Journal of Thermal Analysis and Calorimetry (2024) 149:11337–11348
https://doi.org/10.1007/s10973-024-13625-3

ThermoNet: advanced deep neural network-based thermogram processing pipeline for automatic time series analysis of specific skin areas in moving legs

Daniel Andrés López[1] · Barlo Hillen[2] · Markus Nägele[3] · Perikles Simon[2] · Elmar Schömer[1]

Received: 31 August 2023 / Accepted: 30 August 2024 / Published online: 30 September 2024

Abstract
Infrared thermography is an emerging technique in biomedical research, potentially providing diagnostic insights into psychological stress, physical strain, muscle fatigue, inflammation, tissue damage, and diseases with thermogenic effects. However, manual analysis strategies are frequently applied causing incomparable, non-reproducible results and hampering standardization. Moreover, widely applied manual analysis cannot recognize blood vessel-related thermal radiation patterns during physical exercise. Therefore, an enhanced processing pipeline, "ThermoNet", has been developed to automatically process thermograms captured during running. For acquisition, an automatic temperature calibration technique has been introduced to obtain reliable pixel-temperature mapping. The thermograms are semantically segmented in the processing pipeline to extract the anatomical regions of interest (ROIs) by a state-of-the-art deep neural network rather than considering both legs as a single area. A second neural network further examines the ROIs to identify different venous and arterial (perforator) patterns. Within the segments, advanced statistical features are computed to provide time series data. Separate analysis of venous and perforator vessel patterns is carried out on individual connected components, resulting in the extraction of 276 features for each thermogram. The enhanced ROI extraction achieved a high accuracy for the left and right calf on the manually annotated test set. Each step of the ThermoNet pipeline represents a significant improvement over previous analysis methods. Finally, ThermoNet is a transferable pipeline for automatic, reproducible, and objective analysis of ROIs in thermal image sequences of moving test individuals.

Keywords Artificial neural networks · Image processing pipeline · Pattern recognition · Semantic segmentation · Thermal imaging

Introduction

Infrared thermography (IRT) for non-invasive measurement of human thermal skin radiation (T_{sr}) is an emerging technology in medicine and sports [1–4]. Applications include inflammation detection [5], symmetry in muscle activity [6], tumor detection [7], and internal load assessment [8, 9]. The variety of applications demonstrates the potential of the IRT technology. Further investment in automation, reproducibility and objectivity is essential for wider adoption of the technology [1]. Magalhaes et al. have shown a large number of research studies that incorporate thermal imaging

✉ Daniel Andrés López
daniel.andres@uni-mainz.de

Barlo Hillen
b.hillen@uni-mainz.de

Markus Nägele
markus.naegele@optoprecision.de

Perikles Simon
simonpe@uni-mainz.de

Elmar Schömer
schoemer@uni-mainz.de

[1] Research Group of Computational Geometry, Institute of Computer Science, Johannes Gutenberg University Mainz, Staudinger Weg 9, 55128 Mainz, Germany

[2] Department of Sports Medicine, Disease Prevention and Rehabilitation, Institute of Sports Science, Johannes Gutenberg University Mainz, Albert-Schweitzer-Straße 22, 55128 Mainz, Germany

[3] OptoPrecision GmbH, Auf der Höhe 15, 28357 Bremen, Germany

data [2], but in the reviewed work they focused on generating insights from processed thermal data with machine learning algorithms. However, a description of an automated feature extraction is often omitted, and the region of interest (ROI) extraction performed manually (Perpetuini et al. [3]), which limits the amount of data to a few examples of a large possible data stream. As reported in [3], many studies in sport science, which apply IRT, either select ROIs manually or do not specify the selection. Few studies automate the selection of ROIs and analyze during exercise with moving individuals. Hillen et al. describe in [4] the need to analyze T_{sr} in different areas of the body, and they have shown the appearance of blood vessel-related patterns, such as venous or perforator structures. As it comes to a greater application of IRT in T_{sr} analysis, it is crucial to have an automatic approach to process the images to analyze a complete time period and extract all the necessary features from different ROIs. The variety of methods for obtaining and analyzing ROIs among different research groups is huge, which leads to non-comparable and non-reproducible results between them.

Previously, we proposed and extended a pipeline for the analysis of complete data series of thermograms of human posterior legs [8, 9]. The purpose of this paper is to extend the processing steps to a complete analysis pipeline called ThermoNet. Figure 1 outlines the four main steps. First, thermogram acquisition is automated by an external temperature calibration method that maps pixel intensities to temperature values. The approach involves a two-point calibration body, which eliminates the need to manually change the camera's environmental settings. Improvements are demonstrated in both ROI detection steps. The introduction of new ROI definitions allows the algorithm to focus on comparable ROIs, which can be further segmented into vascular-related patterns that occur at different stages of exercise. The final step of extended statistical evaluation of each ROI completes the automatic thermogram processing. The proposed improvements address the need for comparable thermogram analysis of moving individuals with body part related ROIs and no camera dependent calibration overhead for accurate T_{sr} extraction. Thus, the applied pipeline bridges the gap between the ability to acquire large amounts of thermograms with cameras and the limited manual analysis strategy of research groups by introducing a scalable, automated analysis pipeline.

Related work

Magalhaes et al. [2] had shown in the meta-review paper that there are many groups working on automatic evaluation of insights from thermograms. Several medical applications have been mentioned, most of which are related to breast cancer detection. Therefore, the applications focus on the analysis of a few high quality still images and do not consider moving individuals. In sports medicine, however, the participants are often moving during the analysis. Although in recent work such as [6, 10, 11] still non-moving persons are studied. To obtain high quality images, the participant must pause and stand still in predefined poses. Even the ROIs are not determined fully automatically. Aylwin et al. [12] reported in a study the main problems in evaluating T_{sr} in motion. They emphasized the importance of proper ROI selection as well as the differences between high-speed and mid-speed cameras. Another review [13] summarized the application of machine learning approaches in medical imaging. Despite being based mostly on magnetic resonance and X-ray images, the description of the segmentation parts also includes thermograms as in [5]. Other techniques are also summarized, but without the

