



Benchmarking conventional and machine learning segmentation techniques for digital rock physics analysis of fractured rocks

Marcel Reinhardt¹ · Arne Jacob¹ · Saeid Sadeghnejad² · Francesco Cappuccio³ · Pit Arnold⁴ · Sascha Frank⁵ · Frieder Enzmann¹ · Michael Kersten¹

Received: 9 November 2020 / Accepted: 4 September 2021 / Published online: 25 January 2022
© The Author(s) 2022

Abstract

Image segmentation remains the most critical step in Digital Rock Physics (DRP) workflows, affecting the analysis of physical rock properties. Conventional segmentation techniques struggle with numerous image artifacts and user bias, which lead to considerable uncertainty. This study evaluates the advantages of using the random forest (RF) algorithm for the segmentation of fractured rocks. The segmentation quality is discussed and compared with two conventional image processing methods (thresholding-based and watershed algorithm) and an encoder–decoder network in the form of convolutional neural networks (CNNs). The segmented images of the RF method were used as the ground truth for CNN training. The images of two fractured rock samples are acquired by X-ray computed tomography scanning (XCT). The skeletonized 3D images are calculated, providing information about the mean mechanical aperture and roughness. The porosity, permeability, flow fields, and preferred flow paths of segmented images are analyzed by the DRP approach. Moreover, the breakthrough curves obtained from tracer injection experiments are used as ground truth to evaluate the segmentation quality of each method. The results show that the conventional methods overestimate the fracture aperture. Both machine learning approaches show promising segmentation results and handle all artifacts and complexities without any prior CT-image filtering. However, the RF implementation has superior inherent advantages over CNN. This method is resource-saving (e.g., quickly trained), does not need an extensive training dataset, and can provide the segmentation uncertainty as a measure for evaluating the segmentation quality. The considerable variation in computed rock properties highlights the importance of choosing an appropriate segmentation method.

Keywords Fracture segmentation · Quality · Digital rock physics · Random forest algorithm · Convolutional neural network · Image processing

This article is a part of the Topical Collection in Environmental Earth Sciences on “Sustainable Utilization of Geosystems” guest edited by Ulf Hünken, Peter Dietrich and Olaf Kolditz.

✉ Marcel Reinhardt
mreinh05@uni-mainz.de

¹ Geosciences Institute, Johannes Gutenberg-University, J.-J. Becherweg 21, 55099 Mainz, Germany

² Department of Petroleum Engineering, Faculty of Chemical Engineering, Tarbiat Modares University, Tehran, Iran

³ Department of Geology, University of Otago, P.O. Box 56, 9054 Dunedin, New Zealand

⁴ Department Petroleum Engineering, Montanuniversität Leoben, Parkstraße 27, 8700 Leoben, Austria

⁵ Faculty of Geosciences, Institute of Geology, Mineralogy and Geophysics, Ruhr-University Bochum, Universitätsstraße 150, 44801 Bochum, Germany

Introduction

In low-permeable reservoirs, fractures are the dominant pathways for fluid flow and mass transport, which are influenced by fracture properties like aperture distribution. Therefore, a proper characterization of fractures is crucial for understanding flow in these reservoirs (Bodin et al. 2003; Stoll et al. 2019). Increasing computational power and advances in interdisciplinary scientific fields such as water resources research allowed DRP to become a key tool in studying fractured media (Andrä et al. 2013a, b). This approach is now used in many applications, including contaminant transport carbon capture and storage, enhanced oil recovery, tight gas production, reactive flow (Iassonov et al. 2009; Blunt et al. 2013; Blöcher et al. 2019; Hinz et al. 2019; Stoll et al. 2019; Zhang et al. 2019; Moslemipour and

Sadeghnejad 2021). A typical DRP workflow consists of three steps: (i) image acquisition and processing, (ii) image segmentation (i.e., categorizing each voxel to a specific phase, like pore or matrix), and (iii) numerical simulations (Andrä et al. 2013a, b). None of these steps is trivial, and the predictions of numerical simulations heavily depend on the quality of previous steps (Saxena et al. 2019).

The image acquisition in the DRP workflow is commonly performed by X-ray microtomography or medical XCT scanning depending on sample size and spatial resolution of interest (Ketcham and Carlson 2001; Cnudde and Boone 2013). The acquired 3D images suffer from various artifacts, which are typically corrected during and after the scanning process. Examples of the most common artifacts are noise, beam hardening, ring artifacts, and partial volume effect (PVE). Therefore, the image quality depends on samples, scanner setup, applied filters and reconstruction methods (Ketcham and Carlson 2001; Kaestner et al. 2008; Iassonov et al. 2009; Cnudde and Boone 2013; Schlüter et al. 2014; Garfi et al. 2020). Artifacts make the segmentation process more complicated and can even render specific segmentation methods useless, whereas filtering as well as the choice of reconstruction settings can lead to a loss of critical information (Iassonov et al. 2009; Saxena et al. 2017; Berg et al. 2018; Pot et al. 2020).

Segmentation is the most crucial step affecting the quantitative or qualitative structural analysis of a rock sample (Iassonov et al. 2009; Andrä et al. 2013a; Leu et al. 2014; Saxena et al. 2017; Pot et al. 2020). Ideally, the gray value histogram of rock images would be multi-modal, with distinct peaks representing present phases. However, depending on available phases and imaging artifacts, most histograms are complex with a continuous range of gray values describing two or more phases. This makes precise segmentation challenging (Al-Marzouqi 2018; Berg et al. 2018). Accordingly, there is an enormous variety of different approaches for the segmentation process (e.g., see Iassonov et al. (2009) and Schlüter et al. (2014) for an overview of conventional methods). The conventional methods include simple thresholding or the watershed algorithm, which is based on region growing of seeds and considering local intensity gradients. Generally, the watershed method is more robust against noise and long-range gradients than conventional thresholding methods. The segmentation quality can be significantly impacted by choice of method or small input parameters changes, which introduces an operator bias (Leu et al. 2014; Shou et al. 2020). Thus, the conventional methods are hugely susceptible to user bias (Iassonov et al. 2009).

As an alternative approach, various supervised (Karimpouli et al. 2020; Lu et al. 2020; Lee et al. 2021) and unsupervised (Taibi et al. 2019; Zhang et al. 2021) machine learning techniques were implemented for fracture

identification and segmenting porous samples (Chauhan et al. 2016). Among the unsupervised approaches, encoder-decoder networks in the form of CNNs received much attention during fracture identification in the DRP workflow (Varfolomeev et al. 2019; Hong & Liu 2020; Karimpouli et al. 2020; Kim et al. 2020; Lu et al. 2020; Lee et al. 2021). Studies showed the capabilities of CNNs compared to other segmentation methods (Lu et al. 2021; Lee et al. 2021). However, providing a ground truth for training remains challenging. Recent studies (e.g., Karimpouli et al. 2020; Wang et al. 2021) prepared the ground truth by manually segmenting images. This approach is time-consuming and heavily suffers from user bias. As an alternative approach, segmentations from conventional methods can be used as ground truth (e.g., Niu et al. 2020; Lee et al. 2021). However, since segmentation results from the conventional methods are not free of user bias (e.g., due to the choice of input parameters), the CNN segmentation quality can be affected (Niu et al. 2020). Although more complex CNN structures are expected to perform better, the quality of the training data is critical for utilizing CNNs. When working with unique samples, it can become difficult, if not impossible, to get enough appropriate training data (Hiramatsu et al. 2018; Karimpouli et al. 2020; Varfolomeev et al. 2019; Niu et al. 2020).

In recent years, interactive (supervised) machine learning algorithms like the RF classifier (Ho 1994; Breiman 2001; e.g., implemented in the Trainable Weka Segmentation, TWS: Arganda-Carreras et al. 2017 or in the open-source software package Ilastik by Berg et al. 2019) were primarily developed for biological cases (Babel et al. 2020; Brady et al. 2020; Eyolfson et al. 2020) and offer an alternative to deep learning methods. Berg et al. (2018) showed that depending on the training, the RF classifier performs better than the classical segmentation methods and does not require any prior filtering (Garfi et al. 2020; Kovács et al. 2019; Scanziani et al. 2018). The drawbacks of the TWS implementation include high computational demands and the variety of available filters and classifiers, which makes it difficult to implement without sufficient resources and advanced knowledge of the methods.

This study evaluates the capabilities of the RF classifier as a segmentation method. For this purpose, two fractured rocks are used. For benchmarking, CNN, and conventional segmentation methods (thresholding-based and watershed algorithms) are applied to both structures. Moreover, the breakthrough curves from laboratory experiments are chosen as the ground truth and are quantitatively compared with the breakthrough curves simulated from each segmented image. The influence of the segmentation techniques on the simulated breakthrough curves is studied, and fracture properties are compared.

Materials and methods

Rock samples and XCT scanning

In this study, two synthetically fractured samples (diameter \times length = 100 \times 150 mm), prepared by a Brazilian test from two different sandstones (Remlinger and Flechtinger), were investigated. The Flechtinger sandstone has already been described in many previous studies (Frank et al. 2020a, b; Hassanzadegan et al. 2012; Stanchits et al. 2011; Blöcher et al. 2009; Heiland 2003). A matrix permeability of 10^{-17} – 10^{-16} m² and porosity between 6 and 11% were reported. The Remlinger sandstone is a stratigraphic part of the upper Buntsandstein with the Triassic origin (Schwarzmeier, 1978). A porosity of 12.9% and a matrix permeability of 4 – 7×10^{-17} m² was determined (Frank et al. 2020a, b).

The medical XCT measurements of both fractured sandstone cores were carried out using a SOMATOM Definition AS VA48A instrument. A beam energy of 120 keV and 90 mA was applied. The images were reconstructed by the device software (Somaris/7 syngo CT VA48A), which normalized the gray level values to the Hounsfield scale (ranging from -1000 to over 3000) commonly used in medical fields (Hounsfield 1980). The pre-processing and the conventional watershed/thresholding segmentation procedures were carried out in 3D using Avizo 2019.3 (FEI Visualization Sciences Group, France). The raw image stacks were pre-processed in three steps, (i) resampling to a uniform voxel size of 200 μ m using the Lanczos filter (Lanczos 1950), (ii) image rescaling to 16-bit (unsigned) and shifting the values by 2000 to eliminate negative gray values, and (iii) setting the background to a uniform gray level value of zero. The pre-processed images of dimensions 570 \times 570 \times 789 voxels were segmented in the next step.

Conventional segmentation methods

Since both conventional methods are susceptible to user bias, several segmented image stacks with varying parameters were generated to present a range of plausible segmentation outputs. Additional to the manual segmentation methods, two global thresholding methods were implemented, the Otsu method (Otsu 1979) and K-means clustering (MacQueen 1967), computed by GeoDict 2020 (Math2Market GmbH, Kaiserslautern, Germany). For the manual threshold method, the minimum threshold was set to 10. The maximum gray value for the Remlinger fracture (Fig. 1A) was set to 3350, 3380, 3400, 3410, 3420, 3430, 3440, 3582 (Otsu method), and 3588 (K-means clustering).

For the Flechtinger sandstone, the maximum gray values of 3360, 3380, 3400, 3410, 3430, 3440 (K-means clustering), 3450, 3470, 3490, and 3662 (Otsu method) were used (Fig. 1B).

The watershed segmentation seeds of the Flechtinger sandstone were placed with the thresholds of 0–1177 (background), 1177–3350.4 (fracture) and 3630–5344 (matrix). For the Remlinger sandstone, seeds were placed with the thresholds of 0–130 (background), 131–3310 (fracture) and 3800–5430 (matrix). The thresholds were chosen to produce the least segmentation errors at the boundary between fracture and matrix (i.e., avoiding errors that merged with the fracture) while still segmenting as much of the fracture voxels as possible. The region growing of seeds was conducted by the watershed implementation in Avizo 2019.3 (based on Beucher and Meyer 1993).

