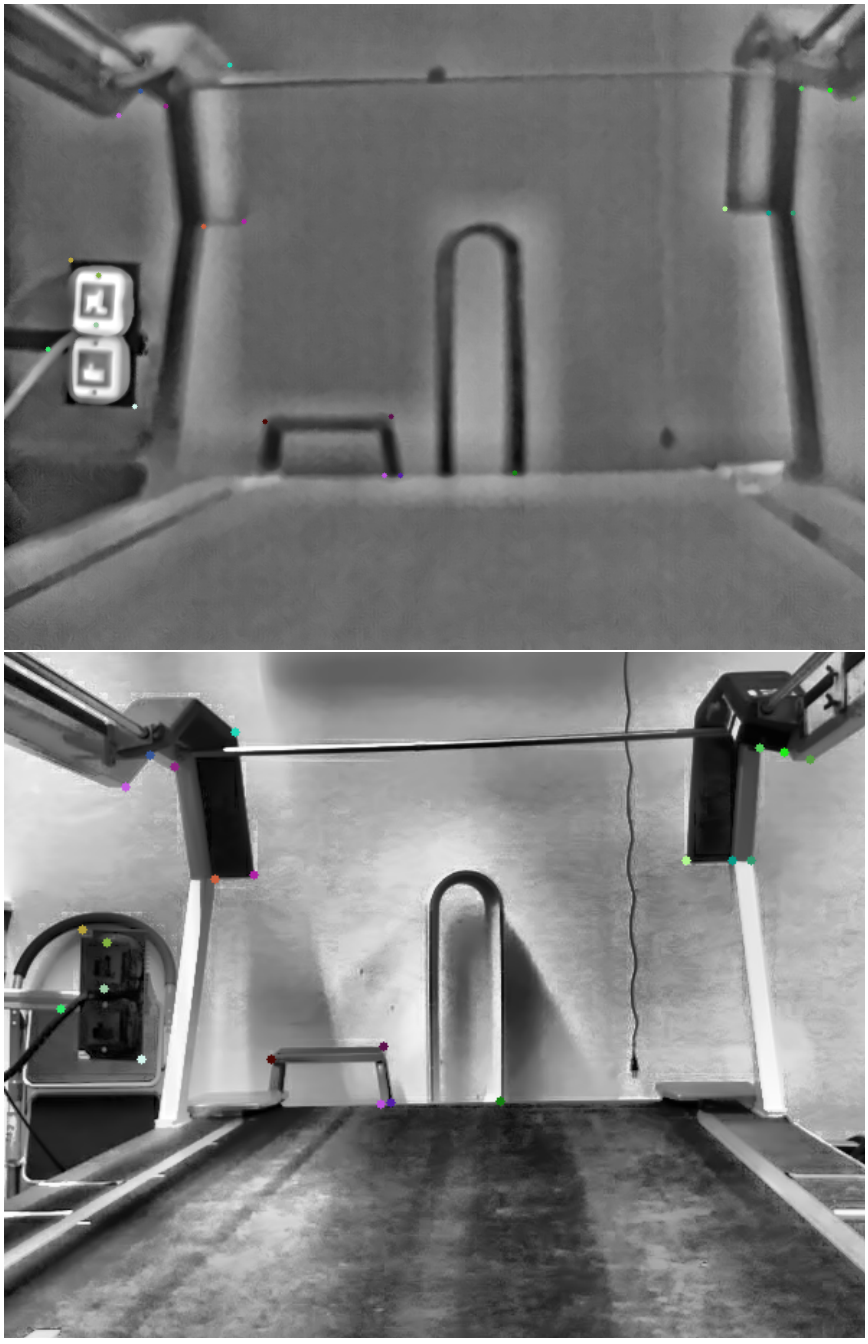