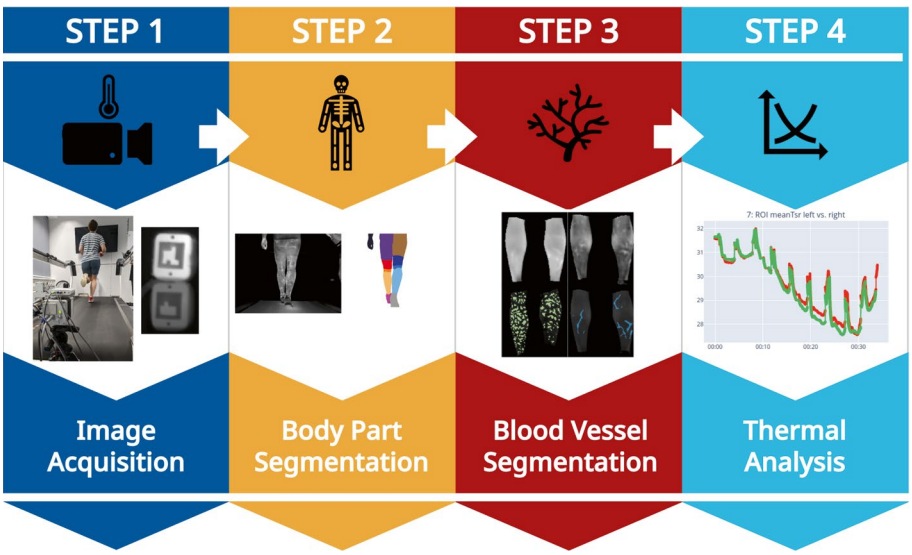


Fig. 1 The ThermoNet processing pipeline consists of four main steps: Step 1 involves image acquisition of a participant running on a treadmill and includes temperature calibration. Step 2 extracts regions of interest (ROIs) in the legs. Step 3 identifies vascular-related structures within the ROIs and step 4 extracts thermal features for all thermograms in an experiment

ability to identify different body regions at once. Bhowmik et al. [14] proposed a modified region growth method for segmentation of inflammatory areas of a knee in thermograms. Magalhaes et al. [15] demonstrated a deep learning approach with a custom-built neural network architecture that outperforms other machine learning approaches for skin cancer segmentation. With their Diabetic Foot Thermograms Network, Cruz-Vega et al. [16] proposed another specialized deep neural network to segment thermograms of humans, in particular the bottom of the feet. In this scenario the person also has to stand still and single images are captured. Unger et al. [17] have developed a Gaussian filter based detection method of the thermal skin surface pattern belonging to the perforator arteries. However, this algorithm is not suitable to distinguish between perforator arteries and other blood vessel structures of the leg, such as veins, and it is not able to do a full segmentation of the analyzed skin surface structure itself. Besides medical applications, He et al. [18] described in their review different applications of IRT and how deep learning methods are applied to perform either classification, feature extraction or semantic segmentation. Although these techniques show potential in the medical field, they lack concrete medical datasets and a fully automated thermal analysis pipeline.

In Hillen et al. [9], a deep learning-based thermogram processing pipeline for analyzing the moving posterior legs of running humans is described. All thermograms of the whole exercise, containing several thousand images, are evaluated. The automatic approach was superior to a manual approach and overcomes the limited number of images analyzed in the manual method. The processing steps include image acquisition, recognition of uncovered legs as ROI, identification of blood vessel patterns within the ROI, conducting a statistical analysis of the results. Based on these findings, the method was further improved in [8] by introducing left and right side differentiation of the body extremities, as well as focusing on the calf instead of the entire leg. However, the paper does not explain these improvements in detail. Table 1 provides a comparison of the two previous methods and the enhancements outlined in this paper. For both [8, 9] the ROI and blood vessel pattern detection is done by two deep neural networks. As described in the paper [9] a modified version of Attention U-Net [19] was developed for the semantic segmentation for both tasks. The work evaluated the validity of the automated pipeline against a manual processing pipeline, showing its superior results and the need for further improvement.

This work involves deep learning to find the ROIs, either of the body parts or of the vascular patterns. The specific task, semantic segmentation, plays an increasingly important role in computer vision and is being continuously improved. Advances in convolutional neural network (CNN) training are most evident in the state-of-the-art architecture

Table 1 A comparison of the improvements to ThermoNet, a four-step pipeline for analyzing thermograms, as proposed in [9], updated in [8], and in this article

	Pipeline in [9] (2022)	Pipeline in [8] (2023)	ThermoNet
Step 1 Image Acquisition	Temperature calibration internally by IRT camera and standardized environment settings	Temperature calibration internally by IRT camera and standardized environment settings	Per thermogram external temperature calibration
Step 2 Body Part Segmentation	One ROI for whole leg parts with visible skin. Network architecture: Modified Attention-U-Net	Different ROIs for left and right calf, but not specified. Network architecture: not specified	Different ROIs for left and right calf side as well for thigh, knee, calf, shoe and pants. Network architecture: DeepLabv3+
Step 3 Vessel Segmentation	Vessels in whole leg and left and right together. Network architecture: Modified Attention-U-Net	Vessels for left and right calf. Network architecture: Modified Attention-U-Net	Vessels for each ROI and side. Network architecture: Modified Attention-U-Net
Step 4 Thermal Analysis	Mean, standard deviation (SD), minimum, maximum T_{sr} and number of pixels for each ROI (in total 20 features)	Additional CP_{sr} for 10% brightest - 10% darkest pixels in ROI (in total 64 features)	Extended statistical features for each ROI including connected components analysis of vessel patterns (in total 276 features)

DeepLabv3+ [20]. Beyond CNNs, vision transformers (ViT) have also recently emerged as a notable alternative, as Thisanke et al. [21] show in their survey. The ViT-based models Segmenter [22] and the recently published foundational segmentation model Segment Anything [23] demonstrate their potential for segmentation tasks. Although these architectures increase the accuracy of predicted segmentation masks in general, they have limited value in specialized medical applications due to the need for large datasets and huge computational load for training and its long inference time. Segment Anything provides a full image segmentation or focuses on specific parts based on input points, making it suitable for interactive applications. The segmentation can be applied to new tasks because the model was trained on a large dataset, but it lacks specific labels for each class. Further modifications are needed to incorporate it into automatic semantic segmentation tasks.