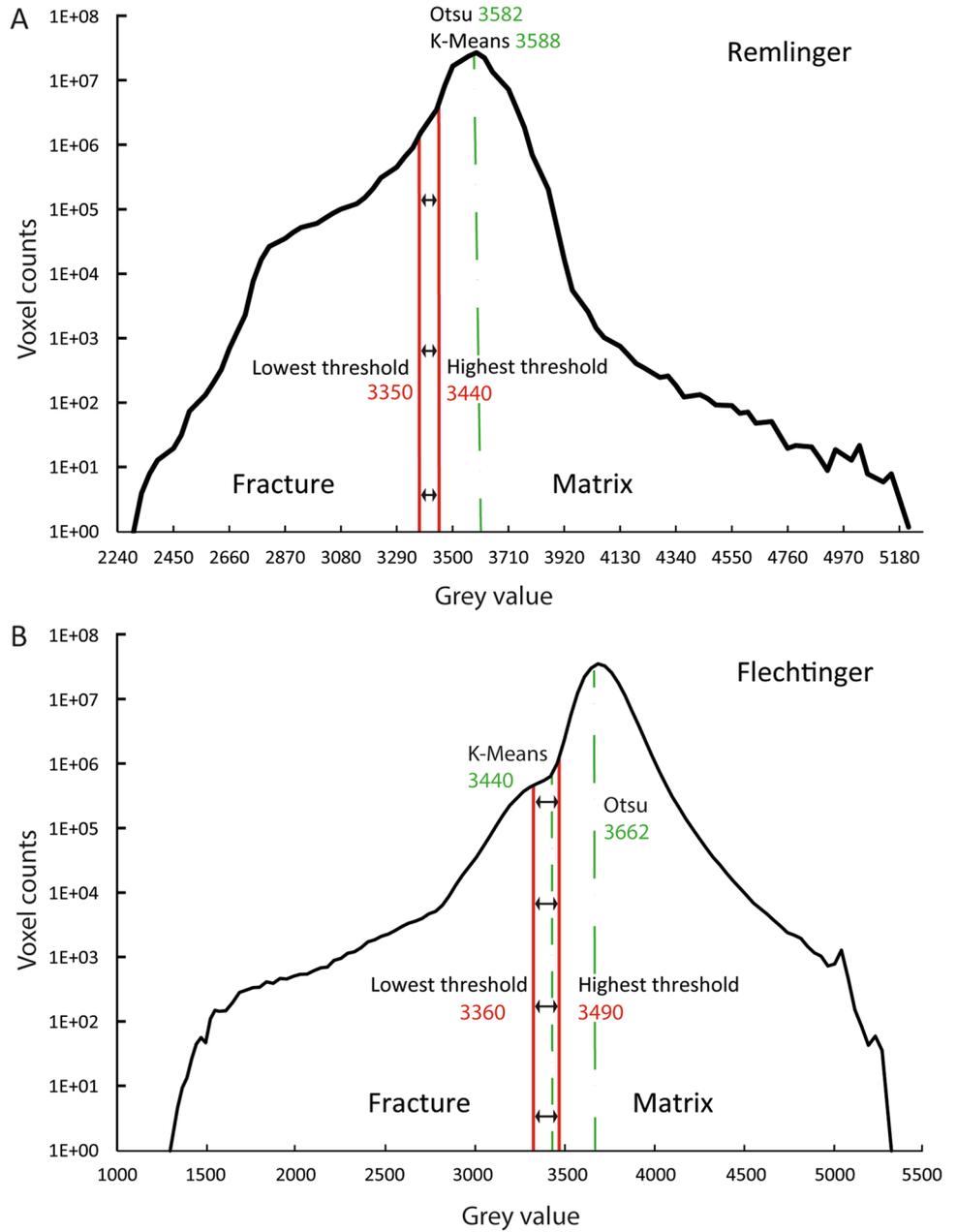
The Remlinger dataset was cropped to a dimension of 570 \times 570 \times 679 voxels to remove the artifacts, produced by the cone-beam X-ray, on the upper (50 slices) and lower end (60 slices) of the dataset. The Flechtinger dataset was cropped by 50 slices on both ends (in z-direction), resulting in dimensions of 570 \times 570 \times 689. The matrix voxels falsely segmented as fracture voxels (islands) were removed with the “cleanse” tool (implemented in GeoDict 2020). This tool searches for connected (by face, edge, or vertex) fracture components smaller than a given size and reassigned their phase to the matrix phase (Fig. 2).

Segmentation using a convolutional neural network

The CNN used in this study is based on a U-net-like architecture (Ronneberger et al. 2015) with the encoder in the form of VGG (Simonyan and Zisserman 2015). Figure 3 depicts the architecture of the network. The roles of the encoder and decoder modules are to extract image features (by down-scaling) and changing the extracted features into predictions (by up-scaling), respectively. The U-net structure links (concatenates) encoder to decoder layers by implementing some symmetric skip connections. These links allow extracted shallow features to be transferred directly to decoding, which speeds up the convergence process and increases the final training accuracy (Wang et al. 2021).

The training input is grayscale 16-bit XCT images with a size of 544 \times 544 and a batch size of 3. Both encoder and decoder modules are distributed over several blocks with different convolutional/deconvolutional layers (Fig. 3). The first and second two blocks of the encoder module have two and three convolutional layers, respectively. All convolutional layers have a kernel size of 3 \times 3, with the same padding of the input layer and various depths (from 64 to 512, increase by a factor of 2). The convolutional layers are followed with a rectifier linear unit (ReLU) function to introduce non-linearity into the model. It returns 0

Fig. 1 The gray value histogram of the fractured **A** Remlinger and **B** Flechtinger sample. The red lines mark the segmentation range of the different gray value thresholds. Values on the right of the threshold are assumed to contain impermeable matrix (solid) voxels, while values to the left contain fracture (pore) voxels. The green dotted lines mark the mean global thresholds



for negative inputs and linear increment up for positives inputs. The max-pooling layers have equal kernel sizes of 2×2 (with 2 stride), which reduce the size of input images by a factor of 2. The decoding blocks contain a deconvolution layer (also called transposed convolution or upscaling layer) followed by 1×1 zero padding, convolution layer, and batch normalization layer. The deconvolutional layers double the size of feature maps. The role of zero paddings with a size of 1×1 in the decoder module is to ensure that the center of each filter corresponds to the first pixel of input images. To regularize the network, batch normalization is implemented in the final layer of each decoder block. The final layer of the network is a convolution layer

with 2 filters (either fracture or matrix phase). The network has 12,321,026 trainable parameters in total.

A typical cost function for binary image segmentation tasks is “binary cross entropy” (Eq. 1) which shows the similarity between the predicted segmentation (s_p) and ground truth (g_t).

$$BCE = -g_t \log(s_p) - (1 - g_t) \log(1 - s_p). \tag{1}$$

However, as a small fraction of each training image is only filled with fracture pixels (fracture porosity is around 1%), we deal with unbalanced class sizes of matrix and fracture phases in each segmented image. In this situation, this

Fig. 2 The thresholding segmentation procedure on the Remlinger sandstone. **A** The gray value image. **B** Pre-processed image. **C** Segmentation using a threshold of 3420. Blue areas contain fracture (pore) voxels. **D** Cropping of top and bottom slices of the structure to remove the cone-beam artifacts. Applying the cleanse tool to remove falsely segmented voxels (“islands”) from the dataset

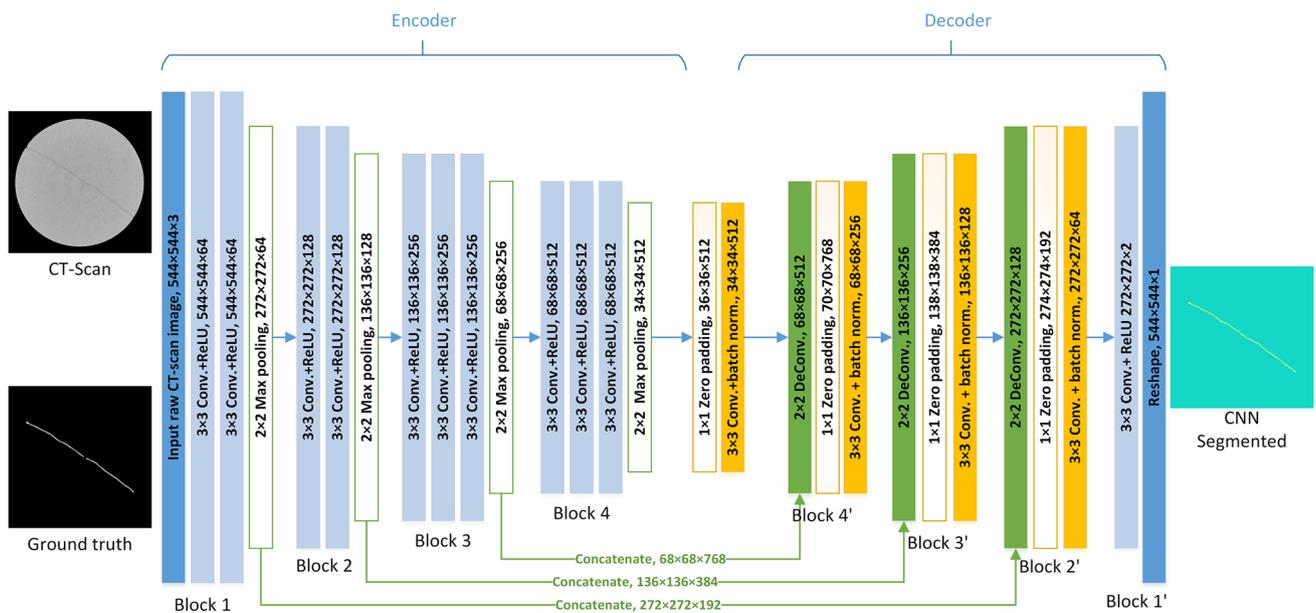
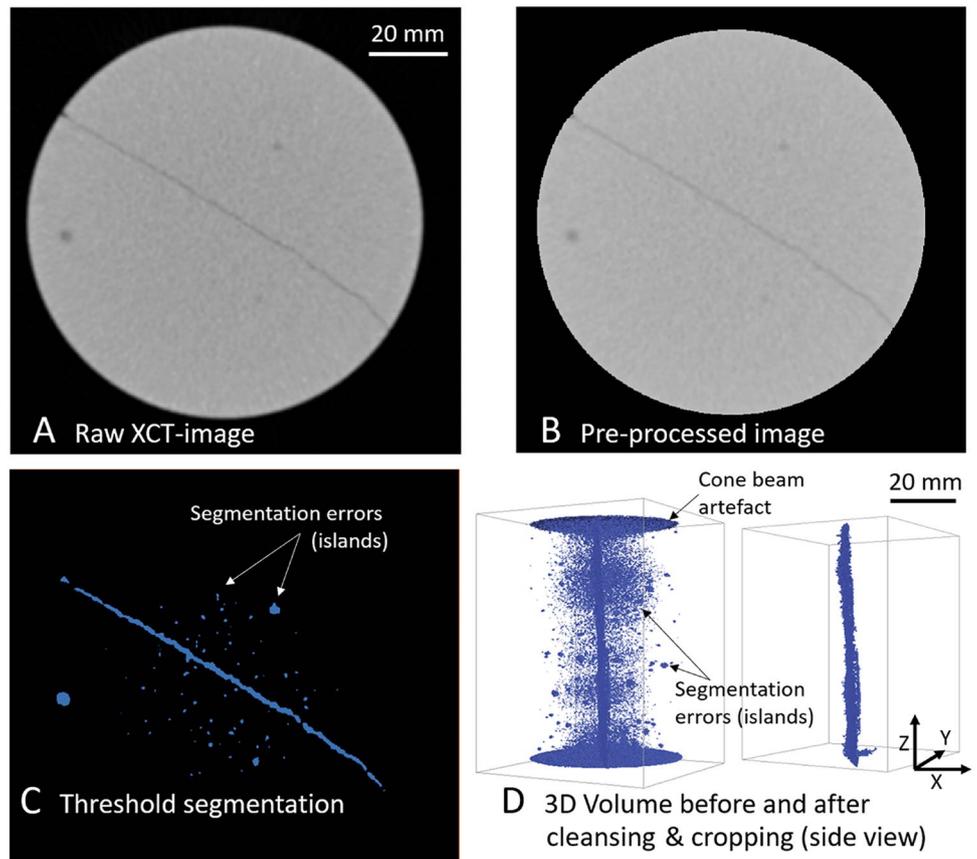


Fig. 3 Architecture of the implemented VGG-U-net composed of encoder and decoder modules, each consists of various blocks. Input to the network are grayscale XCT images and the output is a binary segmented image in which fracture/matrix phases are identified

loss function cannot be solely a good representative of loss values. Therefore, a weighted loss function of binary cross entropy and Jaccard distance loss (J_l) is implemented in this study. J_l measures the dissimilarity between two images of s_p and g_t and is defined as:

$$J_l = 1 - \text{IoU} = 1 - \frac{|s_p \cap g_t|}{|s_p| + |g_t| - |s_p \cap g_t|}, \quad (2)$$

where IoU is the intersection over union (IoU) of two images (also called the Jaccard index), which is defined as the area of overlap between two images ($|s_p \cap g_t|$) divided by their union area ($|s_p \cup g_t| = |s_p| + |g_t| - |s_p \cap g_t|$). $|x|$ indicates the number of specific phase items (e.g., fracture pixels) in x . The IoU index is zero if the two image sets are dissimilar and is one if they are identical. To compute the network accuracy during the validation process, we use the IoU index of the fracture phase. A weighted sum of BCE and J_l with a weight (w) of 0.8 was used for loss function (L),

$$L = wJ_l + (1 - w)\text{BCE}. \quad (3)$$

The network is trained with an initial learning rate of $1e-5$ using an RMSProp optimizer (Hinton et al. 2012), a gradient-based optimization technique. The training phase is carefully monitored to avoid under and overfitting by tracking both training and validation accuracies and losses. The Image-Segmentation-Keras library (<https://github.com/divamgupta/image-segmentation-keras>) is implemented as the base for the CNN training. All trainings were performed on an Nvidia GeForce GTX 1080 Ti GPU with 11 GB RAM.