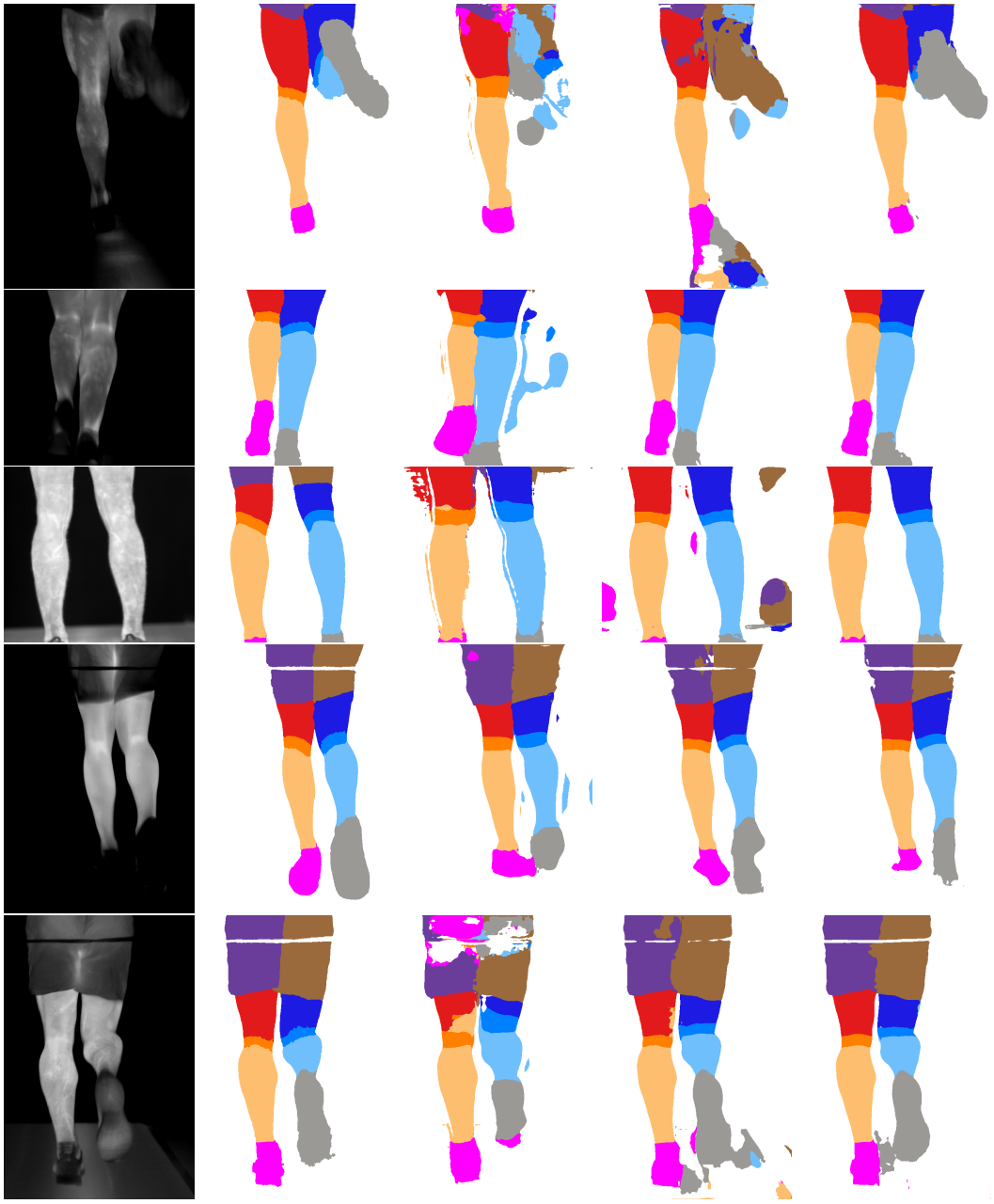


