InterPore JOURNAL

Special Issue – Iran 2024
Original Research

MULTIPHASE IMAGE SEGMENTATION OF NATURALLY FRACTURED MEDIA: BENCHMARKING DEEP LEARNING AND CONVENTIONAL APPROACHES

Behrad Tabrizipour[1], Saeid Sadeghnejad[2], Mastaneh Hajipour[1],
Thorsten Schäfer[2]

[1]Department of Petroleum Engineering, SR.C., Islamic Azad University, Tehran, Iran; [2]Institute for Geosciences, Applied Geology, Friedrich Schiller University Jena, Jena, Germany

Correspondence to:
Saeid Sadeghnejad,
s.sadeghnejad@uni-jena.de

How to Cite:
Tabrizipour, B., Sadeghnejad, S., Hajipour, M., & Schäfer, T. (2026). Multiphase Image Segmentation of Naturally Fractured Media: Benchmarking Deep Learning and Conventional Approaches. *InterPore Journal*, 3(1), IPJ010326-5. https://doi.org/10.69631/j6zv1891

RECEIVED: 21 Aug. 2025
ACCEPTED: 20 Dec. 2025
PUBLISHED: 1 Mar. 2026

ABSTRACT

Characterizing subsurface reservoirs, more specifically naturally fractured subsurface reservoir rocks, is essential for the study of subsurface reservoir properties. Image segmentation is an important aspect of digital rock physics (DRP) workflows. Traditional image segmentation techniques are less accurate than deep learning-based segmentation algorithms. In this paper, we investigate the segmentation accuracy of a convolutional neural network U-net and compare it with traditional methods of Watershed and multi-Otsu thresholding for multiphase segmentation of grayscale images from a naturally fractured coal sample. The segmentation target involved multiphase classification of the matrix, fully-filled fractures with minerals, and open fractures. The results reveal that U-net outperformed the others with an Intersection over Union metric of 94.9%, a Dice metric of 97%, and a Recall metric of 97.5%. The results support the importance of multiphase, deep learning-based segmentation techniques to support DRP studies of naturally fractured rocks.

KEYWORDS

Naturally Fractured Rocks, Digital Rock Physics, Segmentation, U-Net, Watershed, Multi-Otsu

1. INTRODUCTION

Natural fractures control subsurface flow accumulation and retention and affect well productivity (17, 44). From a practical standpoint, it is important to understand how natural fractures affect subsurface formations (39). Evaluating fractured rocks is challenging, and performing segmentation accurately is important for simulating petrophysical properties, such as porosity and permeability, especially at the pore scale. Fractures represent multiple pathways for fluid flow and significantly influence the permeability and deliverability of a reservoir. Open fractures are conduits for fluids, as opposed to closed (mineralized) fractures, which act as a barrier. These interconnected fracture networks increase the complexity and anisotropy of fluid flow through a formation. Therefore, it is essential to know how to separate and identify these fracture types to accurately model and predict rock performance (10, 52). Within the literature, both conventional and artificial intelligence (AI) algorithms have been used extensively in the segmentation of fractured rocks.

Several types of image segmentation techniques have been developed over the years. Multi-thresholding techniques and marker-controlled watershed classification techniques are two traditional examples. These techniques have become common in the segmentation and classification of geologic and reservoir rock images for purposes of clearly defining feature boundaries, separating overlapping components, and enabling more accurate measurements of petrophysical properties (26, 29, 34). Different conventional and machine learning based segmentation techniques in the digital rock physics (DRP) workflow were assessed by Reinhardt et al. to analyze fracture segmentation, with an emphasis on the importance of accurate fracture identification for simulating fluid dynamics (37). Idowu et al. studied segmentation and skeletonization algorithms applied to reservoir rocks, emphasizing segmentation's role in petrophysical predictions (20). Edge detection algorithms for fracture detection in FMI image logs were compared, with the Canny algorithm combined with the Hough transform found to be the most effective (41). While existing methods have shown promise, they often fall short in handling imbalanced data, struggle with complex fracture patterns, and are not well-suited for multiphase segmentation when more than one fracture type (e.g., open and close) coexists in a rock sample. These limitations motivate the need for advanced segmentation strategies capable of addressing imbalanced data and multiphase fracture scenarios.