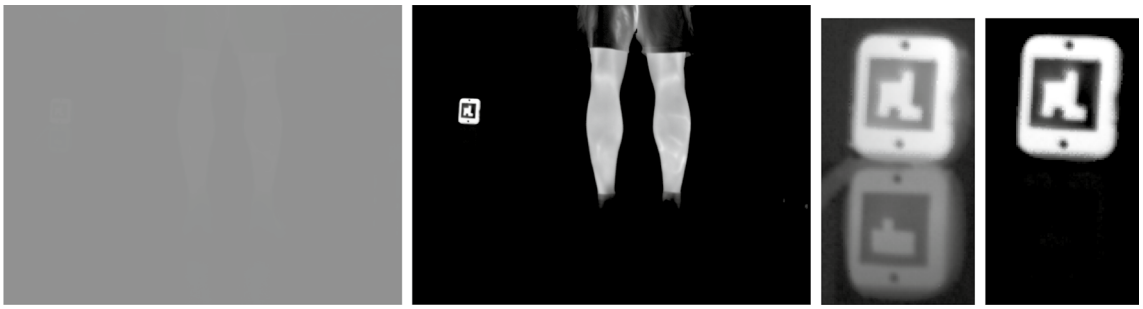
Methods

The ThermoNet pipeline qualitatively improves the previous processing steps in [8, 9] in each part by adding new features or modifying existing ones. Table 1 presents the major improvements over the previous work. The experiments were performed in the Department of Sports Medicine, Disease Prevention, and Rehabilitation of the Institute of Sports Science in Mainz, Germany. All procedures were approved by the Human Ethics Committee Rhineland-Palatinate and are conform to the World Medical Association Code of Ethics (Declaration of Helsinki). The data underlying this analysis were obtained from participants recruited in three different studies (IRB numbers: 837.288.16(10607); 2018–13355 3; 2019–14305; 2021–15713 1). Written informed consent was obtained from all participants. The work focuses on the reliability and reproducibility of the processing system by improving each step individually and increasing the overall performance in terms of accuracy and interpretability.

The first section describes the temperature calibration, which does not rely on the internal camera calibration, but develops an external method related to a blackbody offset calibration. The goal is to reduce the dependence on a camera-specific calibration routine, which is often not fully transparent, by implementing a fully controlled and reproducible system. It automatically determines a mapping from pixel intensities to temperatures. The second section covers the dataset definitions and its extensions over previous work in terms of new segmentation classes within the legs and a side differentiation. The third section discusses body part and vessel segmentation along with its expansions in deep learning algorithms. Additionally, a customized filter algorithm is utilized to ensure consistency checks in body part segmentation; for instance, shoes should not be in the same area as pants. The last section introduces new statistical features in thermogram analysis.

Temperature calibration

The images are captured by the IRT camera VarioCam HD 800 head from Jenoptik with a microbolometer (1024 × 768 IR pixels) at 30 fps and a rolling shutter. To take advantage of the full frame rate, the manufacturer's MicroScan image resolution enhancement unit is not utilized. With MicroScan, the resolution is increased by combining several slightly shifted consecutive frames. For moving people, the combination of multiple frames introduces a lot of motion blur and is not encouraged in this work. The camera captures thermal radiation in the long wave infrared spectrum (7.5–14 µm), has a noise equivalent temperature difference (NETD) of 20 mK at 30 °C and has an accuracy of ±1 K. Each image is stored with 16-bit depth, and a maximum range of -40 to 2000 °C. This work focuses on a specific segment of this spectrum to align with the reasonable T_{sr} for humans from 25 to 35 °C. To convert the captured pixel intensities to T_{sr}, the camera manufacturer offers an interface that enables the conversion from pixel intensities to temperature. The parameters include, among others, ambient temperature, humidity, distance between camera and object, target emissivity, and reflectivity. Because the interface is inaccessible after the image is captured and the internal calibration is not comparable between multiple experiments, images are adjusted and calibrated differently to determine repeatable relative temperature values. A calibration device with two aluminum plates that can be precisely controlled and stabilized at a predefined temperature is introduced. They are set at the limits of the range, one at 25 °C and the other at 35 °C. The plates are colored black to have the same emissivity as the skin surface (0.98). They are placed next to the participant and are always visible in the thermograms. The plates are build as detectable marker to be recognized automatically. Within the specified area, a rectangle of about 250 pixels is taken to calculate the pixel average and correlate it to the known target temperature. Both mappings allow us to linearly interpolate all pixels in the thermogram to obtain temperatures. Figure 2 shows the input image along with the calibration result. Figure 2c displays the temperature plates in a 16-bit image that has been normalized for better visualization. In contrast, Fig. 2d shows them in the calibrated image, where the lower plate disappears because it should have the same intensity value as the lowest possible value of 0, which is also the same as the surrounding and colder background. The calibration method can handle changes in humidity, ambient temperature, or distance without additional



(a) Captured 16-bit thermogram. **(b)** Thermogram in 8-bit (0-255) with calibrated temperature. **(c)** Marker visualization. **(d)** Calibrated marker.

Fig. 2 Original 16-bit thermogram **a** comapred to calibrated 8-bit thermogram **b** in target range. **c** visualizes the reference temperature plates without calibration. The upper marker has a temperature of 35 °C and the lower marker a temperature of 25 °C. **d** shows the reference plates with applied temperature calibration with upper marker as brightest pixel intensity (255), and lower with 0 intensity

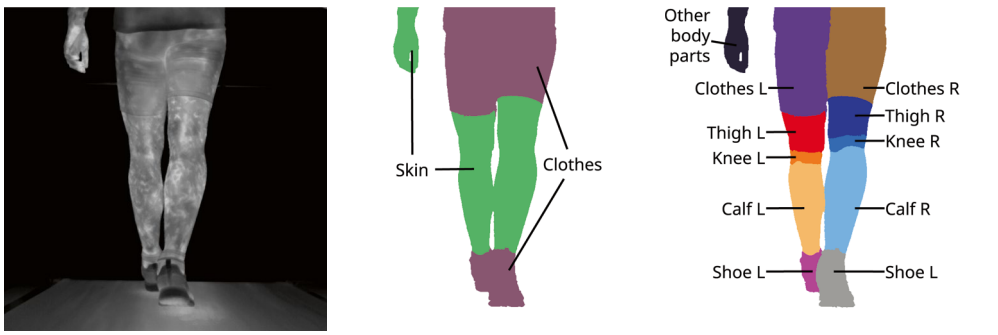
user interaction. Furthermore, this process eliminates the impact of measurement shift in long-running IRT cameras between two nonuniformity calibration (NUC) cycles. NUC refers to the camera's internal recalibration.

In addition, thermograms are also captured with another camera (VarioCam hr head, Jenoptik with 640 × 480 IR pixels at 25 fps, NETD of 50 mK, and ±1.5 K accuracy). These images are stored in a predefined temperature range (25–35°C) as an 8-bit image. The camera API does not allow to specify a file format, images are saved in JPEG format, which has a lossy compression method and image details are slightly altered. The proposed camera calibration technique was not available for this camera.