The preparation of a sophisticated ground truth is time-consuming (usually done on a pixel-by-pixel basis) and suffers from user bias. Thus, we use the cleansed segmented images of the RF method as the ground truth. The CT-scan image of Remlinger and Flechtinger fractured samples are composed of 679 and 689 2D images with a size of 544×544 pixels, respectively. To augment the number of training dataset, the *imgaug* library (<https://github.com/aleju/imgaug>) is used. This process allows us to increase the prediction accuracy and reduce the chance of overfitting during the training phase. The augmentation process artificially increases the number of XCT-images by geometrically modifying them. The geometrical augmenters include, flipping up-down and left-right, cropping from 10 to 30%, scaling to 80–120%, rotating by -45 to $+45^\circ$, shearing by -15 to $+15^\circ$, and translating by -20 to $+20\%$. We increase the number of 2D images of each fractured sample to approximately twofold by randomly applying one or more augmenters on the original images of each data set.

We first train the CNN on augmented images of one fractured sample (e.g., Remlinger with 2037 2D images).

We use 80% of the images for training and the rest 20% for evaluation purposes. We then test the trained network on the XCT images of the other sample (Flechtinger with 689 2D images). In the next step, we switch the training and test rock samples (i.e., training on Flechtinger and testing on Remlinger). The reason for switching rock datasets is to show the CNN's generality when it is applied on various datasets.

Segmentation using a random forest classifier

The RF classifier consists of multiple decision trees, each providing a vote for class prediction of each voxel. The class with the highest vote is assigned to the corresponding voxel. Individual trees are built using bagging and feature randomness to produce uncorrelated trees (Breiman 2001). Filters are applied to the input images. These filtered images act as features, which the classifier uses to distinguish phases in the dataset.

The RF classifier (with 100 trees) was chosen because of its superior generalization performance, robustness, and choice of just a few parameters (Berg et al. 2019). The RF implementation of the Ilastik software (version 1.3.3) provides a predefined, customizable workflow (allowing to show preliminary results and uncertainties in slices) with a set of filters with adjustable parameters. Only one slice at a time is classified during the interactive training to lower the computational demands. This allows applying the workflow on datasets, significantly larger than the RAM [opposed to the TWS implementation, which is RAM limited (Berg et al. 2019)]. After the manual training, the classifier can execute the segmentation process on the entire dataset. Alternatively, the training on 3D images is also possible, but with higher computational demand. The segmentation can be applied to input data in up to 5D (3D, time, and channels).

The pre-processed XCT-images were segmented with the RF classifier using all predefined filters (Gaussian Smoothing, Laplacian of Gaussian, Gaussian Gradient Magnitude, Difference of Gaussians, Structure Tensor Eigenvalues, Hessian of Gaussian Eigenvalues), each with sigma values of 0.3 (only for the Gaussian Smoothing), 0.7, 1.0, 1.6, 3.5, 5.0, and 10.0, to increase the number of training data and obtain a more robust classification. The three classes chosen were 'fracture', 'matrix', and 'background'. All training data were provided manually

Table 1 Training data provided to the RF classifier. Both raw datasets consist of 256,346,100 voxels, which were segmented as background, matrix, or fracture

Sample	Trained matrix voxels	Trained fracture voxels	Trained background voxels
Remlinger	58,806	1,905	6,849
Flechtinger	63,095	4,515	5,481

with brush strokes and refined using the uncertainty map and preliminary segmentation results to minimize segmentation errors visually. Table 1 summarizes the number of trained voxels for each phase in both sandstone samples. For the Remlinger and Flechtinger sandstone, a 0.0285% and 0.0264% of voxels were used as the training data, respectively.

Fracture aperture and roughness

The local fracture apertures were measured by a modified version of the Fracture Trace Point (FTP) algorithm (Cappuccio et al. 2020) on the segmentation output of all algorithms. The algorithm operates on sets of 2D grayscale images by collecting the values across CT profiles (orthogonal to the x- and y-axis). The signal is then analyzed so that a peak–valley–peak shape in the signal, occurring when crossing an open fracture, is recorded as FTPs at the deepest points (valleys). For every point recorded, 2D and 3D measurements of fracture aperture parameters and local orientations are performed. This operation involves several steps. First, the 2D Euclidean distance transform of all non-zero pixels (Fig. 4B) of an input binary image (Fig. 4A) are calculated (Eq. 4),

$$px_i = \sqrt{\sum_{i=1}^n (px_i - b_i)^2}, \tag{4}$$

where n is the number of dimensions. Every pixel px_i then obtains a value equal to the length of its distance to the

nearest background point b_i . Thus, pixels receive higher values (always > 0) when moving towards the middle of a fracture aperture (Fig. 4B). This results in a peak-shaped signal within the fracture (e.g., Fig. 4E), rather than the valley-shaped signal needed for the FTP analysis. Grayscale images with valleys occurring at the middle of the fracture aperture were then obtained by inverting the calculated values in all images (Fig. 4F). During orthogonal traverses at each recorded FTP, the apparent width (note that traverses are not orthogonal to the fracture) was measured as the distance (in pixels) from the two gray value peaks before and after the valley related to the inspected FTP (Fig. 4F). To correct the apparent width and obtain the local actual aperture, the 2D orientation and 3D dip angle measured from the scanning direction as the zenith of the FTP are used to calculate the three-dimensionally orthogonal aperture (Fig. 4C, G). The skeletonized 3D images of all identified FTPs were created at the end of the analysis. These images visually provide information about the true aperture distribution (Fig. 4D, H) and orientation variation using the deviation of the dip angle of each FTP from the overall mean value. Additionally, the fracture roughness was calculated using all estimated local apertures by dividing the standard deviation of measured apertures by the mean aperture (Zimmerman et al. 1991). Low roughness values correspond to smoother fracture surfaces (e.g., two planar planes have a roughness of 0).

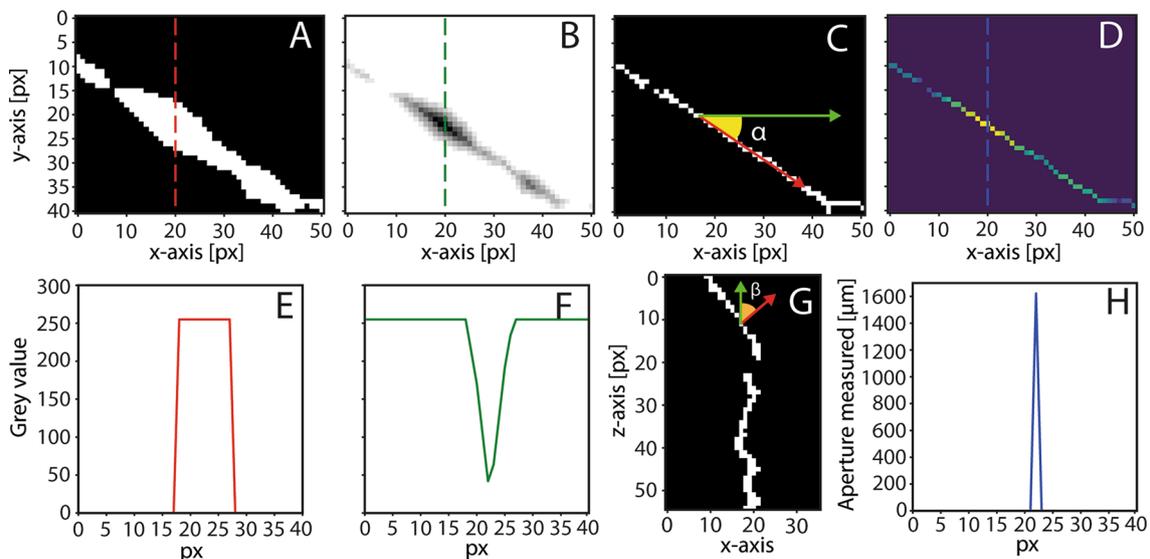
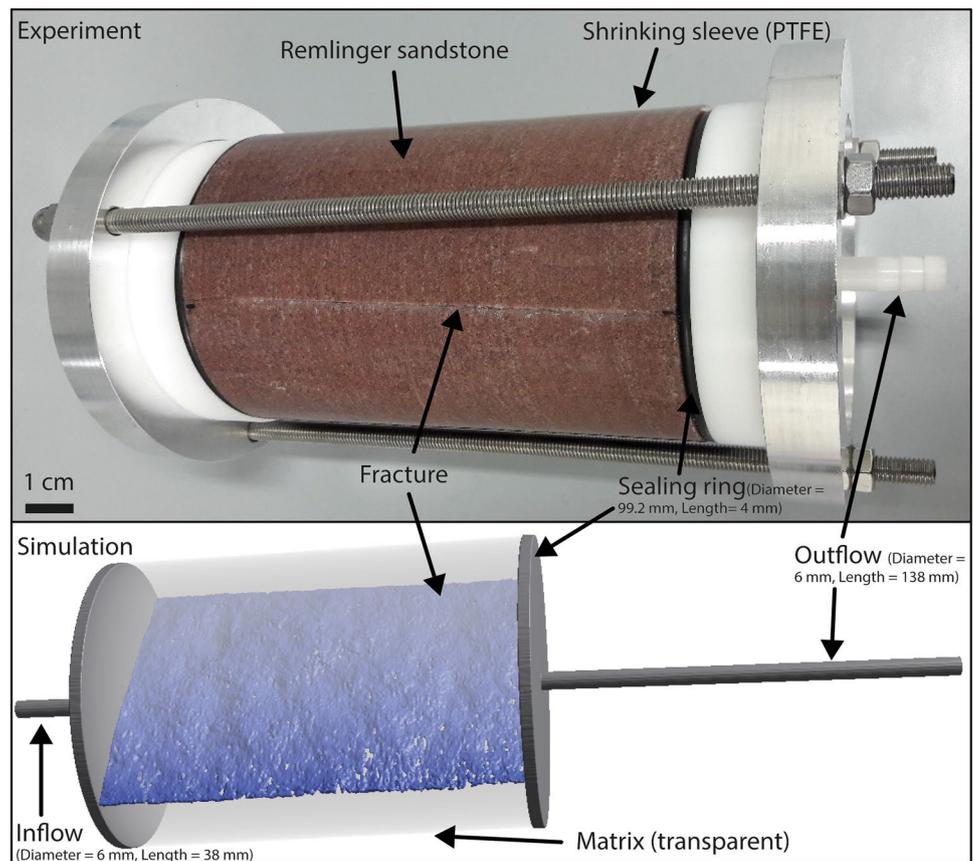


Fig. 4 Illustration of the workflow to determine the 3D aperture of a fracture. Dotted lines are the profiles plotted below the images. A distance map (B) is applied on the binary image in A, where all computed values are inverted to obtain a drop-shaped profile (F) instead of a hump-shaped one (E) during the analysis of traverses. On the

image obtained from the traverse analysis, 2D (C) and 3D (G) local orientations of each FTP are then used to convert the recorded apparent width to the true one (orthogonal to the fracture). At the end, an image showing the aperture of the vertically derived FTPs is returned (D and H)

Fig. 5 Top: A sandstone core built into a flooding cell for conducting the water injection and tracer experiments. Bottom: the translation of the experimental setup into the model. The length of the inflow and outflow tubes of the model corresponds to the distance in the experiment between the core and tracer injection (inflow) and between the core and the measurement location of electric conductivity (outflow)



Water injection and tracer experiments

The core samples were covered with a thin-walled shrinking tube (Polytetrafluoroethylene, PTFE) to avoid fluid escaping during the injection experiments (Fig. 5). The cores were installed in a self-made clamping device. This combination allowed the flow through the fracture during the injection and tracer experiments. The built-in core corresponds to a flooding cell. More details are provided in Frank et al. (2020a, b).