Fractured rocks contain at least two phases of matrix and fracture in their images. Segmenting filled and open fractures plays a crucial role in accurately analyzing fracture networks. This segmentation ensures that filled fractures are not mistaken for open ones. Moreover, the challenge of imbalanced data, in which minority classes (here, fractures) are underrepresented, was emphasized by several authors (21, 22). Several approaches have been developed to address imbalanced and multiphase segmentation in fractured rock images by utilizing methods such as dynamic class balancing (boosting), implementing weighted loss functions and data augmentation to decrease the model's bias toward the majority classes, and leveraging label-distribution-aware margin loss in order to maximize the capability of the algorithm to predict underrepresented fracture types (8, 19, 57). Further, the Jaccard score, a metric that is especially useful in evaluating the segmentation of imbalanced datasets, was introduced to assess voxel-wise segmentation accuracy when images involve the segmentation of multiple phases (1). Generally, deep learning-based segmentation methods account for class imbalance by utilizing data-preprocessing mechanisms built into current deep learning methods.

Deep learning segmentation has transformed the analysis of rock images by directly learning hierarchical feature representations from raw data, leading to accurate multiphase segmentation of rocks (31, 33, 46, 48). The study by Zhao et al. developed a residual network-convolutional block attention module, which is a deep learning network that improves focus on important features for thin-section image segmentation of rocks (55). However, their study did not address the multiphase segmentation challenge. Ma et al. focused on optimizing the Segment Anything Model for segmenting digital rock computed tomography (CT) and scanning electron microscopy (SEM) images, achieving improvements in speed and accuracy on sandstone images (27). Nevertheless, their work was confined to binary segmentation tasks and a small dataset. Accurate segmentation of fractures is essential for subsequent analysis of fractured rock properties, such as permeability and porosity, which were not addressed in the aforementioned studies. Feng et al. introduced

FFSwinNet, a convolutional neural network (CNN)-Transformer with fast Fourier transform (FFT), for the segmentation of shale fracture from SEM images, outperforming other models (14). Recently, Sadeghnejad et al. combined a DRP workflow with a Random Forest classifier for the segmentation of CT scan images to quantify bentonite mass loss in naturally fractured crystalline rocks (38). Furthermore, various types of deep learning models have achieved success in separating fracture and pore structures from subsurface geological images using U-Net–based models and their associated variants (U-Net++, ResNeXt-U-Net, atrous convolutions-U-Net and SegNet-CNNs) (7, 28, 47, 50). They have been shown to perform very well on synthetic rock images, outcrop photographs, 3D CT scans, and unmanned aerial vehicles imagery, particularly when used in complex or low permeability environments. In addition, instance-segmentation techniques such as Mask R-CNN have proved effective for automated identification of fractures from images taken through drill cores and ultrasonic logging of boreholes with high accuracy, often leveraging transfer learning combined with the use of other deep learning techniques (2, 24). Other specialized deep learning models, as well as hybrid pipelines (e.g., super-resolution technique-based pipelines combined with CNN-based models), such as Res-VGG-UNet (with thresholds for segmentation) and EILnet, have been able to further increase accuracy for identifying and separating multiple fractures within carbonate and coal reservoirs even further (18, 25, 56). Despite these advances, many studies still face challenges in accurately segmenting multiple fracture phases (such as open and closed fractures), handling imbalanced datasets, ensuring generalizability, and achieving fully automated workflows.

To address these challenges, our study introduces significant advancements in the segmentation of both open and filled fractures by leveraging a deep learning–based approach built on the U-Net architecture. This AI-driven semantic segmentation model enhances the precision of multiphase classification with imbalanced data and effectively handles complex and numerous fracture networks, offering a more robust and scalable alternative to conventional techniques. The primary objective of this study is not to conduct an exhaustive benchmarking of multiple deep learning architectures, but rather to demonstrate how a representative convolutional network (U-Net) compares with traditional image segmentation methods (watershed and multi-Otsu) in the context of multiphase segmentation of rocks, when both open and closed fractures coexist.

2. MATERIALS AND METHODS

2.1. Sample Description

Coal reservoirs with natural fractures represent a highly heterogeneous and anisotropic geological system, where the distribution, morphology, and connectivity of the fractures are important in controlling the transport properties and performance of the reservoir. In this investigation, we studied coal core samples that contained both natural and induced fractures. The core samples had complex fracture networks ranging from microfractures to fractures at the centimeter scale. The specific samples chosen represented a geologically relevant heterogeneous system, with differences in matrix composition and fracture mineralogy, all of which are typical features found within fractured coal reservoirs. For example, some sets of raw and manual segmented images are shown in **Figure 1**.