Fig. B.3.: Manually clicked points for stereo extrinsics correction (see section 7.1.3). Top: Feature enhanced thermogram. Bottom: Feature enhanced visual image to match a similar field of view (FOV) as the thermogram.

B.3.2 Body Part Network Results on Manual Test Set



(a) Thermogram. (b) Manually annotated label. (c) Prediction BPN S. (d) Prediction BPN B-B. (e) Prediction BPN S-100.

Fig. B.4.: Thermograms (five people from the manually annotated test set) along with their manually annotated label (b) and the result of the BPN S (c), the result of the BPN B-B (d) and the fine-tuned network S-100 (e) in different stages. See section 9.4.3.

B.4 Analysis Dashboard

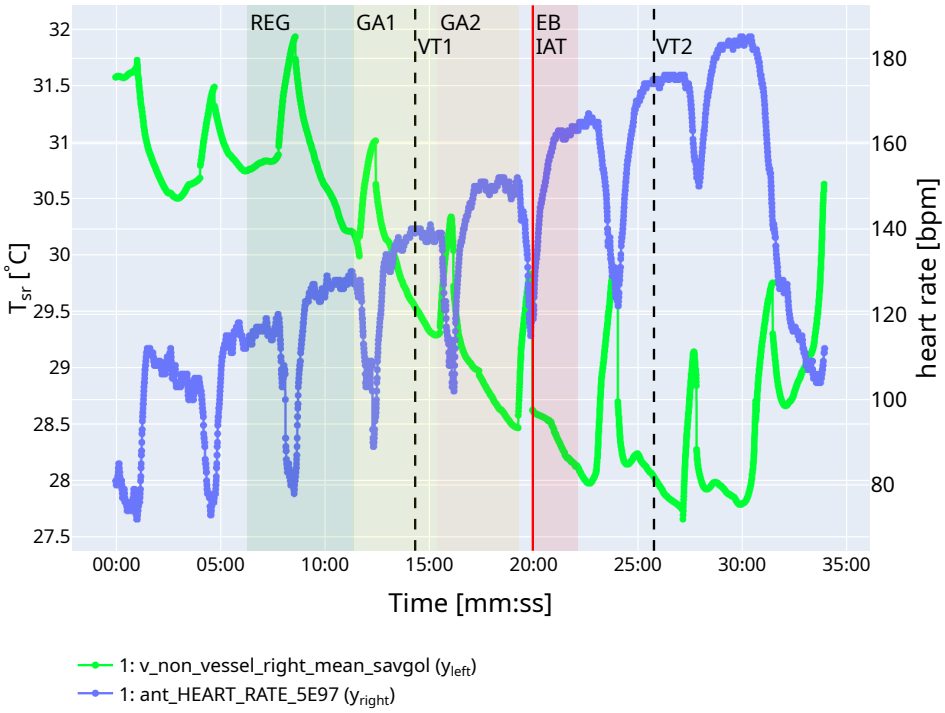


Fig. B.5.: Heartrate and thermal radiation T_{sr} (bigger version of figure 9.22a).

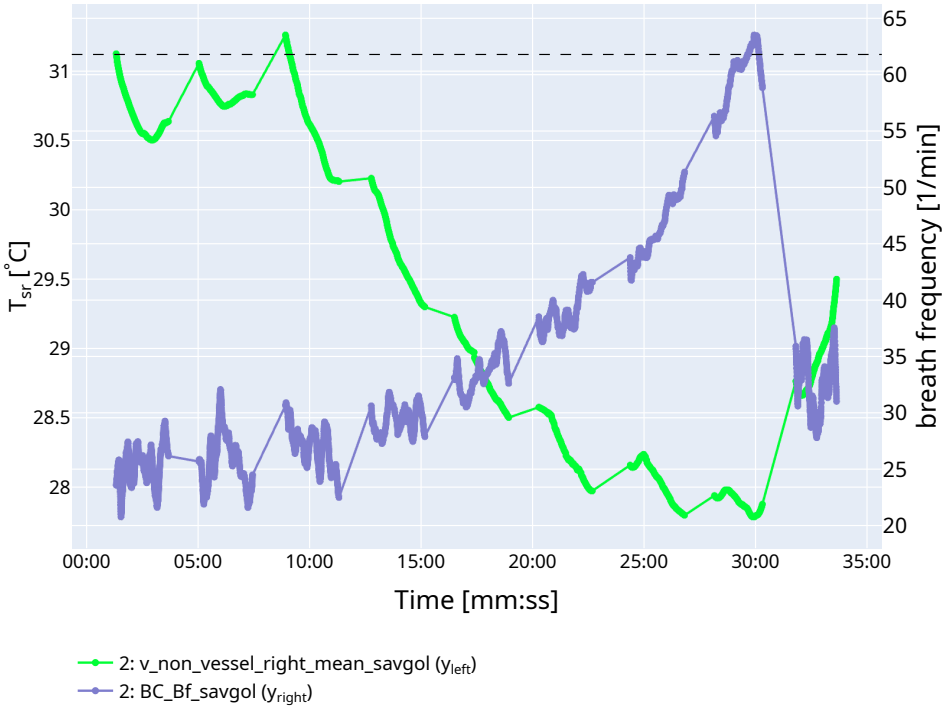


Fig. B.6.: Breath frequency and thermal radiation (bigger version of figure 9.22b).