Water (density of 998.234 kg/m^3 and dynamic viscosity of 0.001 kg/ms at a temperature of $20 \text{ }^\circ\text{C}$) flowed from bottom to top, where it was collected and measured at regular intervals. The calculated flow rate (0.3162 mL/s) was used to determine the permeability and the mean hydraulic aperture of the fractures. To ensure the validity of Darcy's law, a laminar flow (determined by the Reynolds number) was applied. The mean hydraulic fracture apertures were calculated using the cubic law (Tsang 1992; Becker 2008) and the intrinsic permeabilities, K , were calculated by utilizing the relation with the mean hydraulic fracture apertures (de Marsily 1986).

The tracer tests were conducted to determine the transport behavior of the fractures. 1 mL of a 2 molar NaCl solution, acting as the tracer, was injected during one second as a

tracer pulse. To receive a breakthrough curve, the electrical conductivity (EC) was measured each second interval at the outflow with a WTW Multi 3510 conductivity logger and its associated Tetracon 925c conductivity sensor. A detailed description of the procedure and the results of the experiments for both sandstones are given in Frank et al. (2020a, b). To compare the experimental results with the simulations, the EC was also converted into a fractional particle count in the following way. First, the measured EC was corrected by subtracting the background EC from all values. Afterwards, the corrected EC values for each timestep were divided by the sum of corrected EC values of all timesteps.

Permeability and particle migration simulations

The fracture permeability and flow simulations were calculated on the segmented datasets (Fig. 5) with the Left-Identity-Right (LIR) solver by GeoDict2020 (Linden et al. 2015). The LIR solver invokes an iterative finite volume with an adaptive multigrid approach to apply Navier–Stokes's conservation of momentum equation,

$$-\eta \Delta \vec{u} + (\rho \vec{u} \times \nabla) \vec{u} + \nabla p = \vec{f}, \quad (5)$$

where η is the fluid viscosity, \vec{u} the 3D fluid flow velocity field, p is the pressure and \vec{f} is the force density field. Equation 5 is recommended for fluid flow with a low Reynolds number. The calculated force density field is used in Darcy's law to calculate the permeability, K , of the medium (Eq. 6., Darcy 1856),

$$\vec{u} = -\frac{K}{\eta} (\nabla p - \vec{f}). \quad (6)$$

The 500 preferred flow paths in each segmented dataset were calculated by GeoDict 2020. The symmetric boundary conditions and the maximum diameter option were used during the calculations. The algorithm searches for paths through the structure for the biggest possible sphere and iteratively reduces the diameter of spheres, until 500 paths through the structure are found (Wiegmann and Glatt 2021b).

To accurately simulate the tracer and water injection experiments ("Water injection and tracer experiments"), the connecting inflow and outflow areas (sealing rings and tubes) of the setup were added as pore voxels to the segmented fractures (Fig. 5). The size of the tubes was chosen based on the actual distances between the tracer source and the downstream EC logger to the core.

The fracture porosities were calculated by dividing the number of fracture voxels by the number of voxels, that are present in a cube with the same volume as the real samples. The fracture flow fields and permeabilities were calculated with the periodic boundary conditions in the flow direction. In the tangential directions, the no-slip boundary condition was applied. The simulation parameters were set to be identical to the experimental conditions. An error bound of 5% was used as a convergence stopping criterion, i.e., the iteration stops when the relative difference to a predicted permeability within the last 100 iterations is smaller than 5%. The matrix was considered as impermeable (see "Rock samples and XCT scanning"). Fracture permeability was computed by dividing the permeability over the whole simulated domain by the fraction of fracture voxels.

The tracer experiments were simulated in GeoDict2020. To reproduce the experimental conditions, infinitesimally small particles (= molecules) were considered. Particles did not lose their energy when colliding with solid material (Wiegmann and Glatt 2021a). The diffusive motion is based on Eq. (7) and the Stokes–Einstein relation (Eq. 8; Einstein 1956),

$$\Delta x = \sqrt{2D \times t}, \quad (7)$$

$$D = \frac{k_B T}{\gamma}, \quad (8)$$

with the distance traveled by a particle (x), diffusivity (D), time (t), Boltzmann constant (k_B), temperature (T) and the friction coefficient (γ).

The molecular particle motion (\vec{v}) is calculated per time step, dt , by adding the flow velocity (\vec{u}) with random Brownian motion, neglecting electrostatic forces, external forces, and diffusion into the rock matrix (Eq. 9). The diffusive motion is simulated by applying a random walker combining the 3D wiener measure ($\vec{W}(t)$) and (Eq. 7),

$$\vec{v} = \vec{u} + \sqrt{2D} \frac{d\vec{W}(t)}{dt}. \quad (9)$$

The periodic boundary condition was applied in the tangential direction, reflective at the lower end of the domain and the open boundary condition in the upper end of the domain (i.e., particles were able to leave the domain at the outflow). 600,000 particles were generated uniformly during one second at the inflow (the same as the input pulse in the tracer experiment), with a starting velocity equal to the injection velocity. The simulated time was 3,000 s with a time step of 1 s, which corresponds to the measuring interval in the tracer experiments.

Results and discussion

Conventional segmentation results

The gray value histograms of both samples show a complex distribution between the two phases without distinct peaks (Fig. 1). This behavior is expected to be challenging for the thresholding techniques (Al-Marzouqi 2018). Since the segmentation methods of both samples show the same behavior, only the Remlinger sample is discussed here.

Comparing the threshold segmentation results of the Remlinger sample (Fig. 6A–C) highlights the problems of implementing such a simple algorithm. With a lower threshold (3350–3380), large parts of the fracture, especially near the edges, are not segmented as a fracture. The higher thresholds (3400–3440) gradually detect more fracture volume but fail to segment the whole fracture volume correctly. Additionally, raising the threshold leads to more matrix voxels to be falsely segmented as fracture voxels (see Fig. 2D). Removing these errors led to a drop in porosity of 34.8% (from 1.81 to 1.18%) in the Remlinger threshold of 3420 (Fig. 6B, C), highlighting the substantial number of errors present in this approach. These false voxels start to merge into the fracture volume at threshold value of 3420 and cannot be removed (i.e., irrevocably altering the fracture structure). For the thresholding technique, a trade-off emerged between

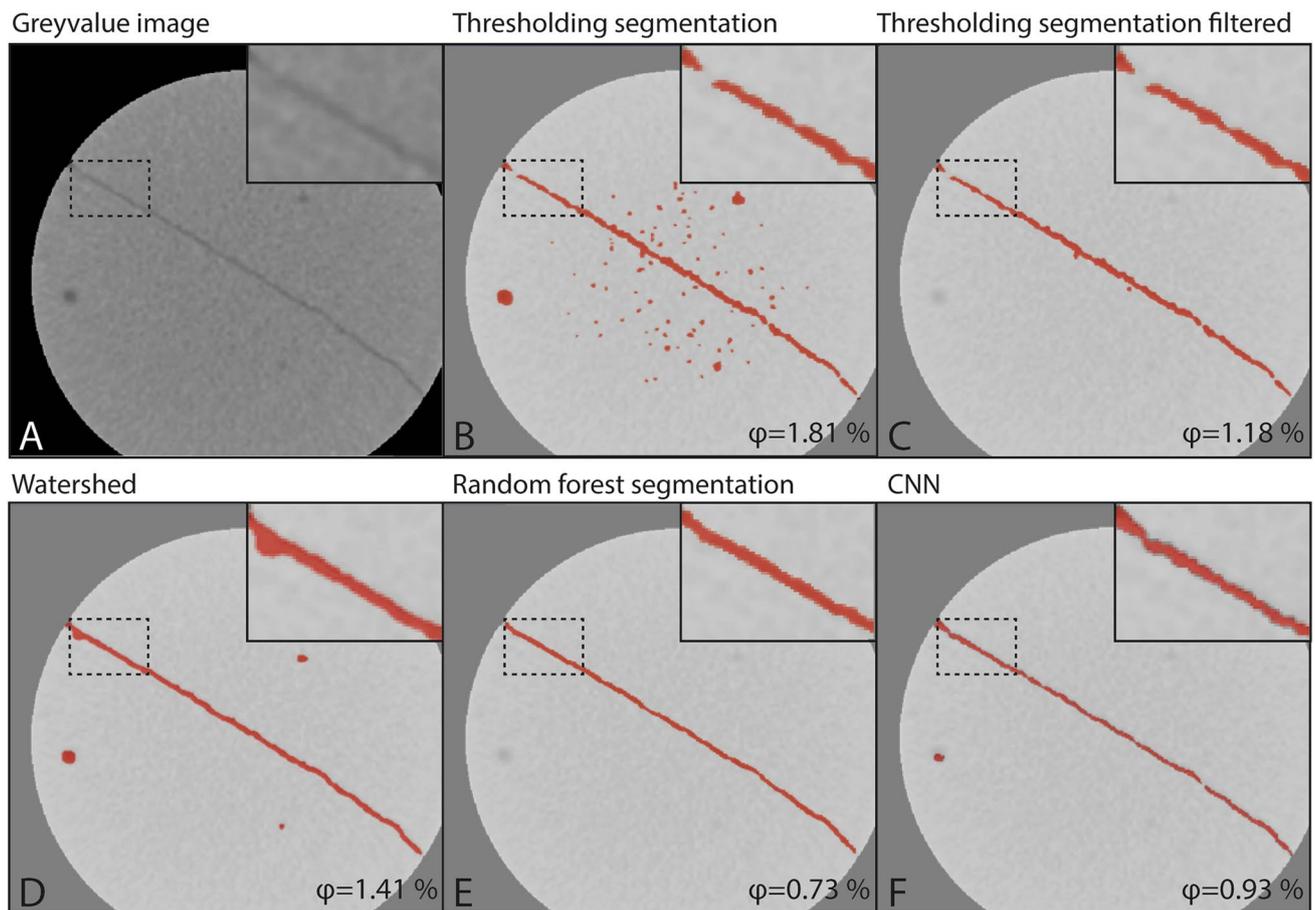


Fig. 6 Comparison of the applied segmentation techniques on the Remlinger sample with the fracture porosity φ . **A** Slice 130 of the grayscale XCT image. **B** Segmented fracture using the threshold method with the range 10–3420. **C** Segmented fracture shown in **B**

after applying the “cleanse tool”. Segmented results by the **(D)** watershed segmentation, **E** RF classifier, and **F** CNN. The machine learning algorithms were not trained on this slice

segmenting numerous fracture voxels (high threshold values) and minimizing errors (low threshold values, also see Figs. 8, 9).

The thresholds calculated by the Otsu and K-means clustering methods for the Remlinger sample were not considered in the analysis because of the extensive errors in these segmentations, which prevented a meaningful segmentation of the fracture.

The watershed algorithm was capable of segmenting much of the fracture volume correctly. However, fracture voxels were inaccurately separated from neighboring matrix voxels with similar gray values, leading to a low gradient at the fracture boundary (Fig. 6D). Furthermore, the algorithm did not segment large parts of the fracture, especially near the edges (Figs. 8, 9). The artifacts produced by the cone-beam measurement were segmented by both conventional approaches (Fig. 2D).