The imaging data in this study were largely derived from the X-ray microtomography (micro-CT) analyses performed by Espinoza et al., which provide non-destructive, high-resolution three-dimensional volumetric imaging at a spatial resolution of approximately 25 microns (12). This is a sufficiently high resolution for pore, fracture, and mineralized vein visualization in the coal matrix. The micro-CT dataset used in this study originates from two analyzed coal cores consisting of one intact and one sheared specimen. Each core was reconstructed as a volume of 2024 × 2024 × 7200 voxels. The raw images are black and white and single-channel in colorspace with a dimension of 1840 x 1840 for height and width. The collection of raw images was extensive in spatial coverage across reservoir samples, which contained areas with isolated fractures as well as areas that were densely fractured (also considered for this study as a fracture "network") with differing degrees of connectivity. The many features present will support modeling for generalized behavior across different morphologies of fractures and reservoir conditions in the future. The dataset includes fractures that are filled with distinct minerals, which also present a useful challenge for segmentation.

2.2. Data Processing, Augmentation and Manual Segmentation

Correct ground truth annotations are essential building blocks for successful supervised deep learning segmentation models, supplying the target labels that these networks are designed to match. In the case of natural fractured subsurface coal reservoirs, accurately delineating fractures and matrix regions is challenging due to differences in grayscale intensity, complete fracture