To train deep neural networks, we annotate the images with ground truth segmentation masks (see Sect. 3.2). The training of the neural networks does not rely on an accurate temperature calibration, since it is applied for pattern recognition in the image itself. The temperature analysis is performed within the found ROIs and is not part of the pattern recognition. Therefore, it is not necessary to train the algorithms with calibrated images. Furthermore, it could be advantageous to include different temperature ranges. Most of the manually labeled segmentation masks were created without temperature calibration.

Data preparation

To train two artificial neural networks for both tasks, a predetermined set of images in the specific domain is manually labeled. For body part segmentation (Body Part Network, BPN), the following new classes are introduced for each side (left and right): shoe (including socks), calf, knee, thigh, pants, and other body parts if they are visible, along with a common background class. Figure 3 shows the improvement in class definitions according to [9]. Although in [8] left and right classes are implemented for the calves, a complete



(a) Thermogram. **(b)** Classes according to [9]. **(c)** New classes within the leg.

Fig. 3 Class definitions for separate body parts within the posterior leg

class definition was not mentioned. Each label also contains a single class assignment for each pixel to perceive a dense pixelwise segmentation mask. The hand-annotated BPN dataset consists of 472 training, 164 validation and 160 test samples. The second task, vessel segmentation (Vessel Network, VN), is only applied to the skin parts of the body. Therefore, to simplify the task, only these regions are examined in the second model, while all other regions are masked with 0. Four segmentation classes are defined: background, non-vessel parts, vein-associated patterns, and perforator-associated patterns. The VN dataset is a subset of the BPN dataset, which includes 374 thermograms for training, 75 for validation and 94 for testing.

Image normalization is a standard technique to ensure a more stable training process. However, two different normalization parameters are necessary. The BPN works with the mean and standard deviation (SD) of the complete acquired and calibrated thermograms. Mean and SD are estimated from about 1 million thermograms to ensure a large variety. The mean is 0.1498 and the SD is 0.2309. Since the input to the VN is a masked version of the original image, the normalization parameters have also to be recomputed. During development, the mean and SD are estimated from a rough segmentation of the legs using a threshold and a morphological closure approach. However, this simplified segmentation includes some parts of the treadmill and clothing. Therefore, future improvements of the pipeline must include the application of the BPN model to estimate the masks for the mean and SD calculation. The necessary normalization parameters are determined from the resulting masked images, yielding $mean = 0.1237$ and $SD = 0.2354$.

Body part and blood vessel segmentation

DeepLabv3+ is applied as current segmentation network architecture for the BPN. The model is initialized with masses from a pretrained model based on the ImageNet [24] data to perceive an initial model with good visual feature detectors. The training procedure includes also randomly applied data augmentation: elastic grid and artificial optical deformations, small rotations and brightness changes, and coarse dropouts. Additionally, images will be randomly zoomed and then resized to 640×480 pixels. As target objects could possibly not be inside the ROI, the cropping process will take care of including a certain amount of non-background class pixels.

As a reference metric, the commonly applied segmentation metric Intersection over Union (IoU), also known as the Jaccard index, is employed. It is defined as the ratio between the intersection of the prediction and ground truth labels with the union of both labels. The average IoU and also the individual values to show the strengths and weaknesses of the model with respect to the different classes are reported.

Dice [25] is chosen as the loss function because it performs well for segmentation tasks, and it is related to the IoU metric. The network masses are updated by the advanced gradient descent optimizer AdaBelief [26] with a learning rate of 0.001. The model is trained with a batch size of 4. The network performance is monitored during training with a validation set and evaluated against a test set of 160 images. The results with the new classes reach an overall IoU of 0.68, the left calf 0.84 and the right side 0.87 (see Table 2).

During running, participants let their legs fall back until they actively move them forward and take a new step. The camera setup is designed to see the straight line of the process. However, the participants do not perform the movements perfectly in line, some have small variations. This is at most the case when the legs overlap and some parts are covered by the other leg or shoe. For each body ROI, the entire area should be considered for feature extraction. However, if parts are occluded, they will not be comparable as they may miss blood vessel structures, which are not evenly distributed in the legs. Therefore, these ROIs in the image are excluded from additional processing by the following approach during post-processing. The BPN is a simple segmentation task and does not consider any special requirements. The post-processing algorithm is briefly described in Algorithm 1. Based on the predicted segmentation mask and the algorithm analyzes the classes for each side for special rules, e.g., the size of a calf must be greater than 10,000 pixels to be valid. If the conditions are not met, the part is removed from further processing. As a result, a filtered prediction mask with a single instance for the left and right calf will be returned.

Table 2 IoU of the BPN network on the test set with 160 test images for each class and overall

Class	IoU left side	IoU right side
Mean	0.68 (no side)	
Background	0.97 (no side)	
Clothes	0.50	0.43
Thigh	0.67	0.73
Knee	0.55	0.61
Calf	0.84	0.87
Shoe	0.61	0.67

Algorithm 1 Filter predicted BPN-mask with predefined rules.

```
for each shoe class do
    Connected component analysis to find the best candidate by filtering for a reasonable size and lower left or right position.
end for
for each other class do
    Find connected components
    for each component do
        if component ≤ 10000 pixels then
            remove component                                   ▷ Remove small components.
        end if
        if component position is below shoe of same side then
            remove component                                   ▷ Remove misplaced detections.
        end if
        if component has big convexity defects then
            remove component    ▷ Remove deformed body parts due to incorrect segmentation or other artifacts.
        end if
        Compare the position with the instance of the same class in the previous image. Select the nearest instance, remove others.
    end for
end for
Compute the convex hull of the remaining components.
Calculate the intersection of the convex hulls of the remaining components.
if left calf and right calf intersect then
    remove both components ▷ Avoid overlapping areas as they can obscure important parts and distort statistical analysis.
end if
```

After post-processing, the calf mask is passed to the VN. With this technique, some images do not have a calf class due to the high angle of the legs resulting in a small calf area, overlap, or misclassification. Therefore, these images are not included in the statistical analysis.