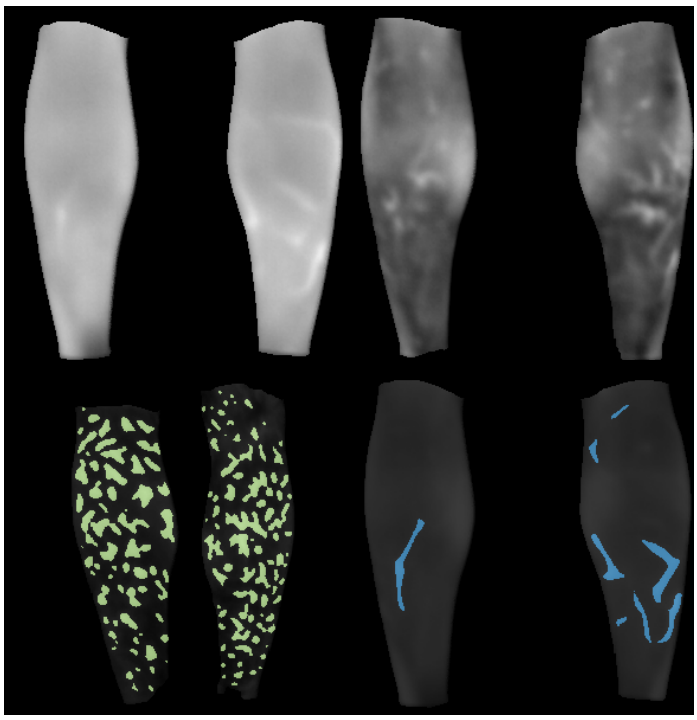


Fig. B.7.: Region of interest and individual thermal radiation distribution patterns (bigger version of figure 9.22c).

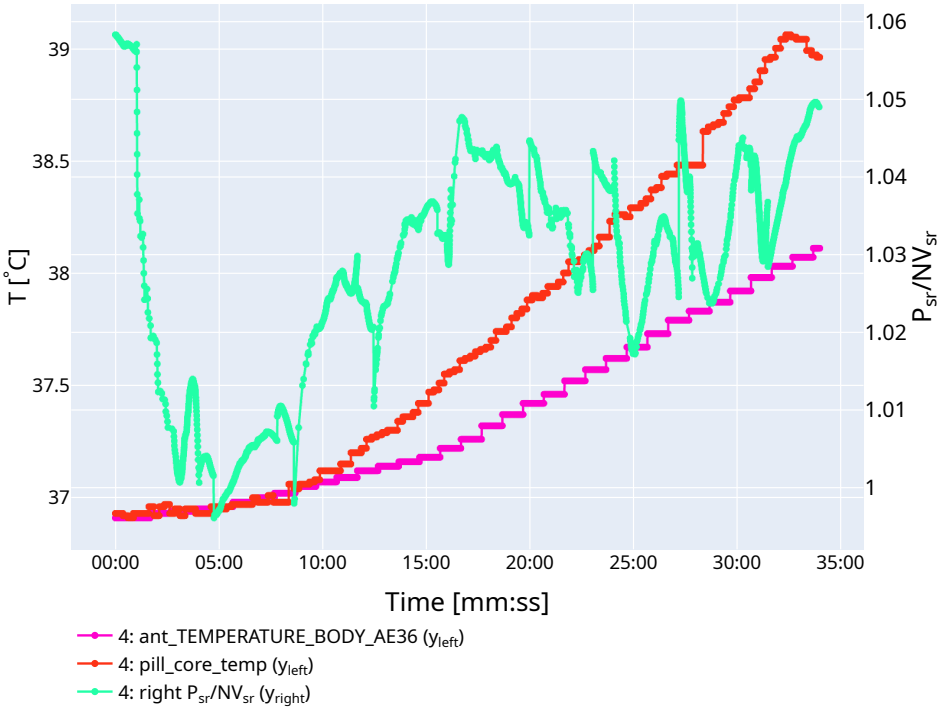


Fig. B.8.: Core temperature and perforator P_{sr} vs. non-vessel NV_{sr} (bigger version of figure 9.22d).