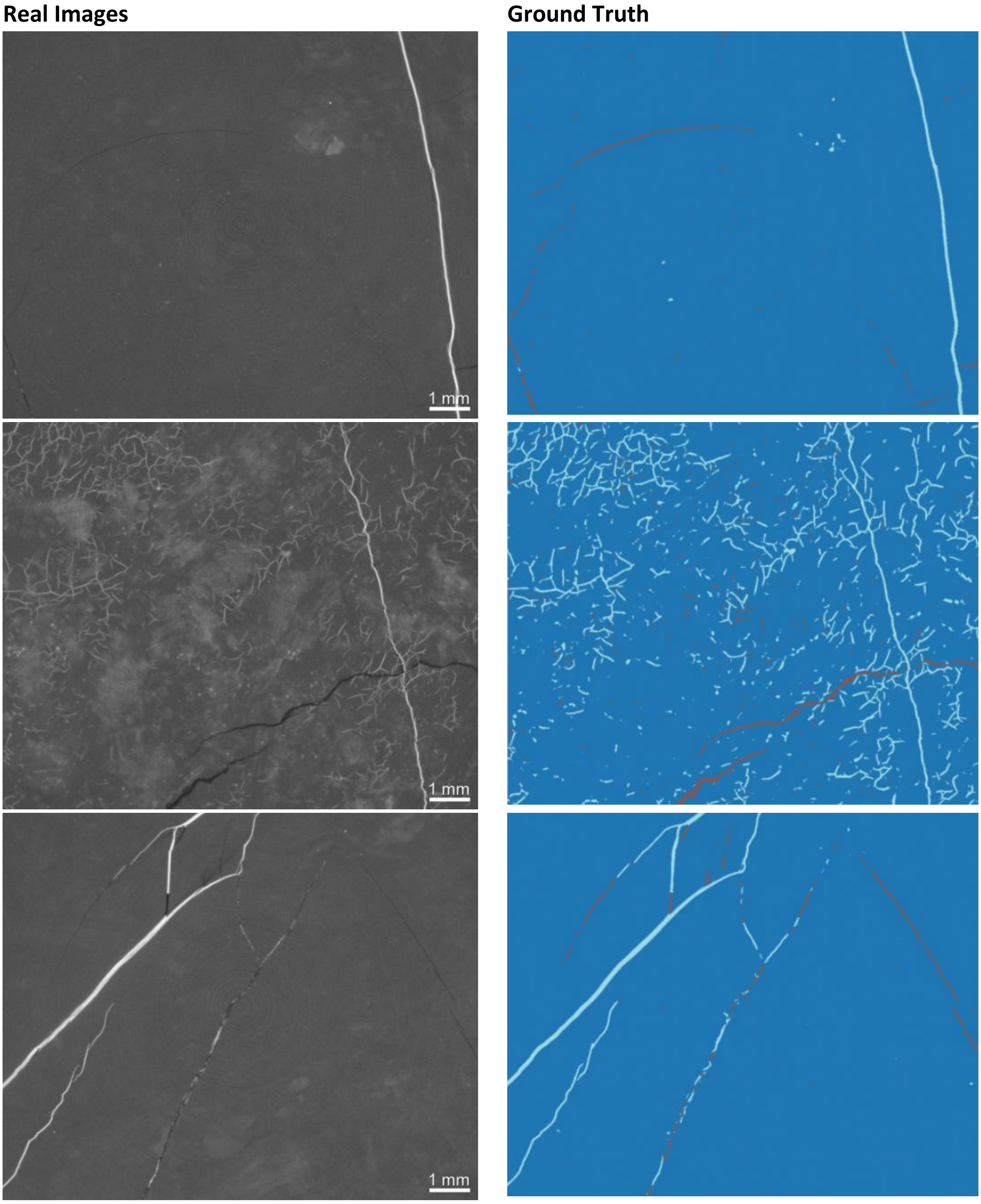

Figure 1: Natural fractured coal samples. The white pixels are filled natural fractures, and the black pixels are open fractures. In the right column, the representative segmented image is highlighted. Matrix, closed fractures, and open fractures are depicted in blue, white and red color, respectively.