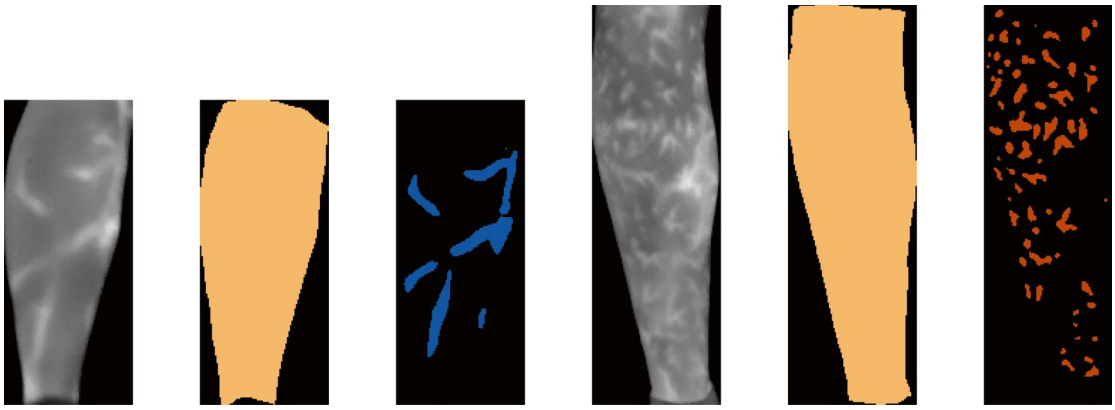
The VN follows the same strategies as the BPN, but has an adopted Attention U-Net architecture. As an additional simplification of the problem during training, the input images are masked by the ground truth data of the body parts. Only skin classes remain in their values, while clothing, shoes, and background are set to zero. This also corresponds to the background class of the vessel labels. It should also help the network to focus on the necessary features. As above, a random cropping approach is employed to get smaller parts for training, which could lead to cropping into a pure background area. To avoid this, a small area must be non-background.

Thermogram analysis

With both segmentation masks, the thermogram is analyzed to obtain the radiation properties in different ROIs. The task focuses on the left and right calf. These two classes are the most comparable classes for the analysis of this work because they are usually not affected by different trousers lengths. Furthermore, the calves do not touch each other and a clear segmentation is more reliable than for the thighs. Within the two ROIs, VN detects the non-vascular areas, the veins and the perforator. These three sub-ROIs are analyzed together with the full ROI. This paper presents a small subset of the extracted features and leaves the full investigation to future studies, as the selection of the most insightful indicators depends on the task and the ROI. The values are calculated based on the pixel intensities and then converted to a temperature value with the calibration values obtained in Sect. 3.1. The extracted features include number of pixels, mean, SD, median, lowest pixel intensities (`min`), highest pixel intensities (`max`), mean and median of the 10% lowest pixel intensities (`min10`), mean and median of the 10% highest pixel intensities (`max10`), $\texttt{max10} - \texttt{min10}$ (`diff10` for mean and median), entropy, skewness and kurtosis (Fisher and Pearson).

Additional details are provided for the vein and perforator classes. Both classes have many smaller instances that are isolated and not connected to other instances. These connected components are analyzed individually with the spaghetti labeling algorithm [27]. The analysis is performed on all components and on a filtered subset. Filtering is achieved by excluding all components that have a size less than 8 pixels. Features for the connected component analysis include all of the statistics listed above, the number of connected components, the average and median area of each component, and the range of component sizes.

Different numbers of features are calculated for the different ROIs. For a calf, 22 features are extracted for the lower leg and 20 features for the non-vessel area. The vein and perforator ROIs each have the base features (20) plus four for the connected components. The filtered connected components also add 24 features. Therefore, the final number of features per class is 48. The total number of features is 138 for one side and 276 for both sides. Figure 4 shows two



(a) Thermogram with veins. **(b)** Calf segmentation. **(c)** Vein segmentation. **(d)** Thermogram with perforators. **(e)** Calf segmentation. **(f)** Perforator segmentation.

Fig. 4 Two examples of vein and perforator patterns. The images are cropped to show only the left calf

Table 3 Example statistics for the left calf and left veins in Fig. 4a–c

Feature	Left calf	Left veins
Number of pixels	12,903	1637
Mean/°C	29.387	30.742
Standard deviation/°C	1.025	0.859
Median/°C	29.275	30.725
Entropy	13.608	10.661
Skewness	−0.534	0.469
Kurtosis-pearson	7.346	3.792
min10-mean/°C	27.551	29.37
max10-mean/°C	31.279	32.444
diff10-mean/°C	3.728	3.074
Number of connected components	–	7
Mean-area of connected components	–	233.857

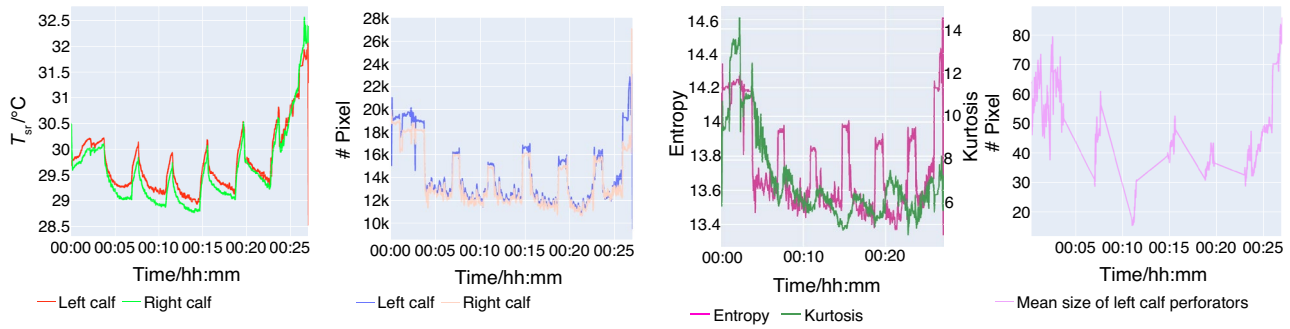
sample thermograms from different time points in a measurement of a single person. One focuses on high amount of venous patterns, and the other on the perforators. The first image is captured during movement, and the leg is not fully extended, resulting in a smaller detected area compared to the other thermogram, where the person is standing still. Table 3 displays a portion of the statistics for the ROIs of a sample image Fig. 4a: the left calf Fig. 4b and its veins Fig. 4c.

Figure 5 shows sample plots for a time series. The experiment was performed by an individual walking and running several stages of increasing speed for three minutes each. After each stage, a short pause was introduced for the test protocol. At the end, a slow recovery speed of 4 km h^{-1} was set for three minutes. The curves are smoothed with a Savitzky-Golay-filter for better visualization. These time series help to examine the differences between left and right (a,b). The first shows the average temperature over time for each side. The peaks represent the breaks. The curves are slightly different, but comparable. The second plot compares the ROI sizes for both sides. Leg sizes are larger in the stance phase than in the movement phase. The third plot (c) shows the entropy and kurtosis over time for the left calf. And the last plot (d) visualizes the mean size of the connected components of the left calf perforators. There are long areas without data in the motion phases, where small and fewer perforators are difficult to detect. However, the average size of the perforators increases, making them more visible and easier for the algorithm to detect.