Machine learning segmentations

After providing proper training data to the RF classifier (Table 1), it avoided most of the problems occurring in the conventional methods. The whole fracture volume was segmented for both samples (Figs. 6E, 8, 9). User bias was minimized due to the constant interaction between the user and the RF classifier, which provided preliminary segmentation results and uncertainty maps. Figure 7C shows an exemplary uncertainty map, which shows that the highest uncertainties of the trained classifier appear at the transition between the phases. The segmentation of the transition voxels highly influences the fracture properties (like aperture and roughness). High uncertainties occur due to, e.g., the partial volume effect. Therefore, it is crucial to provide the RF classifier with proper training data at the fracture boundaries to ensure a correctly segmented fracture despite

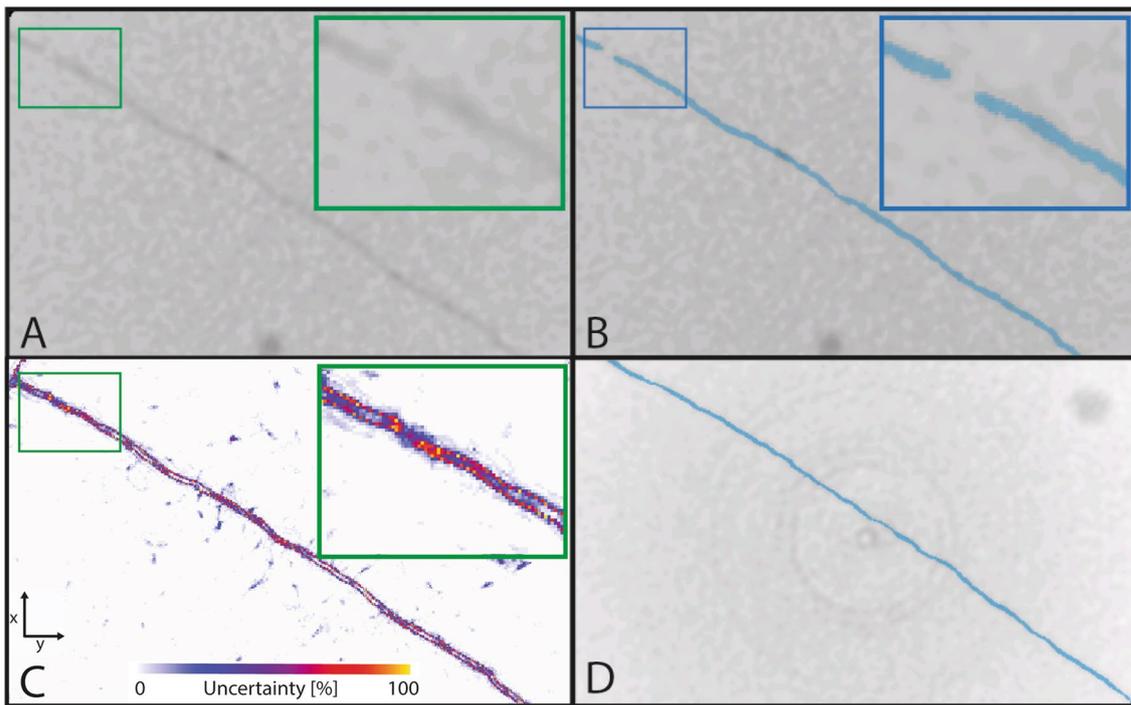


Fig. 7 Segmentation results of RF classifier. The classifier was not trained on these slices. **A** Slice 273 of the gray value image showing an interruption in the Remlinger fracture. **B** Segmented fracture is shown in **A**. **C** Uncertainty map of the segmentation shown in **B**.

D Segmented fracture voxels (blue) with the underlying gray value 2D image in slice 760, in which the RF classifier was able to correctly segment the fracture despite of the availability of ring artifacts

Table 2 Loss and accuracy values

	Training I		Training II	
	Training	Validation	Training	Validation
Loss	0.0043	0.0037	0.0030	0.0032
Accuracy	0.9975	0.8326	0.9982	0.8873

In Training I, CNN was trained on the Remlinger sample and the validation was applied on the Flechtinger sample. In Training II, the training and validation datasets of Training I were substituted

the high uncertainty at the phase boundaries. Moreover, the RF classifier was able to segment the cone-beam artifacts correctly (Fig. 7D), which was impossible with the conventional methods. Cavities in the fracture area were interpreted as locations where the fracture aperture becomes too small to be detected by the XCT scanner or as contact areas of the two opposing fracture surfaces (Fig. 7A, B).

Table 2 summarizes the final accuracy and loss functions of both trainings processes applied during the CNN segmentation. Like the RF classifier, the CNN provided a superior segmentation compared to the conventional methods (Fig. 6F). The results indicate that the segmented images have around 83–89% IoU similarity with the ground truth used during the CNN training phases. It means that upon

completing the training phase, the CNN network can be applied to other validation images to segment fractures. The training phase required approximately 10 h which should be implemented once. However, the validation process only takes few minutes on new image datasets. This method was also able to correctly segment the cone-beam and ring artifacts (see “Conventional segmentation methods”).

Fracture properties

Roughness, mean mechanical aperture, porosity and permeability were calculated for each segmentation on both rock samples (Table 3). The fracture properties showed distinct deviations, depending on the method used. However, different segmentations showed similar permeability and porosity values: e.g., for the Flechtinger sample, the machine learning segmentations resulted in similar permeability and porosity values as the thresholding at 3380 and 3400. Since the segmented fractures visually showed obvious differences in segmentation quality (Fig. 6), these parameters are not sufficient to quantify the quality of the different segmentation approaches. For this reason, flow fields, preferred flow paths, and breakthrough curves were calculated and compared with the experimental results as a ground truth.

Table 3 Results of the flow simulations and fracture analysis

Segmentations	Permeability $\times 10^{-8}$ [m ²]	Porosity (fracture portion) [%]	Roughness [-]	Mean mechanical aperture [μ m]	STD [μ m]
Flechtinger					
Thresholding at 3360	3.32	0.90	0.38	885	331
Thresholding at 3380	4.15	0.99	0.36	935	336
Thresholding at 3400	5.05	1.09	0.35	985	343
Thresholding at 3410	5.45	1.15	0.34	1003	339
Thresholding at 3430	6.55	1.26	0.33	1017	354
Thresholding at 3440 (K-means clustering)	7.07	1.31	0.33	1065	361
Thresholding at 3450	7.48	1.39	0.37	1082	399
Thresholding at 3470	8.41	1.55	0.43	1089	464
Thresholding at 3490	8.96	1.79	0.57	1092	576
Watershed	10.20	1.44	0.22	1201	269
RF classifier	4.67	1.07	0.22	827	179
CNN	4.19	1.03	0.22	816	177
Remlinger					
Thresholding at 3350	0.65	0.53	0.41	692	286
Thresholding at 3380	1.92	0.78	0.39	805	314
Thresholding at 3400	3.11	0.98	0.39	887	346
Thresholding at 3410	3.79	1.08	0.39	929	370
Thresholding at 3420	4.53	1.18	0.41	973	396
Thresholding at 3430	5.28	1.31	0.44	1009	443
Thresholding at 3440	6.00	1.46	0.49	1032	513
Watershed	6.91	1.41	0.76	1248	952
RF classifier	2.24	0.73	0.22	571	126
CNN	2.98	0.93	0.26	787	206

Flechtinger sample

Figure 8 depicts the segmented fractures, calculated flow fields and 500 preferred flow paths through the fractures for each segmentation method. The low threshold values were not able to segment the upper fracture section (in the x-direction), whereas the high thresholds resulted in significant biases near the inflow (see also Fig. 6). The higher the threshold, the more evenly spread is the flow (and the preferred flow paths) and the higher the flow velocity is. The watershed segmentation was also unable to segment the upper fracture part, resulting in the concentration of the flow in the middle and lower parts of the fracture. On the other hand, both the RF and CNN segmentation were able to segment the whole fracture, resulting in an evenly distributed flow field and flow paths.

Remlinger sample

The outer sections (x-direction) of the low thresholding fractures were not connected to the middle sections of the fracture, focusing the flow in the Remlinger sample (Fig. 9).

The higher the threshold was chosen, the wider was the connected fracture section, with higher the flow velocities and wider spreading of the preferred flow paths. The main flow paths remained concentrated in the middle. The watershed segmentation showed high fluid velocities with preferred flow paths distributed over the middle and lower parts of the fracture. Due to large areas of falsely segmented voxels, the fracture sections near the edges were truncated with impermeable matrix clusters, which led to the convergence of the flow paths.

On the other hand, the segmentation by the RF classifier resulted in a more continuous fracture area with smaller and fewer impact of matrix voxels. Accordingly, the water flowed through a wider area, and the preferred flow paths spread more evenly across the fracture. The CNN structure showed a continuous fracture, but with more matrix voxels (i.e., areas where both fracture sides are in contact) than the RF segmentation. This resulted in a fewer but more concentrated flow path with locally higher flow velocities.

The differences in fracture parameters for both rock samples highlights the enormous influence of the segmentation method on the fracture structure. Meaningful

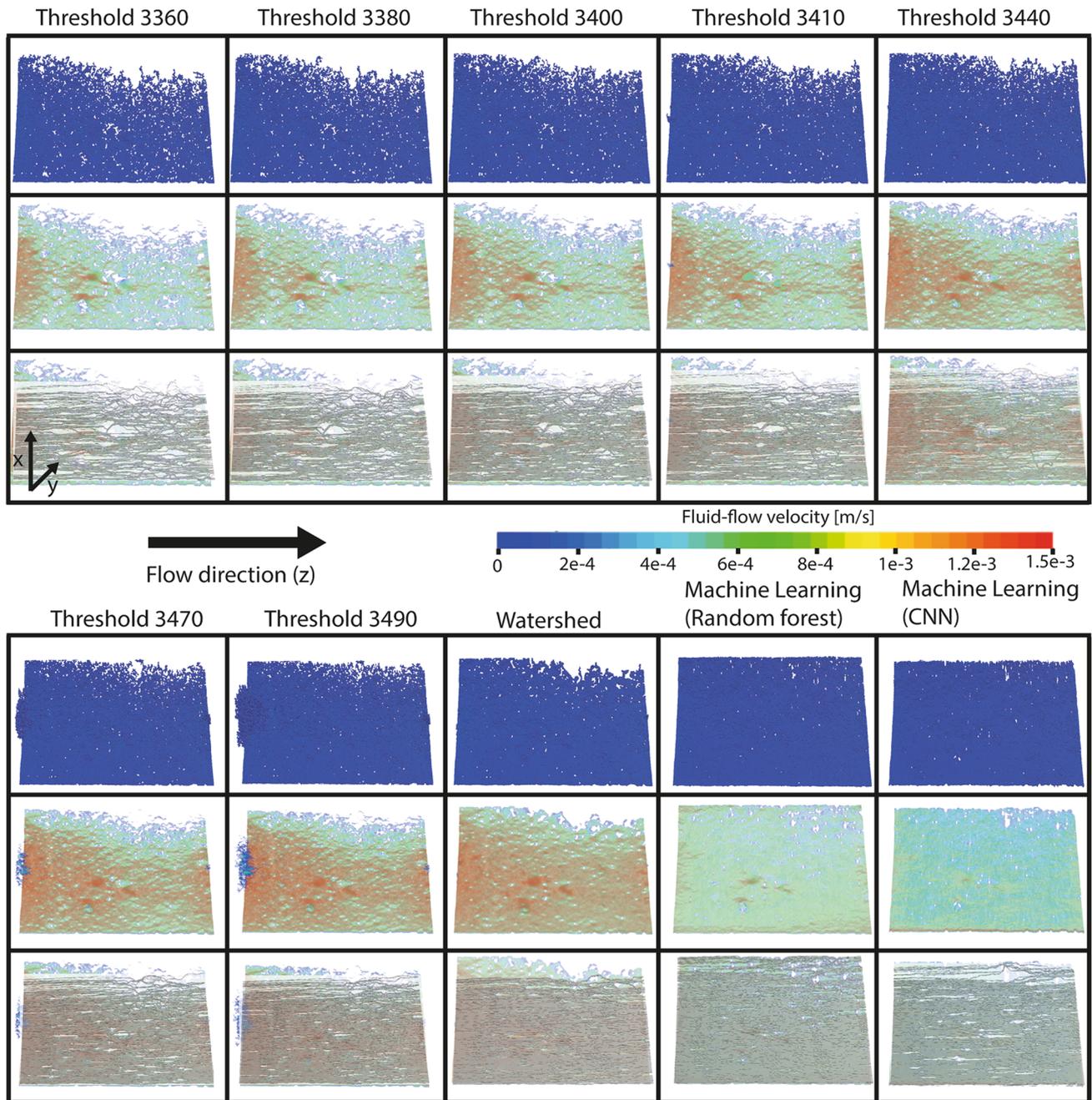


Fig. 8 Comparison of the segmented structures (top), flow fields (middle) and preferred flow paths (bottom) of each Flechtinger segmentation result. Inflow and outflow areas are not shown

segmentation is crucial to obtain a representative digital model of a fracture and credible modeling outcomes. Permeability and porosity alone are insufficient to compare the quality of the segmented fractures quantitatively, e.g., both parameters of the RF segmentation and the thresholding 3400 segmentation are similar for the Flechtinger sample; however, the structure, computed flow fields, and

preferred flow paths show significant variations. To fully assess the quality and reliability of the calculated fracture properties for each segmentation, this study focuses on the comparison of breakthrough curves from tracer experiments and simulations to quantify the quality of the different segmentation results.