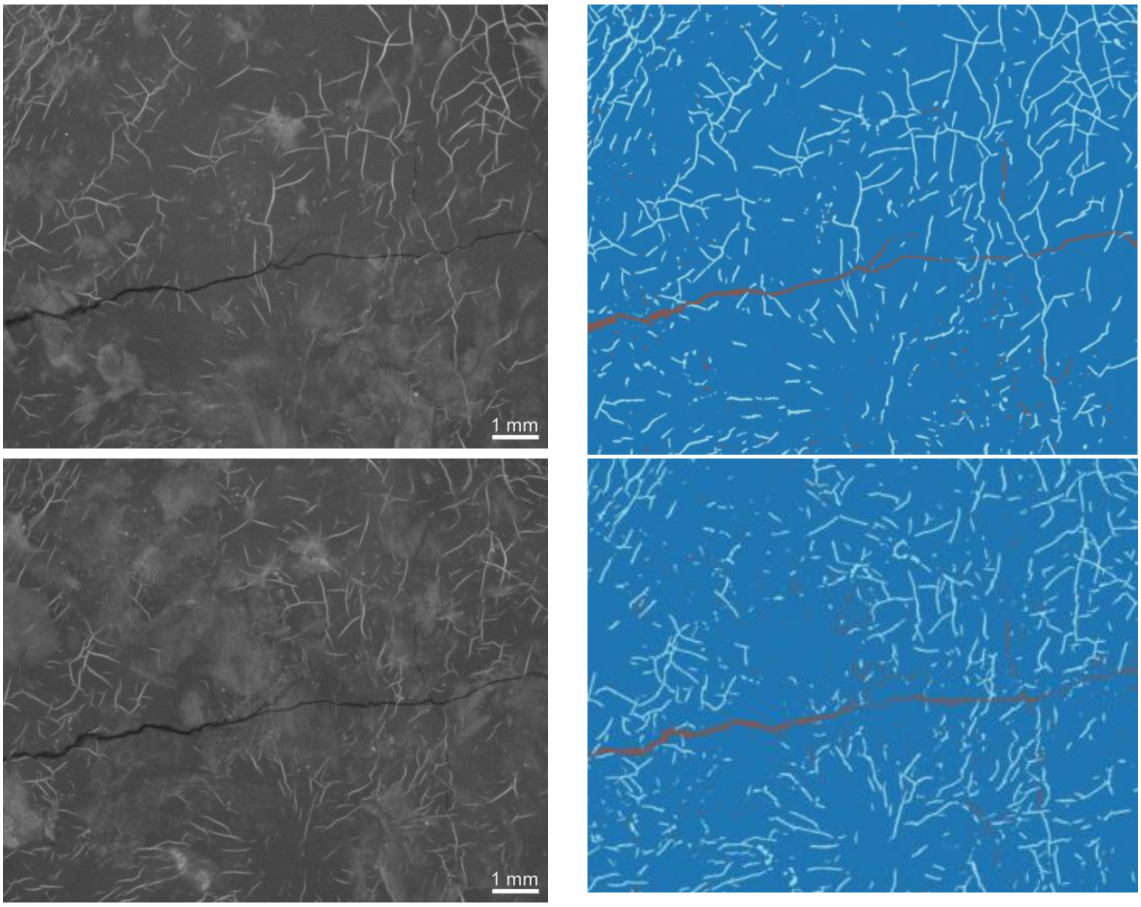

Figure 1 (cont.): Natural fractured coal samples. The white pixels are filled natural fractures, and the black pixels are open fractures. In the right column, the representative segmented image is highlighted. Matrix, closed fractures, and open fractures are depicted in blue, white and red color, respectively.

mineralization, and fracture morphological characteristics. The labeling effort focused on three distinct classes, which are defined in **Table 1**.

Several methods can be used to evaluate labeling consistency, such as measuring inter-rater agreement (e.g., Cohen's κ or Fleiss' κ) between multiple annotators (9, 15). These approaches are widely used to assess annotation reliability in segmentation studies. In our study, the annotation process followed a structured and repeatable protocol. Each image volume was first preprocessed to enhance contrast and visualize fracture boundaries. The annotator (an experienced geoscientist specializing in fracture studies) then manually annotated the classes of matrix, filled fractures, and open fractures in the selected images. All annotations were reviewed and refined under expert supervision to ensure consistency. Since labeling was performed by a single expert, quantitative inter-rater agreement metrics could not be calculated. Nevertheless, we acknowledge that involving multiple annotators or employing soft or probabilistic labels could further improve robustness and help quantify uncertainty.

Table 1: Rock image phases explanations.

Class	**Explanation**
Matrix	The coal rock matrix constituting the majority of the image, typically characterized by relatively uniform grayscale intensities.
Open Fractures	Void spaces or open cracks serving as primary conduits for fluid flow.
Filled Fractures	Fractures partially or fully filled by minerals or other deposits, which can significantly reduce their permeability and fluid dynamics.

The manual ground truth segmentation dataset, thus developed, retains the detailed spatial heterogeneity characteristic of fractured coal rocks, including both macroscopic fractures evident at the centimeter scale and finer microfractures that are only observable with high-resolution imaging techniques. The pixel-wise segmentation maps are representative of the detailed labeling of multiple fracture phases and help address the development of new and advanced multiphase segmentation models. The segmentation masks are encoded as integer tensors for multi-class classification (e.g., 0 = matrix, 1 = filled fractures, 2 = open fractures).

Preprocessing is essential to transform raw image data into a format that deep learning models can accept. It helps improve model convergence, robustness, and generalization, particularly with images of heterogeneous geological materials. Grayscale images from micro-CT scans often contain intensity variances due to illumination or the heterogeneity of the samples themselves. To resolve this, image pixel data were standardized through normalization to zero mean and unit variance, reducing intensity variances. The normalized images and segmentation masks were also converted into a tensor format, which is accepted in PyTorch without affecting associated metadata. The tensor format further allows for efficient loading as batches, data augmentation protocols, and access to GPU acceleration capabilities for computation.

Importantly, baseline augmentation methods, such as random horizontal flip and random vertical flip with a 0.5 probability, were only applied to the data used for training, and hence the resulting changes in orientation, scale and lighting conditions simply improved robustness. This way, CNNs can learn invariant features that improve generalization to heterogeneous and noisy datasets. To lessen spatial autocorrelation (where adjacent patches may have too many similar features), the patches were shuffled randomly. Randomizing the patches decreases the likelihood that the model learns redundant information by learning from samples that are too close together in space.

To avoid spatial leakage among subsets (13), the dataset was comprised of 2,638 selected images and was first divided at the CT slice level into non-overlapping training, validation, and test subsets in an 80/10/10 split ratio. This partition was done before any tiling, shuffling, or augmentation had taken place so that all slices (and adjacent slices from the same volume/sample) were assigned to a single subset. After the splitting process, images within each subset were further split into 128 × 128 patches based on a custom Python script utilizing image processing techniques. The patch size was considered a compromise that was large enough to preserve important contextual information needed to make fracture decisions, but small enough to keep computational times reasonable. This further subdivision expanded the dataset size to more than 11,000 paired patches (raw images and ground truth) as part of the augmentation phase, significantly enhancing the number of training samples and assisting in mitigating overfitting. Because the dataset splitting was conducted before image cropping, the final subsets were composed of 8,800 training, 1,100 validation, and 1,100 test samples using an 80/10/10 split ratio to provide balanced evaluation and consistent comparisons at each stage of the models.