Discussion

The pipeline consists mainly of four steps to exploit the full potential: automatic temperature calibration, body part segmentation, blood vessel segmentation, and thermogram statistics and time series generation. The previous processing pipeline has already been applied to several studies [8, 9] and is improved in this paper.

As a first step, an automatic temperature calibration procedure with a two-plate calibration object next to the participant is proposed. To the best of our knowledge, this is a novel technique to calibrate the temperature mapping for each image, including adaptation to changing ambient temperature and humidity. Compared to the previous approach, which relied on the camera's internal mapping, the temperature scale of thermograms is now interchangeable with



(a) T_{sr} of the left and right calves.

(b) Evolution of the size of these ROIs.

(c) Entropy and kurtosis for the left calf.

(d) Mean area of the perforators in the left calf.

Fig. 5 Four plots showing different extracted features for calves for an entire experiment

respect to the two reference points. This allows us to decide the best range during the analysis time instead of the measurement time. The main drawback of this approach is the lack of proven reliability. For comparison, it is necessary to evaluate the standard blackbody calibration procedure against the proposed technique, as recommended in [12, 28]. Blackbody calibration involves a single blackbody with a reference temperature. The calibration includes the systemic shift correction by applying the difference between the measured and expected temperature. Otherwise, the new approach includes the shift correction along with a linear interpolation between the two reference points. For each image, a new calibration is calculated from the temperature plates. Only a small area within these plates is averaged as a reference for pixel intensities. Also image denoising would lead to more stable results over time, but is not considered in this work.

With BPN, an extended method is implemented to extract body-related ROIs of a moving person. As stated in [12], it is highly necessary to have a fixed ROI definition that does not change over time. The calves are less likely to be covered by other parts or clothing than the thighs, so further work will only evaluate the calves. In addition, Aylwin et al. [12] discussed the role of thermograms of moving persons and non-moving persons. The camera employed in this work captures long-wave infrared (LWIR) with an uncooled microbolometer, which requires about 10 ms to integrate the sensor data for each pixel row (the camera has a rolling shutter). The authors of [12] compare these types of cameras with a high-speed mid-wave infrared camera. As expected, the increase in speed affects the sharpness of the images. The resulting motion blur with the LWIR camera, together with rolling shutter effects, also affects the ROI detection of both the BPN and the VN. Since the deep neural networks are trained with 472 annotated training images and 164 validation images, including many images of people standing still and fewer moving, the dataset needs to be increased to include more moving situations. The same applies to the test set of 160 images. According to Aylwin et al. the mid-speed camera can measure comparable results as a high-speed one, but with less accuracy of the temperature. The cyclic movement of the leg ensures that there are repeatedly measurements with lower speed of the leg (relative to the camera) which can be captured with more precision. Figure 6 shows a sample step and how the ROI changes within a step. The effect of motion blur can be seen in the later images, while the first images remain sharp.

The results of the deep neural networks are measured by the IoU. The most important results are the IoUs for the evaluated classes, the calves. Calves have the best coverage in the data because they are always shown and visible. Thighs can be covered by pants that vary in length and tightness. The tighter the pants, the more thermal radiation will pass through from the underlying thigh, resulting in a less clear separation from the uncovered thigh. The shoe classes are not always fully visible in the image because they may have a lower T_{sr} than the temperature scale. Therefore, they cannot be recognized well. This is also true for manual annotation, it is up to the annotator to decide where to mark the complete shoe if it is not visible. However, it is necessary to improve the prediction of all classes for a better generalization of the models. Better results in other classes can be leveraged to focus on ROIs other than calves directly. To improve the segmentation results, a hyperparameter search including a decision on different network architectures has to be performed. The current approach focuses on proof of principle, but not on DNN optimization. In this part of the research, recently published technologies such as the Segment Anything Model [23] or the Segmenter Model [22] should be considered, although these models will increase the inference time. In the image pipeline shown here, the images are processed individually, without any knowledge of

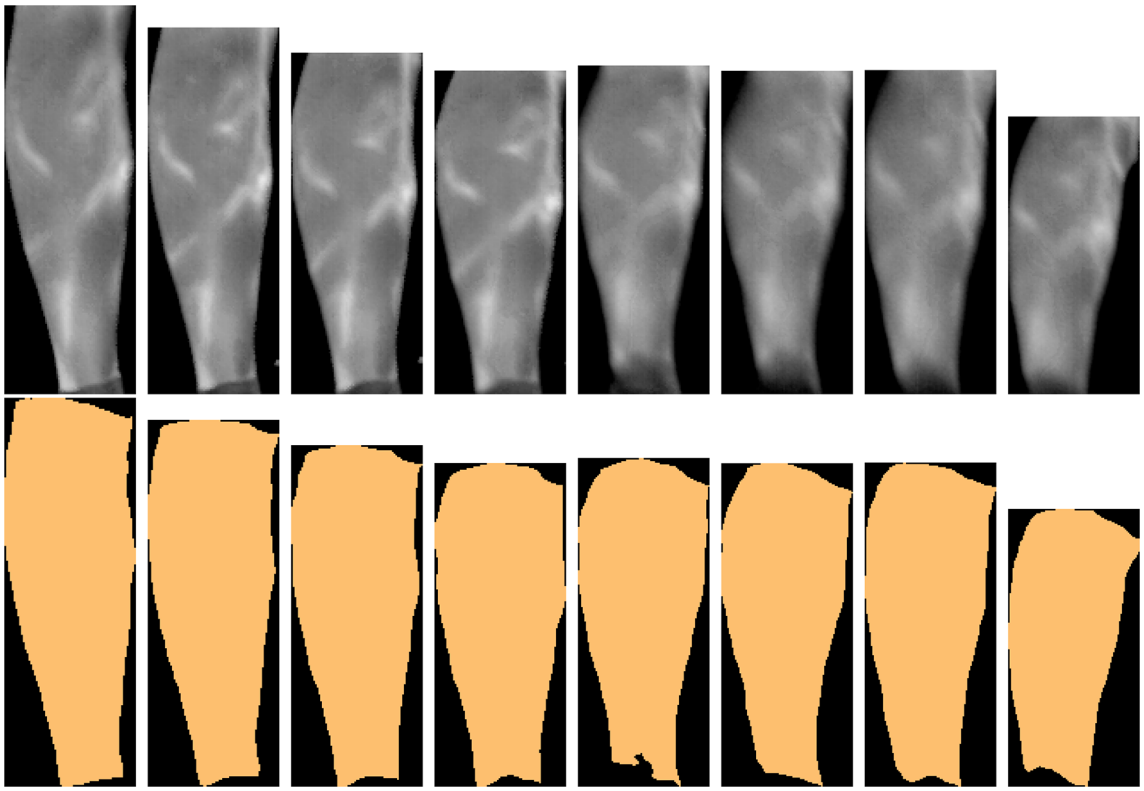


Fig. 6 The first row shows the left calf of 8 consecutive thermograms of a single step within 230 *ms*, and the second row the corresponding automatically detected ROI