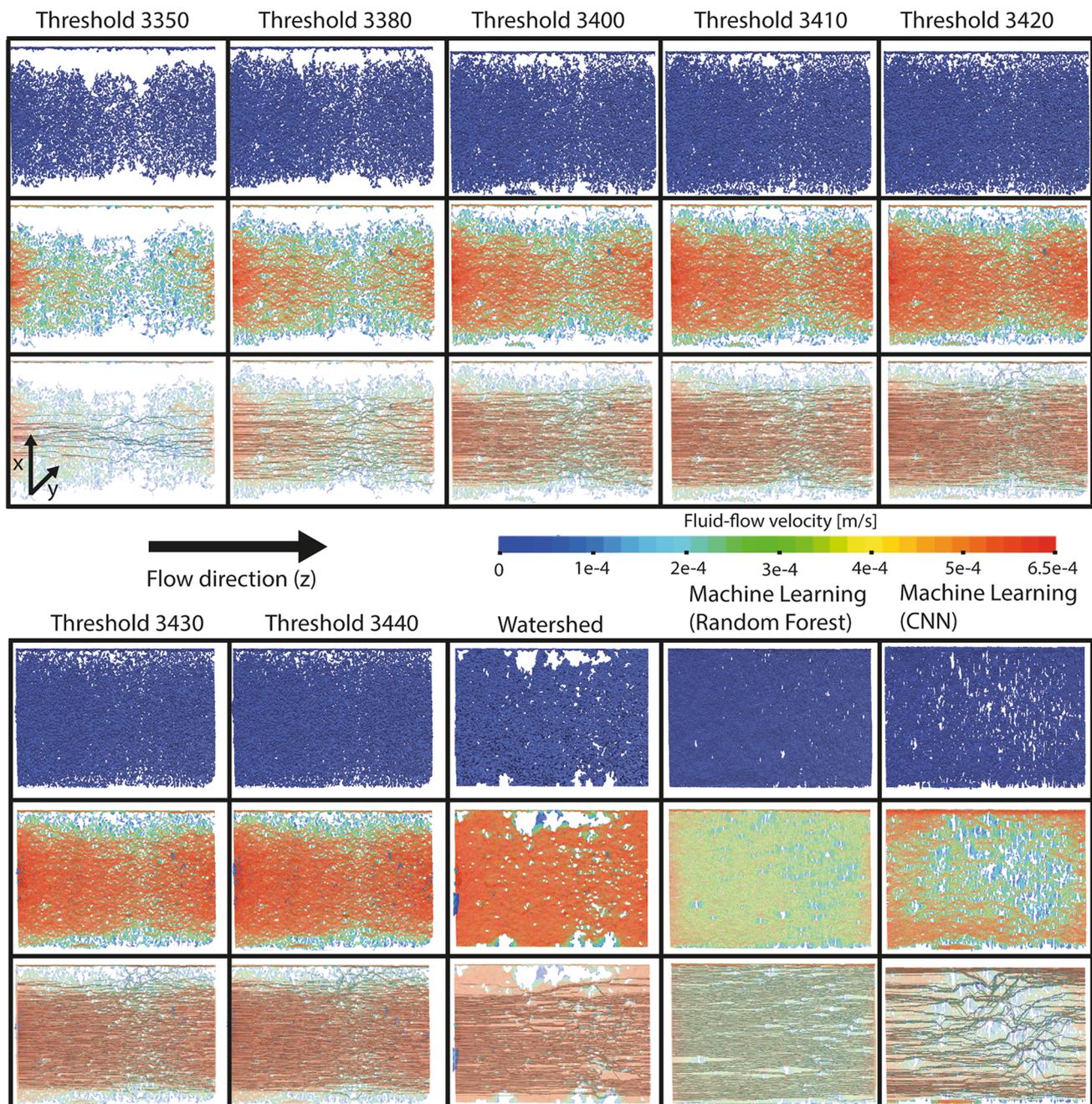


Fig. 9 Comparison of segmented structure (top) computed flow fields (middle) and preferred flow paths (bottom) of each segmentation result. Inflow and outflow areas are not shown

Tracer simulations

Flechtinger sample

The experimental tracer results show the first breakthrough of particles at 132 s and a maximum peak height of effluent particle fractions at the 0.6% (Fig. 10). None of the thresholding structures was able to reproduce both breakthrough

and similar maximum peak height. The thresholding structures showed either a similar time for the first breakthrough (thresholding at 3360, deviation of 14.4% from the tracer experiment) or a similar maximum peak height (thresholding at 3410, deviation of 0.2% from the tracer experiment, Fig. 10). The watershed segmentation was also not able to reproduce the experimental breakthrough results. On the other hand, the RF segmentation resulted in a breakthrough

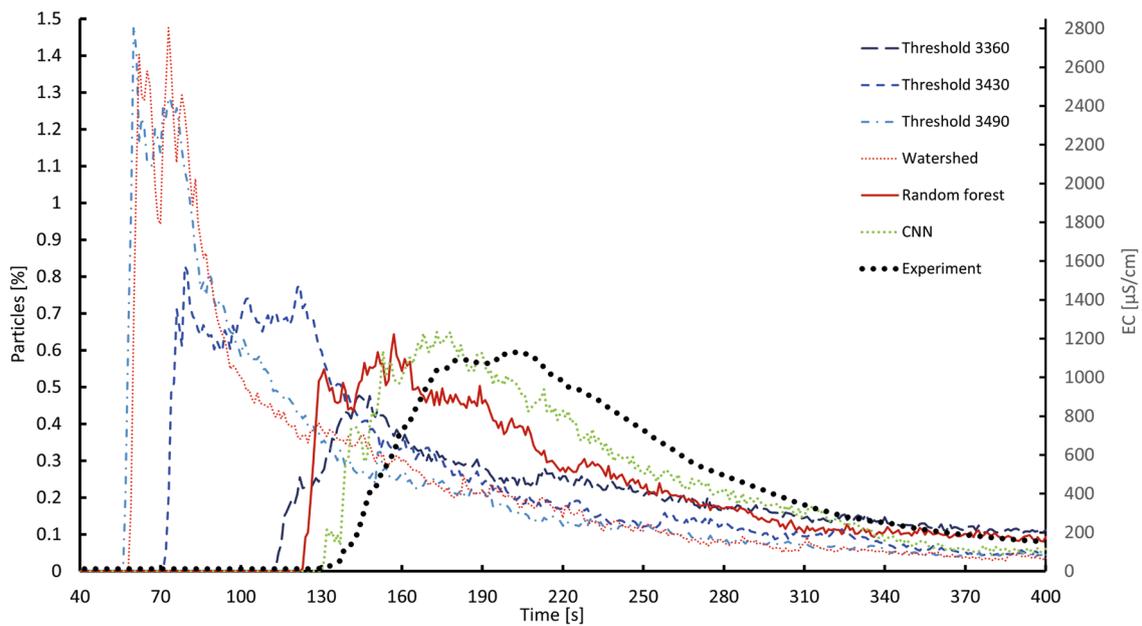


Fig. 10 Comparison between simulated breakthrough curves and experimentally measured breakthrough curves. For a better overview, only three breakthrough curves from the thresholding approach are shown

at 125 s and a peak height of 0.58%, which is in close agreement with the experimental data (error < 5%). Furthermore, this simulation was able to reproduce the tailing of the tracer

curve at later time steps. The experimental breakthrough curves were also reproduced by the CNN structure (with a peak height of 0.65% and a breakthrough time of 135 s).

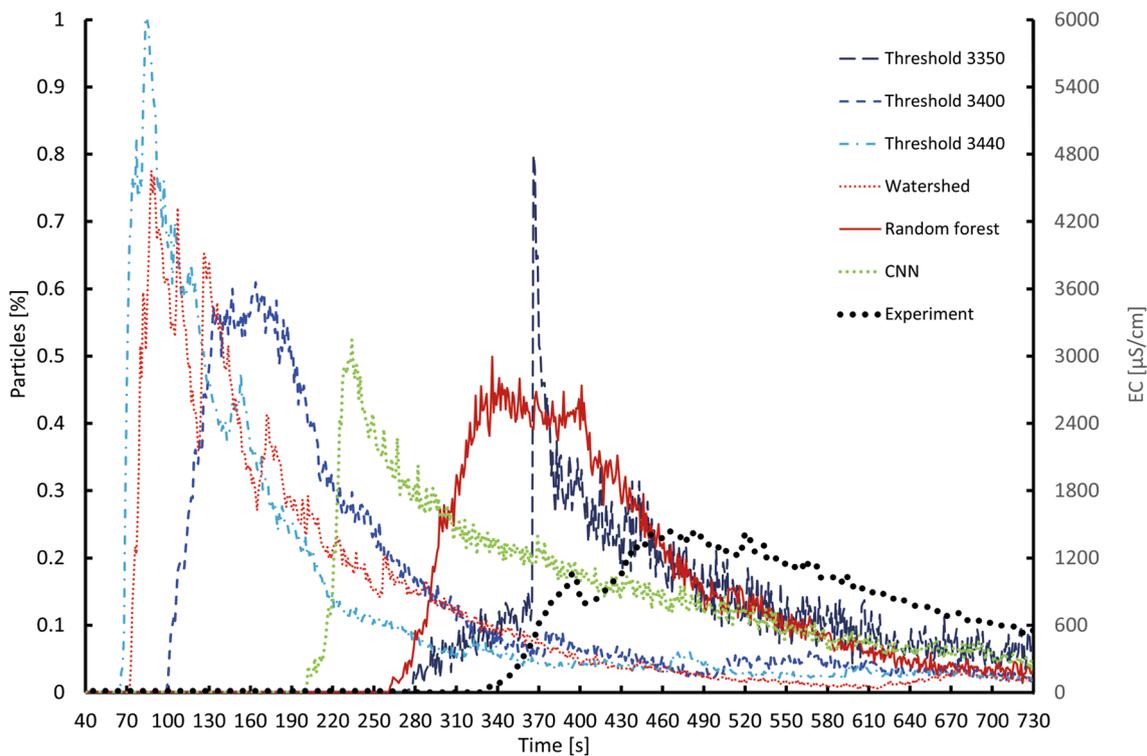


Fig. 11 Comparison between simulated breakthrough curves and experimentally measured breakthrough curves. For a better overview, only three breakthrough curves from the thresholding approach are shown

Remlinger sample

The particle breakthrough was at 335 s with a peak height of 0.23% for the Remlinger fracture. The simulated particle migrations on the watershed and threshold structures showed breakthrough times in the range of 69–274 s (79.4–18.2% earlier than the tracer experiment) and peak heights of 1.00–0.78% (334–75% more than the tracer experiment, Fig. 11). Generally, higher thresholds in the segmentation showed earlier first breakthroughs and higher peaks (Fig. 11). The conventional methods were not able to reproduce the experimental tracer results and showed great deviations from both breakthrough time and peak height. The particle migration simulations on the machine learning segmentations did not reproduce the experimental data. However, it resulted in a significantly closer breakthrough curve (in terms of breakthrough time, peak height, and curve tailing), as the other methods. The CNN structure showed significant deviations from the experimental curves (Fig. 11).