In the training set, patches were randomly shuffled and augmented, while making sure that no patches originating from different or neighboring slices were assigned outside of their designated subsets, and the validation and test datasets were not augmented. All preprocessing steps (e.g., channel matching, normalization) were done per image or from calculations based only on the training data, and therefore from training to validation and test sets, the same preprocessing was applied to avoid data leakage. Notably, we intentionally left the 11,000 selected image pairs imbalanced to preserve the natural characteristics of the dataset and to reflect geological conditions in real-world cases where fractures are naturally sparse and/or imbalanced. Retaining this natural imbalance will allow the model to learn the representative patterns associated with fracture classes, to yield better generalization and consistency with prior studies of class imbalance in deep learning (6).

Lastly, the use of smaller patches allows the model to better concentrate on local variations of texture, intensity, and discontinuities in structure, among other features, of the fracture patterns. In order to resolve the class imbalance caused by the unequal number of instances available for each of the three classes (with minority classes, i.e. fractures, being underrepresented) during optimization, we applied a weighting factor to the Dice loss function based on the inverse of class frequency, which allowed us to apply greater influence during training to the data associated with the minority class (fractures). In addition, we would also point out that the U-Net model, through its unique encoder/decoder design and skip connection properties, provides an ideal learning environment for the training of the minority classes. The model retains detailed spatial information associated with the minority class (fractures) and allows for better identification of small-dimensional fracture regions in areas that are thin or sparse. During training, image patches and labels were batched into increments of 32 to use the GPU appropriately, but the batching was increased to 64 for the validation and test phases. The PyTorch DataLoader optimized computations by loading data asynchronously.

2.3. Automatic Segmentation Methods

2.3.1. Watershed Segmentation

Watershed segmentation is a traditional, region-based method analogous to topographic maps of image intensity (5). Specifically, the grayscale image is thought of as a topographical surface, with bright pixels corresponding to peaks, and dark pixels, valleys. The algorithm conceptually floods the surface from the regional minima to identify different regional basins, using watersheds to define the boundaries. Watershed segmentation works particularly well for separating connected or overlapping objects. Conversely, it is also sensitive to noise and runs the risk of over-segmentation unless aided by preprocessing or marker-controlled watersheds. Therefore, watershed segmentation is helpful for detecting rock fracture boundaries by taking advantage of intensity gradients.

For the Watershed algorithm, several gradient thresholds (0.1–0.3) and minimum basin size thresholds (20–100 pixels) were tried, and the best configuration was a 0.18 gradient and a 50-pixel minimum basin size.

2.3.2. Multi-Otsu Thresholding

Multi-Otsu thresholding is an extension of the original Otsu method, which determines a single threshold by maximizing inter-class variance (32). The Otsu thresholding technique operates directly on image histograms, as it determines thresholds to maximize between-class variances (or equivalently, to minimize within-class variances) across all possible gray-level partitions. Each possible combination of thresholds is exhaustively evaluated, and the combination that provides the greatest statistical separability between the classes is selected.

The multi-threshold extension is simply the multi-class extension of Otsu thresholding; thus, it can segment images into multiple classes. This makes it suitable for geological images with distinct intensity regions, such as pore space, matrix material, and mineral phases. In addition, this method is computationally efficient and does not require any training data. However, it struggles when the intensity distributions overlap, or when the boundaries are poorly defined, which is often the case in fractured rock images. There are several automatic threshold selection techniques for grayscale images based on discriminant analysis that optimize class separability by maximizing between-class variances (40, 45, 51). This method has been modified and used for a variety of images from textures to cells, and is a relatively simple, nonparametric approach for performing effective image segmentation.