the previous or subsequent images. To improve the consistency of the results, it should be investigated how to extend the analysis to several consecutive images. In this area, a recent paper on tracking multiple moving objects in a video shows promising results [29].

Step 2, the body part segmentation also includes a refinement step. It is an algorithm with consistency checks from the world, such as a calf must have a minimum size, must not overlap with other calves, or should not have large deformation artifacts. The processing pipeline assumes that there is a single, well-defined ROI for each calf. During the execution of this approach, calves are often removed and no data can be generated. One reason for this is often the intersection of the calves. A calf should not be occluded by other parts and cannot be included if it is. However, the segmentation of the calf in front of the other one cannot be found exactly and the ROI is no longer well-defined. Therefore, it is also removed. Checking these rules is done after the first neural network and may reduce the inference speed of the whole pipeline. Another approach to enforce the rules would be to integrate them into the training process and optimize the network to ensure the single instance criteria according to its relationships with the other classes to further improve the IoU metric.

As a final step, thermogram analysis provides statistical insight into each ROI. This allows a deep analysis of the changing properties of each ROI during the exercise. Many features are introduced with each ROI based on the analysis strategy, but the feature set has not been systematically evaluated for the most insightful features. Within the vessel pattern ROIs, a connected component analysis detects structural properties of individual components. These features can be helpful in analyzing the pattern itself without the need for reliable temperature mapping. How the body adapts to its vascular system by analyzing the distribution behavior over time is an underexplored area of research. The calculated features can be plotted as a time series to compare them over time. However, the signals are not completely stable and fluctuate. One reason is the previously mentioned inconsistent ROI detection. If the ROIs differ in their area, they may include or exclude some significant parts, which may affect the features. Therefore, a refinement step is applied after BPN to reduce inconsistent ROIs. A second possibility is the cyclic motion effect. The size and angle of the ROIs change periodically over time due to the walking motion. The analyzed legs move cyclically and the extracted features reflect the cycles. An analysis strategy must be aware of and deal with the recurring movement, either by including elements only when a certain property is met, such as the size of the

ROI being above a threshold. Alternatively, the time series can be smoothed with a low-pass filter to reduce the effect of noise. The time series itself can be merged with other sensors such as a heart rate monitor, a breath analyzer, as well as environmental data, e.g., ambient temperature and humidity, treadmill incline, elapsed time, speed, and received personal extinction (RPE). It is necessary to pay attention to the correct merging of these data because they do not have the same frequency, precision, and reliability as the thermograms.

Conclusions

ThermoNet enables sophisticated automatic processing of thermograms acquired from humans during physical exercise. The analytical approach includes end-to-end processing from image acquisition to statistical feature extraction. In this way, non-invasive IRT can be established more efficiently in medical and sports science use cases. In our work, all essential processing steps of the ThermoNet pipeline have been improved. The first step, image acquisition with automatic calibration, allows automatic image pixel-temperature mapping without requiring environmental settings acquisition. Hence, thermograms are directly fed into subsequent steps. In steps 2 and 3, BPN and VN detect the thermal patterns within more specific ROIs now. Thereby, the BPN outcome is refined by a rule-based system. Subsequently, ThermoNet achieves higher accuracy and consistent measurements in consecutive images, and exclude occluded or misclassified ROIs from further processing. In step 4 connected component analysis has been added and therefore allows a precise analysis of the ROIs concerning specific thermal pattern variation over time (appearance and size). Finally, all of these four steps build the essential foundation for automatic and objective thermogram time series analysis.

In future, we intend to implement classification algorithms to identify physiological features within the processed time series to support decision-making in medical and sports science applications of IRT.

Author contributions Daniel Andrés López contributed in study design, implementation of algorithms, data labeling and selection. He also wrote the first manuscript draft. Barlo Hillen contributed in study design, data interpretation and manuscript editing and reviewing. Markus Nägele contributed in infrared camera instrumentation, thermal imaging camera calibration and reviewing the manuscript. Perikles Simon helped in study design, interpretation and manuscript reviewing. Elmar Schömer contributed in algorithm design, interpretation and manuscript reviewing.

Funding Open Access funding enabled and organized by Projekt DEAL. The authors gratefully announce that this research study was funded by the Federal Ministry for Economic Affairs and Energy based on a decision of the German Parliament (ZF4211603GR9) and by the Johannes-Gutenberg University research program "Inneruniversitäre Forschungsförderung (Stufe I)" with title: "Extend DeepSpoMed: Computer-aided domain adoption for human skin body segmentation in thermal images." We would like to thank all participants for their voluntary participation in this study.

Declarations

Conflict of interest The authors declare that they have no Conflict of interest. All authors read and approved the final manuscript.