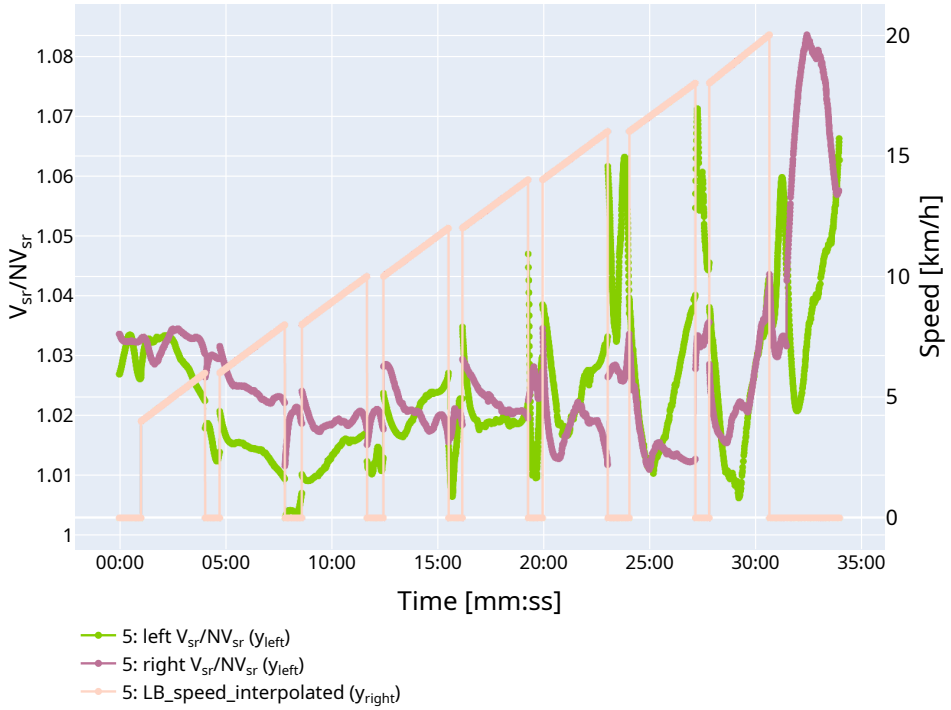


Fig. B.9.: Venous V_{sr} vs. non-vessel NV_{sr} (bigger version of figure 9.22e).