It is vital to understand that Otsu's Method and its multi-threshold enhancement have no dependency on user definitions or requirements, as they are completely parameter-free. The entirety of the calculated threshold values has been derived mathematically from the statistics of histogram data alone, with no requirement of user-specified input or preprocessing and training data. This concept is extended in multi-Otsu to include κ classes, where a simultaneous selection of all multiple thresholds is conducted on the basis of maximum variance among all classes. It is on this basis, a completely data-driven and deterministic approach, that we establish the framework upon which we derive accurate and consistent multi-region image segmentation.

2.4. U-Net Segmentation Architecture

Originally, U-Net was developed for biomedical image segmentation (35). U-Net has since been embraced in the geoscience community for geological image analysis in part due to its effectiveness in capturing complicated structures. The U-Net architecture is defined by a symmetric encoder–decoder design. At the encoder, hierarchical features are extracted via convolution layers and pooling layers, while the decoder reconstructs the segmentation map via upsampling layers. Importantly, skip connections integrate the encoder and decoder layers of the same spatial resolution, thus allowing the model to retain fine spatial resolution that may improve the quality of segmentation.

The U-Net is robust to limited training data and is particularly well-suited for the purpose of fracture detection in rock images where it is critical to accurately delineate fracture boundaries. Its fully convolutional nature enables efficient training and inference on fixed-size image patches. A detailed schematic of the U-Net architecture is shown in **Figure 2**.