References

1. Ramirez-GarciaLuna JL, Bartlett R, Arriaga-Caballero JE, Fraser RDJ, Saiko G. Infrared thermography in wound care, surgery, and sports medicine: a review. Front Physiol. 2022. https://doi.org/10.3389/fphys.2022.838528.
2. Magalhaes C, Mendes J, Vardasca R. Meta-analysis and systematic review of the application of machine learning classifiers in biomedical applications of infrared thermography. Appl Sci. 2021;11(2):842. https://doi.org/10.3390/app11020842.
3. Perpetuini D, Formenti D, Cardone D, Filippini C, Merla A. Regions of interest selection and thermal imaging data analysis in sports and exercise science: a narrative review. Physiol Meas. 2021;42(8):08–01. https://doi.org/10.1088/1361-6579/ac0fbd.
4. Hillen B, Pfirrmann D, Nägele M, Simon P. Infrared thermography in exercise physiology: the dawning of exercise radiomics. Sports Med. 2020;50(2):263–82. https://doi.org/10.1007/s40279-019-01210-w.
5. Das K, Bhowmik MK, Prasad Mukherjee D. Segmentation of knee thermograms for detecting inflammation. In: 2019 IEEE International Conference on Image Processing (ICIP), IEEE, Taipei, Taiwan 2019;1550–1554. https://doi.org/10.1109/ICIP.2019.8803094
6. Szurko A, Kasprzyk-Kucewicz T, Cholewka A, Kazior M, Sieroń K, Stanek A, Morawiec T. Thermovision as a tool for athletes to verify the symmetry of work of individual muscle segments. Int J Environ Res Public Health. 2022. https://doi.org/10.3390/ijerph19148490.
7. Tayel MB, Elbagoury AM. An efficient and reliable method for regional analysis of breast thermographic images. Glob Sci J. 2020;8(9):1508–18.
8. Hillen B, Andrés López D, Pfirrmann D, Neuberger EW, Mertinat K, Nägele M, Schömer E, Simon P. 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. J Therm Biol. 2023;113: 103498. https://doi.org/10.1016/j.jtherbio.2023.103498.
9. Hillen B, Andrés López D, Schömer E, Nägele M, Simon P. Towards exercise radiomics: deep neural network-based automatic analysis of thermal images captured during exercise.

IEEE J Biomed Health Inform. 2022;26(9):4530–40. https://doi.org/10.1109/JBHI.2022.3186530.
10. Hillen B, Andrés López D, Marzano-Felisatti JM, Sanchez-Jimenez JL, Cibrián Ortiz de Anda RM, Nägele M, Salvador-Palmer MR, Pérez-Soriano P, Schömer E, Simon P, Priego-Quesada JI. Acute physiological responses to a pyramidal exercise protocol and the associations with skin temperature variation in different body areas. J Therm Biol. 2023;115: 103605. https://doi.org/10.1016/j.jtherbio.2023.103605.
11. Bogomilsky S, Hoffer O, Shalmon G, Scheinowitz M. Preliminary study of thermal density distribution and entropy analysis during cycling exercise stress test using infrared thermography. Sci Rep. 2022. https://doi.org/10.1038/s41598-022-18233-5.
12. Aylwin PE, Racinais S, Bermon S, Lloyd A, Hodder S, Havenith G. The use of infrared thermography for the dynamic measurement of skin temperature of moving athletes during competition; methodological issues. Physiol Meas. 2021;42(8): 084004. https://doi.org/10.1088/1361-6579/ac1872.
13. Parashar A, Rishi R, Parashar A, Rida I. Medical imaging in rheumatoid arthritis: a review on deep learning approach. Open Life Sci. 2023;18(1):15–7. https://doi.org/10.1515/biol-2022-0611.
14. Bhowmik MK, Das K, Bhattacharjee D. Temperature profile guided segmentation for detection of early subclinical inflammation in arthritis knee joints from thermal images. Infrared Phys Technol. 2019;99(January):102–12. https://doi.org/10.1016/j.infrared.2019.04.011.
15. Magalhaes C, Tavares JMRS, Mendes J, Vardasca R. Comparison of machine learning strategies for infrared thermography of skin cancer. Biomed Signal Process Control. 2021. https://doi.org/10.1016/j.bspc.2021.102872.
16. Cruz-Vega I, Hernandez-Contreras D, Peregrina-Barreto H, Rangel-Magdaleno JJ, Ramirez-Cortes JM. Deep learning classification for diabetic foot thermograms. Sensors. 2020;20(6):1–22. https://doi.org/10.3390/s20061762.
17. Unger M, Markfort M, Halama D, Chalopin C. Automatic detection of perforator vessels using infrared thermography in reconstructive surgery. Int J Comput Assist Radiol Surg. 2019;14(3):501–7. https://doi.org/10.1007/s11548-018-1892-6.
18. He Y, Deng B, Wang H, Cheng L, Zhou K, Cai S, Ciampa F. Infrared machine vision and infrared thermography with deep learning: a review. Infrared Phys Technol. 2021. https://doi.org/10.1016/j.infrared.2021.103754.
19. Oktay O, Schlemper J, Le Folgoc L, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, Glocker B, Rueckert D. Attention U-Net: Learning where to look for the pancreas. In: 1st Conference on Medical Imaging with Deep Learning, Amsterdam, Netherlands (2018)
20. Chen L.-C, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Computer Vision – ECCV 2018, vol. 11211 LNCS, 2018;833–851. Springer, Cham. https://doi.org/10.1007/978-3-030-01234-2_49
21. Thisanke H, Deshan C, Chamith K, Seneviratne S, Vidanaarachchi R, Herath D. Semantic segmentation using vision transformers: a survey. Eng Appl Artif Intell. 2023;126: 106669. https://doi.org/10.1016/j.engappai.2023.106669.
22. Strudel R, Garcia R, Laptev I, Schmid C. Segmenter: Transformer for Semantic Segmentation. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021;7242–7252. https://doi.org/10.1109/ICCV48922.2021.00717
23. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo W-Y, Dollár P, Girshick R. Segment Anything (2023) https://doi.org/10.48550/arXiv.2304.02643
24. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009;248–255.
25. Bertels J, Eelbode T, Berman M, Vandermeulen D, Maes F, Bisschops R, Blaschko MB. ptimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory and Practice. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.-T., Khan, A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, vol. 11765 LNCS, pp. 92–100. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_11
26. Zhuang J, Tang T, Ding Y, Tatikonda S, Dvornek N, Papademetris X, Duncan JS. AdaBelief Optimizer: Adapting Stepsizes by the Belief in Observed Gradients. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, editors. Advances in Neural Information Processing Systems, vol. 33. Curran Associates Inc; 2020. p. 18795–806.
27. Bolelli F, Allegretti S, Baraldi L, Grana C. Spaghetti labeling: directed acyclic graphs for block-based connected components labeling. IEEE Trans Image Process. 2020;29(1):1999–2012. https://doi.org/10.1109/TIP.2019.2946979.
28. Ring EFJ, Ammer K. The technique of infrared imaging in medicine. In: Infrared Imaging. 2015;2053-2563, 1–10. IOP Publishing, Bristol, UK . https://doi.org/10.1088/978-0-7503-1143-4ch1
29. Wang Q, Chang Y-Y, Cai R, Li Z, Hariharan B, Holynski A, Snavely N. Tracking everything everywhere all at once. 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 2023;19738–19749.