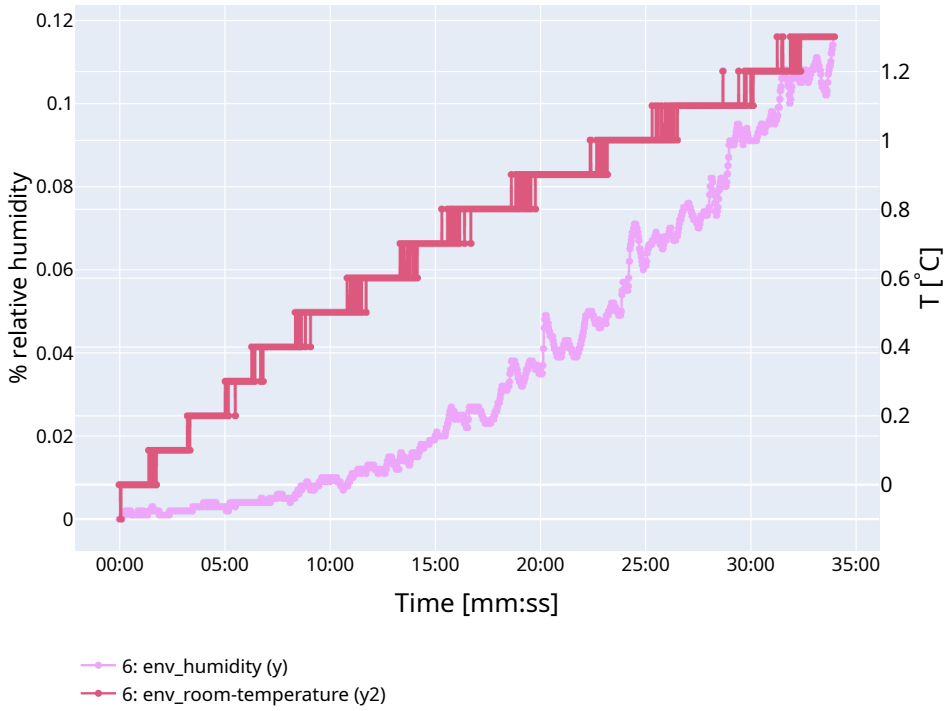


Fig. B.10.: Environmental conditions (bigger version of figure 9.22f).

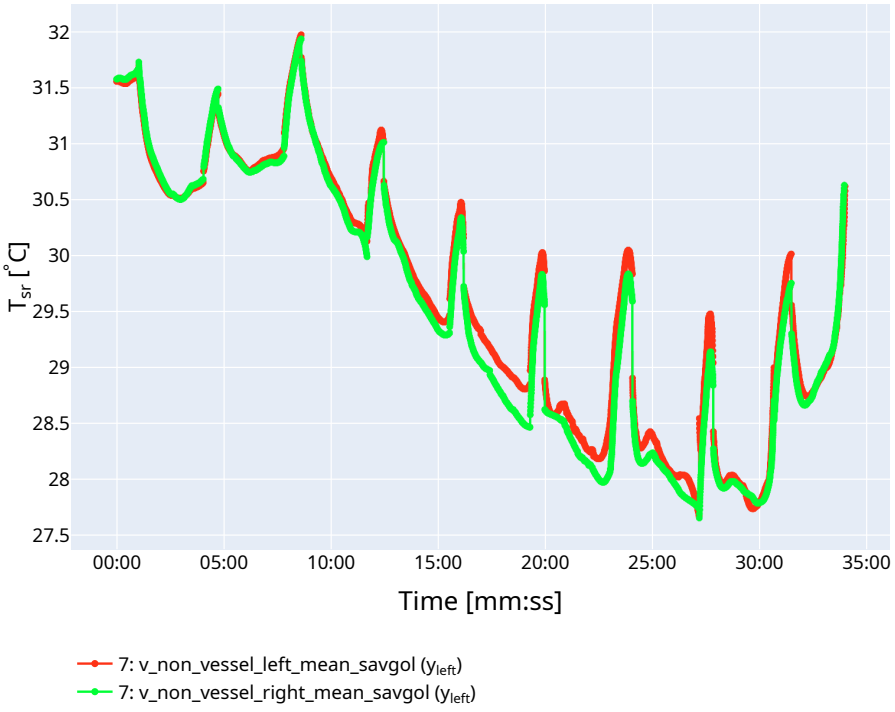


Fig. B.11.: ROI mean T_{sr} left vs. right (bigger version of figure 9.22g).