Discussion on segmentation results

The flow fields, preferred flow paths, and the visualization of the segmented fractures (in 2D and 3D) demonstrated the benefit of the machine learning-based segmentation. The threshold algorithms failed to segment significant parts of the fractures, especially near the boundaries and showed issues in distinguishing between fracture and matrix voxels with a similar gray value. The latter error was particularly troubling when these matrix voxels (segmented as fracture voxels) lied in the vicinity of the fracture thus merged into it. The watershed algorithm showed similar problems in which significant fracture parts were segmented as belonging to the matrix voxels. Moreover, the fracture neighboring matrix voxels with a similar gray value were erroneously segmented as fracture voxels. These errors resulted in a higher fracture porosity, higher mean mechanical aperture, and flawed flow fields. The machine learning approaches (CNN and RF) showed a considerably better segmentation performance without the above-discussed biases. The whole fractures were segmented with small intrusions of matrix voxels, which were interpreted as contact areas between both fracture sides or as narrow openings, which the XCT scanner could not detect. Small and few clusters (islands) of falsely segmented voxels occurred in these methods and were easily removed (see “Conventional segmentation methods”). The uncertainties maps of the RF classifier showed that proper training by an experienced user is crucial at the phase transitions. Through the constant feedback with the RF classifier, user bias is reduced during training.

The particle transport simulations reveal the superior segmentation of the Flechtinger sample, provided by both machine learning approaches. Contrarily to the conventional

methods, the experimental breakthrough curve was reproduced in terms of first breakthrough time, peak height, and tail by both implemented machine learning approaches. Since these methods dealt better with the cone-beam artifacts, it was not necessary to crop the upper and lower parts of the dataset as much as in the segmentation of the conventional methods. Therefore, a more representative fracture model could be achieved.

The simulated breakthrough curves for the Remlinger sample showed deviations from the experimental data. The RF segmentation led to the closest breakthrough curve; however, no good match could be achieved. The reason may be explained by the difference between fracture aperture and imaging resolution. The experiments (Frank et al. 2020a) revealed that the mean hydraulic aperture of the Remlinger fracture (93.7 μm) is smaller than that of the Flechtinger fracture (143.9 μm) (Figs. 10, 11). Ketcham et al. (2010) discussed that the smaller the fracture aperture is (at a given resolution), the more challenging it is to extract a representative fracture model from the XCT images. One possible reason can be the partial volume effect, which gets more important with smaller apertures, low XCT scan resolutions, and near structure boundaries (Ketcham et al. 2010; Garfi et al. 2020; Yang et al. 2020). This can result in an over- or underestimation of fracture aperture, which can be corrected if sufficient data is available (e.g., Kato et al. 2013; Ramandi et al. 2017; Yang et al. 2020). Since the Remlinger fracture has a lower mean hydraulic aperture than the Flechtinger fracture and the spatial image resolution is low (200 μm), the partial volume effect is assumed to significantly influence the segmentation of the Remlinger fracture more than the other sample. An overestimation of the Remlinger fracture aperture could explain the early first breakthrough and the higher peaks of all simulated breakthrough curves. The Flechtinger segmentations were also expected to suffer from the influence of the partial volume effect. However, because of the larger mean aperture, the influence on the segmentation could be less critical, and it was therefore still possible to reproduce the experimental breakthrough curves with the particle simulations.

One of the advantages of CNN methods over the traditional image processing and thresholding-based methods is efficiently handling artifacts of XCT-images. However, if a sufficient number of training images with artifacts are available in the dataset, the machine learning algorithm learns how to segment them correctly in the prediction phase. Moreover, this method can be applied on XCT-images with low resolution or high noise-to-signal ratio, in which other segmentation methods (like the watershed algorithm) need a prior image pre-processing (e.g., denoising) step (Iassonov et al. 2009; Lee et al. 2021).

Despite these promising advantages, CNNs have some bottlenecks. Determining the optimum network architecture

and parameter tuning is time-consuming and computationally expensive (Lee et al. 2021). Usually, different architectures should be tested to find the best network. Moreover, the selected network parameters (e.g., number of filters, number of batches and epochs, learning rate, optimizer) should be optimized to increase the prediction efficiency. Furthermore, this process is heavily dependent on the availability of enough computational resources. Most of the deep learning libraries are now running on GPUs to boost training time. The availability of enough GPU onboard memory is another challenge that can limit input image sizes or batch numbers. Apart from these limitations, providing a sufficient number of training images and their ground truth is another critical challenge. Although a large training dataset increases the reliability of the network (Wang et al. 2021), the acquisition of such a dataset by imaging techniques can be expensive.

The RF approach is also able to handle complex images (Berg et al. 2018). This method is accessible through open-source software like Ilastik, which provides a pre-adjusted workflow. Additionally, this method does not need a ground truth. Instead, training data is provided, directly on the gray value images by the user. User input and bias are minimized by the preliminary results and uncertainty maps provided during the training phase. The training can be applied on one 2D slice at a time to lower the computational demands. Thus, the RF approach has significant advantages over other methods, especially when working with unique samples. However, since the RF classifier depends on training by the user, a user bias, although minimized, is not eliminated from this method.

Conclusions

In this study, we demonstrated the potential of using the RF classification for segmentation purposes. We compared the segmented digital images of two fractured sandstones using the CNN, several manual and global thresholding methods, and a watershed algorithm. User bias was minimized by the machine learning techniques used. Compared with the conventional segmentation methods (manual thresholding, Otsu method, K-means clustering, and watershed) distinguishing of the different phases was possible with the machine learning approaches, despite the complexity of the gray value histograms.

The physical rock properties of both samples depended on the segmentation method applied, which demonstrates the importance of a reliable segmentation in the DRP workflow. For both fractured samples, a segmentation of the whole fracture area could be achieved using machine learning approaches (RF and CNN). Due to the similar porosity and permeability of different segmentation methods, these parameters were insufficient to quantitate the

segmentation results. Comparing the simulated fluid flow fields, preferred flow paths, and fracture structures qualitatively showed the superior segmentation of the RF classifier. Whereas the conventional segmentation methods were susceptible to various errors, the machine learning segmentations handled all difficulties. Comparing simulated breakthrough curves with experimental data quantitatively showed the superior segmentation performed by the machine learning approaches. Only with these structures it was possible to reproduce the experimental tracer curve of the Flechtinger sample. For the Remlinger fracture, breakthrough curves could not be reproduced with the same precision. However, the RF classifier still performed the best among all tested methods. Furthermore, because the machine learning approaches handled the cone-beam artifacts well, it was possible to extract a larger structure (in Z-direction) as close as possible to the acquired XCT image. Furthermore, the simulated breakthrough curves of these structures increased the match with the experimental data and highlighted that a more realistic structure was obtained.

Although the machine learning segmentations showed similar quality results, the RF classifier had inherent advantages over the CNN in terms of computational demands, accessibility, and training data. However, CNNs can potentially surpass the RF approach if a comprehensive database with sufficient training data and corresponding ground truths (free of user bias) could be provided in the future. To fully assess the reliability of the machine learning segmentation, future studies should validate fracture parameters with additional laboratory tests or implementing scans at high resolution.

Acknowledgements This work was supported by the German Federal Ministry for Economic Affairs and Energy (BMWi). It is part of the project ReSalt (Reactive Reservoirsystems – Scaling and Erosion and its Impact on Hydraulic and Mechanic Reservoirproperties, grant number: 0324244A). We would like to thank Dustin Hering, Rainer Seehaus and Rafael Schäffer from the TU Darmstadt for providing the fractured cores. We thank Olga Moravcova for her contribution to the first manuscript. The third author (S. S.) gratefully acknowledges financial support from the Alexander von Humboldt Foundation for his visiting research at the Johannes Gutenberg University at Mainz, Germany. We thank Olaf Kolditz and the two anonymous reviewers for their valuable suggestions that helped to improve our manuscript.

Author contributions Conceptualization: MR, AJ, SS, FC, PA, FE and MK; Methodology: MR, AJ; Software: MR, SS, AJ, FC, FE; Validation: MR, SS, AJ, FC, PA, FE and MK; Formal analysis and Investigation: MR, SS, AJ, FC, PA, SF; Writing—original draft preparation: MR, SS, AJ; Writing—review and editing: MR, AJ, SS, FC, PA, SF, FE and MK; Visualization: AJ, MR; Resources: PA, FE and MK; Data curation: MR, AJ, SS, FC, PA, SF, FE and MK; Supervision: FE and MK; Project administration: FE and MK; Funding acquisition: FE and MK.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the German Federal Ministry

for Economic Affairs and Energy (BMWi). It is part of the project ReSalt (Reactive Reservoirsystems—Scaling and Erosion and its Impact on Hydraulic and Mechanic Reservoirproperties, grant number: 0324244A). This open-access publication was funded by Johannes Gutenberg University Mainz.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Al-Marzouqi H (2018) Digital rock physics: using CT scans to compute rock properties. *IEEE Signal Process Mag* 35:121–131. <https://doi.org/10.1109/MSP.2017.2784459>
- Andrá H, Combaret N, Dvorkin J et al (2013a) Digital rock physics benchmarks—part II: computing effective properties. *Comput Geosci* 50:33–43. <https://doi.org/10.1016/j.cageo.2012.09.008>
- Andrá H, Combaret N, Dvorkin J et al (2013b) Digital rock physics benchmarks—Part I: imaging and segmentation. *Comput Geosci* 50:25–32. <https://doi.org/10.1016/j.cageo.2012.09.005>
- Arganda-Carreras I, Kaynig V, Rueden C et al (2017) Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics* 33:2424–2426. <https://doi.org/10.1093/bioinformatics/btx180>
- Babel H, Naranjo-Meneses P, Trauth S et al (2020) Ratiometric population sensing by a pump-probe signaling system in *Bacillus subtilis*. *Nat Commun* 11:1176. <https://doi.org/10.1038/s41467-020-14840-w>
- Becker TA (2008) Zur Strömungssimulation in Einzelklüften: Gegenüberstellung von numerischen Methoden und Experiment; Dissertation; Rheinisch-Westfälische Technische Hochschule Aachen: Aachen, Germany.
- Berg S, Saxena N, Shaik M, Pradhan C (2018) Generation of ground truth images to validate micro-CT image-processing pipelines. *Lead Edge* 37:412–420. <https://doi.org/10.1190/tle37060412.1>
- Berg S, Kutra D, Kroeger T et al (2019) Ilastik: interactive machine learning for (bio)image analysis. *Nat Methods* 16:1226–1232. <https://doi.org/10.1038/s41592-019-0582-9>
- Beucher S, Meyer F (1993) The morphological approach to segmentation: the watershed transformation. In: Dougherty ER (ed) *Mathematical morphology in image processing*. Marcel Dekker Inc., New York, pp 433–481
- Blöcher G, Zimmermann G, Milsch H (2009) Impact of poroelastic response of sandstones on geothermal power production. *Pure Appl Geophys* 166:1107–1123. <https://doi.org/10.1007/s00024-009-0475-4>
- Blöcher G, Kluge C, Milsch H et al (2019) Permeability of matrix-fracture systems under mechanical loading: constraints from laboratory experiments and 3-D numerical modelling. *Adv Geosci* 49:95–104. <https://doi.org/10.5194/adgeo-49-95-2019>
- Blunt MJ, Bijeljic B, Dong H et al (2013) Pore-scale imaging and modelling. *Adv Water Resour* 51:197–216. <https://doi.org/10.1016/j.advwatres.2012.03.003>
- Bodin J, Delay F, de Marsily G (2003) Solute transport in a single fracture with negligible matrix permeability: I. Fundamental mechanisms. *Hydrogeol J* 11:418–433. <https://doi.org/10.1007/s10040-003-0268-2>
- Brady EL, Kirby MA, Olszewski E et al (2020) Guided vascularization in the rat heart leads to transient vessel patterning. *APL Bioeng* 4:016105. <https://doi.org/10.1063/1.5122804>
- Breiman L (2001) Random forest. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Cappuccio F, Toy V, Mills S, Adam L (2020) Three-dimensional separation and characterization of fractures in X-ray computed tomographic images of rocks. *Front Earth Sci*. <https://doi.org/10.3389/feart.2020.529263>
- Chauhan S, Rühaak W, Khan F et al (2016) Processing of rock core microtomography images: using seven different machine learning algorithms. *Comput Geosci* 86:120–128. <https://doi.org/10.1016/j.cageo.2015.10.013>
- Cnudde V, Boone MN (2013) High-resolution X-ray computed tomography in geosciences: a review of the current technology and applications. *Earth-Sci Rev* 123:1–17. <https://doi.org/10.1016/j.earscirev.2013.04.003>
- Darcy H (1856) *Les Fontaines Publiques de la Ville de Dijon*. Dalmont, Kraljevica
- De Marsily G (1986) *Quantitative hydrogeology-groundwater hydrology for engineers*, 1st edn. Academic Press, Orlando (ISBN 0-12-208915-4)
- Einstein A (1956) *Investigations on the theory of the Brownian movement*. Courier, New York
- Eyolfson E, Yamakawa GR, Griep Y et al (2020) Examining the progressive behavior and neuropathological outcomes associated with chronic repetitive mild traumatic brain injury in rats. *Cereb Cortex Commun*. <https://doi.org/10.1093/texcom/tgaa002>
- Frank S, Heinze T, Ribbers M, Wohnlich S (2020a) Experimental reproducibility and natural variability of hydraulic transport properties of fractured sandstone samples. *Geosciences* 10:458. <https://doi.org/10.3390/geosciences10110458>
- Frank S, Heinze T, Wohnlich S (2020b) Comparison of surface roughness and transport processes of sawed, split and natural sandstone fractures. *Water (switzerland)*. <https://doi.org/10.3390/w12092530>
- Garfi G, John CM, Berg S, Krevor S (2020) The sensitivity of estimates of multiphase fluid and solid properties of porous rocks to image processing. *Transp Porous Media* 131:985–1005. <https://doi.org/10.1007/s11242-019-01374-z>
- Hassanzadegan A, Blöcher G, Zimmermann G, Milsch H (2012) Thermoelastic properties of Flechtinger sandstone. *Int J Rock Mech Min* 49:94–104. <https://doi.org/10.1016/j.ijrmms.2011.11.002>
- Heiland J (2003) Permeability of triaxially compressed sandstone: influence of deformation and strain-rate on permeability. *Pure Appl Geophys* 160:889–908. <https://doi.org/10.1007/PL00012571>
- Hinton G, Srivastava N, Swersky K (2012) Neural networks for machine learning: Lecture 6a—Overview of mini-batch gradient descent. Available at https://www.cs.toronto.edu/~tjmen/csc321/slides/lecture_slides Lec6.pdf