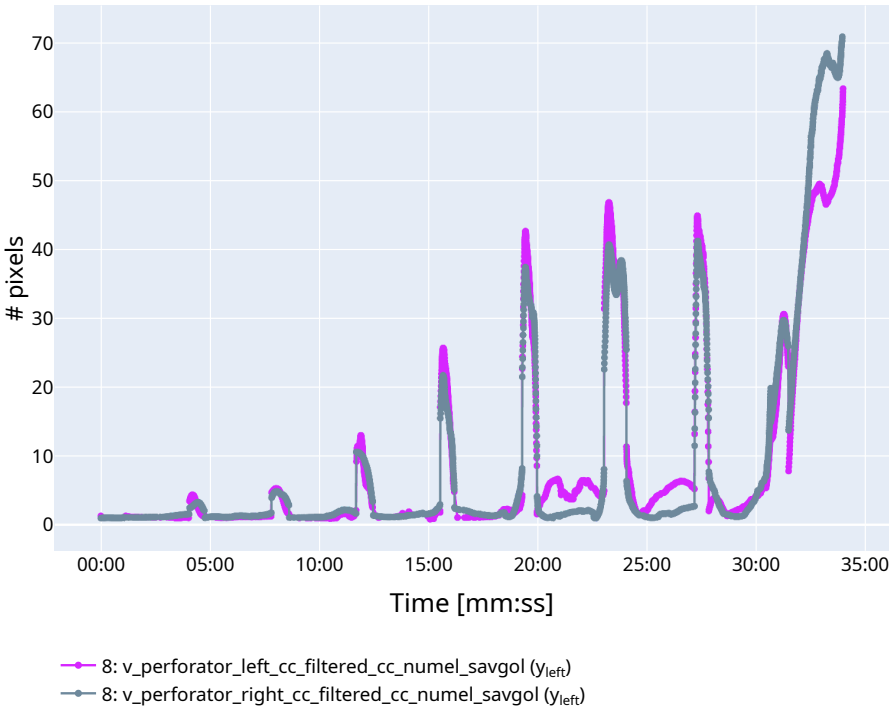


Fig. B.12.: ROI perforator T_{sr} left vs. right (bigger version of figure 9.22h).

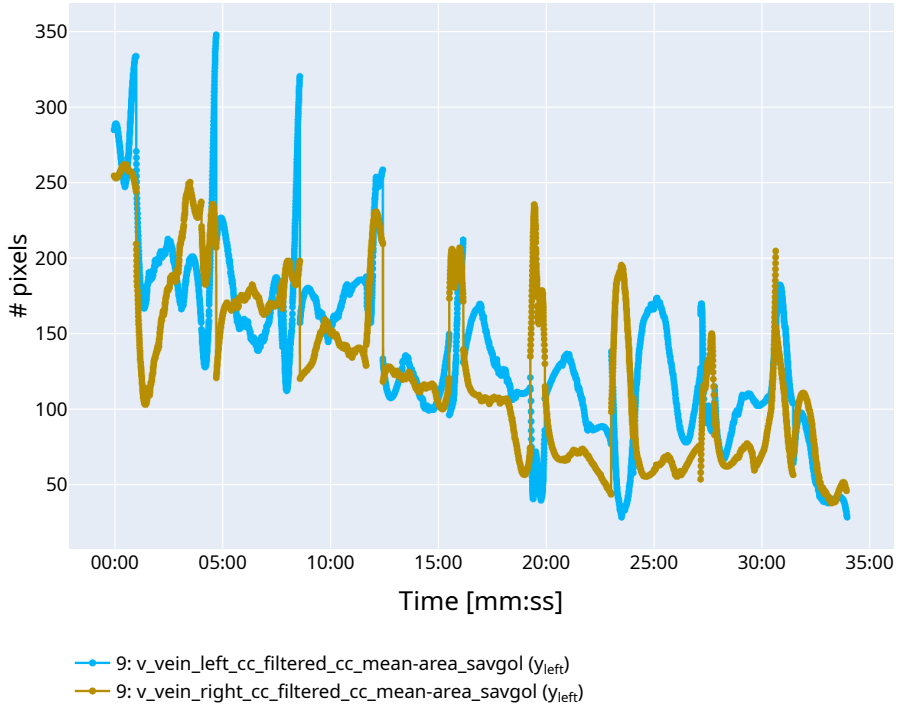


Fig. B.13.: ROI veins T_{sr} left vs. right (bigger version of figure 9.22i).

Colophon

This thesis was typeset with LaTeX 2_ε. It uses the *Clean Thesis* style developed by Ricardo Langner (`http://cleanthesis.der-ric.de/`). The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Adaptations to the style of the Institute of Computer Science can be found at `https://gitlab.rlp.net/institut-fur-informatik/cleanthesis-jgu`.

Declaration

I hereby declare that I have written the present thesis independently and without use of other than the indicated means. I also declare that to the best of my knowledge all passages taken from published and unpublished sources have been referenced. I have documented the AI tools used in the table below.

The thesis has not been submitted for evaluation to any other examining authority, nor has it been published in any form whatsoever.

I duly noted the Regulations for Good Scientific Practice and Dealing with Scientific Misconduct.

Mainz, June 24, 2024

Daniel Andrés López

AI Tool	Used for	Reason	Applied to
DeepL Write	Grammar, spelling, punctuation, and wording correction and optimization.	Readability, correct orthography.	Throughout the entire text.

Use of AI tools.