- Hinz C, Enzmann F, Kersten M (2019) Pore scale modelling of calcite cement dissolution in a reservoir sandstone matrix. *E3S Web Conf* 98:05010. <https://doi.org/10.1051/e3sconf/20199805010>
- Hiramatsu Y, Hotta K, Imanishi A et al (2018) Cell image segmentation by integrating multiple CNNs. *IEEE Comput Soc Conf Comput vis Pattern Recogn Work*. <https://doi.org/10.1109/CVPRW.2018.00296>
- Ho TK (1994) Random decision forests. *Proc Int Conf Doc Anal Recogn* 1:278–282. <https://doi.org/10.1109/ICDAR.1995.598994>
- Hong J, Liu J (2020) Rapid estimation of permeability from digital rock using 3D convolutional neural network. *Comput Geosci* 24:1523–1539. <https://doi.org/10.1007/s10596-020-09941-w>
- Hounsfield GN (1980) Nobel lecture, 8, December 1979. *Comput Med Imaging J Radiol* 61:459–468
- Iassonov P, Gebrenegus T, Tuller M (2009) Segmentation of X-ray computed tomography images of porous materials: a crucial step for characterization and quantitative analysis of pore structures. *Water Resour Res*. <https://doi.org/10.1029/2009WR008087>
- Kaestner A, Lehmann E, Stampanoni M (2008) Imaging and image processing in porous media research. *Adv Water Resour* 31:1174–1187. <https://doi.org/10.1016/j.advwatres.2008.01.022>
- Karimpouli S, Tahmasebi P, Saenger EH (2020) Coal cleat/fracture segmentation using convolutional neural networks. *Nat Resour Res*. <https://doi.org/10.1007/s11053-019-09536-y>
- Kato M, Takahashi M, Kawasaki S et al (2013) Evaluation of porosity and its variation in porous materials using microfocus x-ray computed tomography considering the partial volume effect. *Mater Trans* 54:1678–1685. <https://doi.org/10.2320/matertrans.M-M2013813>
- Ketcham RA, Carlson WD (2001) Acquisition, optimization and interpretation of X-ray computed tomographic imagery: applications to the geosciences. *Comput Geosci* 27:381–400. [https://doi.org/10.1016/S0098-3004\(00\)00116-3](https://doi.org/10.1016/S0098-3004(00)00116-3)
- Ketcham RA, Slotke DT, Sharp JM (2010) Three-dimensional measurement of fractures in heterogeneous materials using high-resolution X-ray computed tomography. *Geosphere* 6:499–514. <https://doi.org/10.1130/GES00552.1>
- Kim Y, Ha SJ, Yun TS (2020) Deep learning for extracting microfracture: pixel-level detection by convolutional neural network. *E3S Web Conf*. <https://doi.org/10.1051/e3sconf/202020503007>
- Kovács D, Dabi G, Vásárhelyi B (2019) Image processing for fractal geometry-based discrete fracture network modeling input data: a methodological approach. *Cent Eur Geol* 62:1–14. <https://doi.org/10.1556/24.61.2018.09>
- Lanczos C (1950) An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J Res Natl Bur Stand*. <https://doi.org/10.6028/jres.045.026>
- Lee D, Karadimitriou N, Ruf M, Steeb H (2021) Detecting microfractures with X-ray computed tomography. *arXiv preprint arXiv:2103.12821*
- Leu L, Berg S, Enzmann F et al (2014) Fast X-ray micro-tomography of multiphase flow in berea sandstone: a sensitivity study on image processing. *Transp Porous Media* 105:451–469. <https://doi.org/10.1007/s11242-014-0378-4>
- Linden S, Wiegmann A, Hagen H (2015) The LIR space partitioning system applied to the Stokes equations. *Graph Models* 82:58–66. <https://doi.org/10.1016/j.gmod.2015.06.003>
- Lu F, Fu C, Zhang G, Shi J (2020) Adaptive multi-scale feature fusion based residual U-net for fractures segmentation in coal rock images. *Abstract*. <https://doi.org/10.21203/rs.2.23959/v2>
- Lu F, Fu C, Zhang G, Shi J (2021) Adaptive Multi-Scale Feature Fusion based U-net for fracture segmentation in coal rock images. *J Intell Fuzzy Syst*. <https://doi.org/10.3233/JIFS-211968>
- MacQueen J (1967) Some methods for classification and analysis of multivariate observations. *Proc Berkeley Symp Math Stat Probab* 1:281–297
- Moslemipour A, Sadeghnejad S (2021) Dual-scale pore network reconstruction of vugular carbonates using multi-scale imaging techniques. *Adv Water Resour* 147:103795. <https://doi.org/10.1016/j.advwatres.2020.103795>
- Niu Y, Mostaghimi P, Shabaninejad M et al (2020) Digital rock segmentation for petrophysical analysis with reduced user bias using convolutional neural networks. *Water Resour Res*. <https://doi.org/10.1029/2019WR026597>
- Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* 9:62–66. <https://doi.org/10.1109/TSMC.1979.4310076>
- Pot V, Zhong X, Baveye PC (2020) Effect of resolution, reconstruction settings, and segmentation methods on the numerical calculation of saturated soil hydraulic conductivity from 3D computed tomography images. *Geoderma* 362:114089. <https://doi.org/10.1016/j.geoderma.2019.114089>
- Ramandi HL, Mostaghimi P, Armstrong RT (2017) Digital rock analysis for accurate prediction of fractured media permeability. *J Hydrol* 554:817–826. <https://doi.org/10.1016/j.jhydrol.2016.08.029>
- Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF (eds) *Medical image computing and computer-assisted intervention—MICCAI 2015*. Springer International Publishing, Cham, pp 234–241
- Saxena N, Hofmann R, Alpak FO et al (2017) Effect of image segmentation and voxel size on micro-CT computed effective transport and elastic properties. *Mar Pet Geol* 86:972–990. <https://doi.org/10.1016/j.marpetgeo.2017.07.004>
- Saxena N, Hows A, Hofmann R et al (2019) Rock properties from micro-CT images: digital rock transforms for resolution, pore volume, and field of view. *Adv Water Resour* 134:103419. <https://doi.org/10.1016/j.advwatres.2019.103419>
- Scanziani A, Singh K, Bultreys T et al (2018) In situ characterization of immiscible three-phase flow at the pore scale for a water-wet carbonate rock. *Adv Water Resour* 121:446–455. <https://doi.org/10.1016/j.advwatres.2018.09.010>
- Schlüter S, Sheppard A, Brown K, Wildenschild D (2014) Image processing of multiphase images obtained via X-ray microtomography: a review. *Water Resour Res* 50:3615–3639. <https://doi.org/10.1002/2014WR015256>
- Schwarzmeier J (1978) *Geologische Karte von Bayern 1:25 000 Erläuterungen zum Blatt Nr. 6024 Karlstadt und zum Blatt 6124 Remlingen*. Bayerisches Geologisches Landesamt, München
- Shou YD, Zhao Z, Zhou XP (2020) Sensitivity analysis of segmentation techniques and voxel resolution on rock physical properties by X-ray imaging. *J Struct Geol* 133:103978. <https://doi.org/10.1016/j.jsg.2020.103978>
- Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations ICLR 2015, Conference Track Proceedings*, pp 1–13
- Stanchits S, Mayr S, Shapiro S, Dresen G (2011) Fracturing of porous rock induced by fluid injection. *Tectonophysics* 503:94–104. <https://doi.org/10.1016/j.tecto.2010.09.022>
- Stoll M, Huber FM, Trumm M et al (2019) Experimental and numerical investigations on the effect of fracture geometry and fracture aperture distribution on flow and solute transport in natural fractures. *J Contam Hydrol* 221:82–97. <https://doi.org/10.1016/j.jconhyd.2018.11.008>
- Taibi F, Akbarizadeh G, Farshidi E (2019) Robust reservoir rock fracture recognition based on a new sparse feature learning and data training method. Springer, Berlin
- Tsang YW (1992) Usage of “equivalent apertures” for rock fractures as derived from hydraulic and tracer tests. *Water Resour Res* 28:1451–1455. <https://doi.org/10.1029/92WR00361>

- Varfolomeev I, Yakimchuk I, Safonov I (2019) An application of deep neural networks for segmentation of microtomographic images of rock samples. *Computers* 8:1–21. <https://doi.org/10.3390/computers8040072>
- Wang YD, Shabaninejad M, Armstrong RT, Mostaghimi P (2021) Deep neural networks for improving physical accuracy of 2D and 3D multi-mineral segmentation of rock micro-CT images. *Appl Soft Comput* 104:107185. <https://doi.org/10.1016/j.asoc.2021.107185>
- Wiegmann A, Glatt E (2021a) AddiDict handbook, GeoDict 2021. Math2Market GmbH, Kaiserslautern. <https://doi.org/10.30423/userguide.geodict2021-addidict>
- Wiegmann A, Glatt E (2021b) PoroDict handbook, GeoDict 2021. Math2Market GmbH, Kaiserslautern. <https://doi.org/10.30423/userguide.geodict2021-porodict>
- Yang E, Kang DH, Yun TS (2020) Reliable estimation of hydraulic permeability from 3D X-ray CT images of porous rock. *E3S Web Conf* 205:1–5. <https://doi.org/10.1051/e3sconf/202020508004>
- Zhang P, Il LY, Zhang J (2019) A review of high-resolution X-ray computed tomography applied to petroleum geology and a case study. *Micron* 124:102702. <https://doi.org/10.1016/j.micron.2019.102702>
- Zhang W, Wu T, Li Z et al (2021) Fracture recognition in ultrasonic logging images via unsupervised segmentation network. *Earth Sci Informatics*. <https://doi.org/10.1007/s12145-021-00605-6>
- Zimmerman RW, Kumar S, Bodvarsson GS (1991) Lubrication theory analysis of the permeability of rough-walled fractures. *Int J Rock Mech Min Sci Geomech* 28:325–331. [https://doi.org/10.1016/0148-9062\(91\)90597-F](https://doi.org/10.1016/0148-9062(91)90597-F)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.