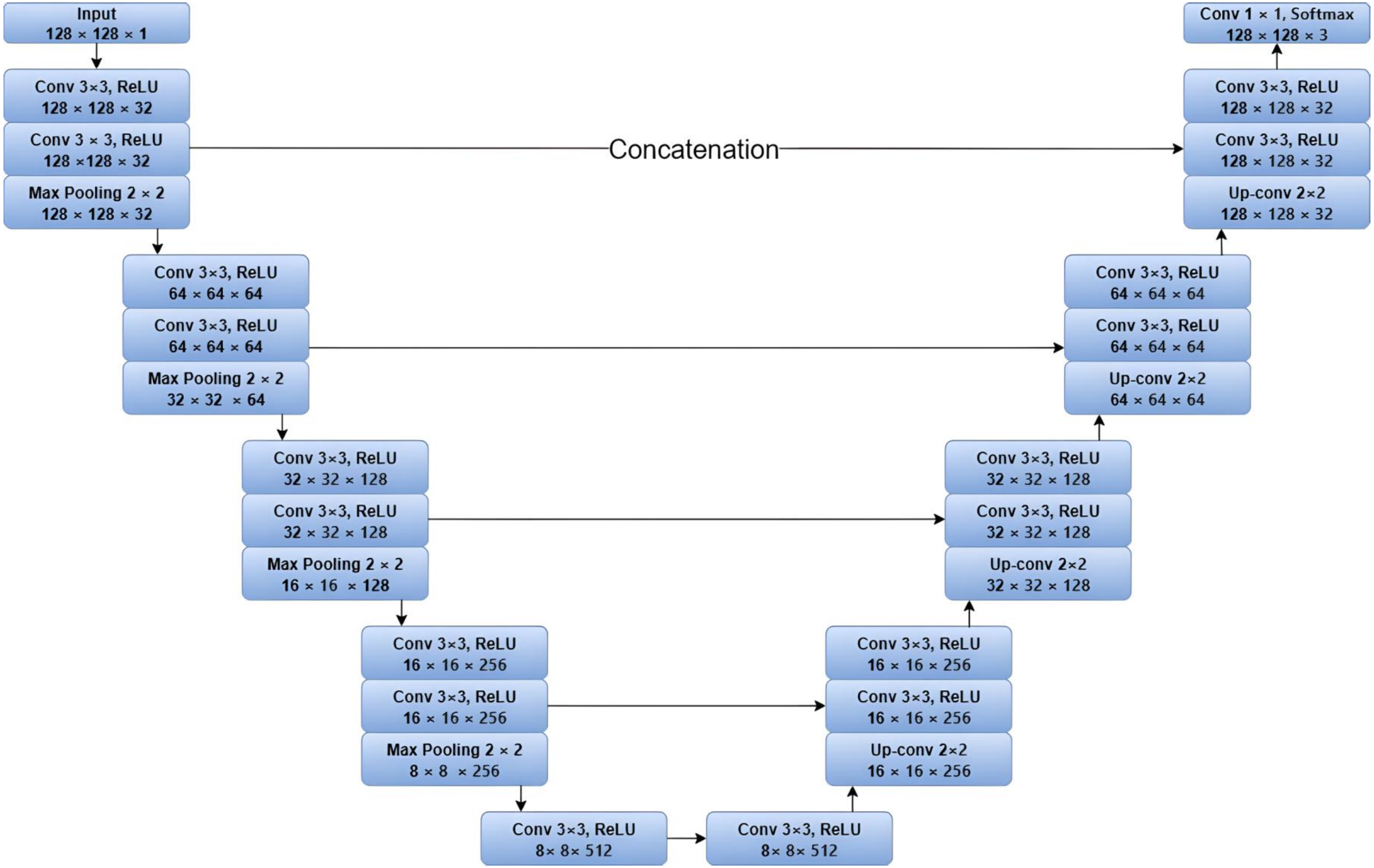

Figure 2: U-net model architecture.

The convolutional blocks, present at the encoder side, extract features at progressively coarser spatial resolutions (54). Each convolutional block usually consists of a pair of small convolutional filters that are followed by non-linear activations. After extraction, a pooling operation reduces the spatial dimensions. While the model goes deeper, the number of feature channels increases, which allows more complex patterns to be represented. Early layers detect simple edges and textures, while deeper layers detect higher-order structures and contextual cues that are critical for accurately segmenting the image.

At the lowest layer of the architecture is the bottleneck, which serves as a link between the encoder and decoder sides (4). At this level, convolutions without any down-sampling occur to merge information into a compact representation. This compact representation encodes global context as well as relationships across the entire input, preparing the model for detailed reconstruction during the subsequent upsampling layers. Each upsampling layer utilizes a learnable up-projection operation to enlarge the spatial dimensions of the feature maps while also reducing their depth (53). The upsampled feature maps are then concatenated with the corresponding output from the encoder through lateral connections to bring back more localized details lost through the down-sampling.

The U-Net architecture was chosen due to its popularity and established validation as an effective baseline model for image segmentation. Additionally, many improved architectures have been developed from U-Net, including but not limited to UNet++, Attention U-Net, and DeepLabv3+ to improve segmentation accuracy (3). The chosen model allows for findings to be comparable, interpretable, and reproducible in the wider literature.

2.4.1. Training Procedure and Hyperparameters Tuning

Training deep CNNs with geological images requires careful optimization of the training strategies, hyperparameters, and usage of hardware so that robust models can be developed without overfitting (11, 30). The hyperparameters of the U-Net model were obtained by grid search in a systematic manner to achieve an optimal configuration for accurate and robust fracture segmentation. In particular, learning rates were tested in the range 10^{-1} to ۱۰$^{-7}$, and weight decay values $|\lambda|$ between 0 to ۱۰$^{-5}$. Several optimization methods were benchmarked (Adam, AdamW, RMSProp, and SGD) in a fair setting regarding convergence trajectory and generalization performance.

Training lasted for up to 20 epochs. This time was sufficient for convergence, as shown by the stabilization of training and validation losses, along with the segmentation metrics leveling off. To avoid overfitting, we used an early stopping method, which stopped training if the validation loss did not improve for five straight epochs.

To protect against training problems or overfitting, we saved model checkpoints at the epochs with the best validation performance. This way, we could roll back if needed. We kept detailed training logs, which included loss curves, accuracy measurements, Intersection Over Union (IoU) and Dice scores, to track progress and adjust hyperparameters.

2.4.2. Model Evaluation

This study utilized a set of complementary metrics to offer a comprehensive evaluation of model performance. Essential metrics, like IoU and Recall, are vital for imbalanced datasets, such as the one examined in this study. Precision calculates the proportion of pixels correctly predicted as positive out of all pixels predicted as positive. High precision reflects the model's ability to minimize false positives, which is critical in fracture detection to avoid over-segmentation. It is calculated with **Equation 1** (43):

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

where TP (True Positive) refers to how many of the model's predictions correctly align with the ground truth as positive. FP (False Positives) is when the model incorrectly predicts a positive result while the true label is negative.

Recall determines the fraction of positive pixels that were correctly classified by the model. High recall provides a higher likelihood that most of the true fracture sections were captured, ensuring an overall lower chance of missing fracture sections. This metric is defined in **Equation 2**:

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

where FN (False Negatives) refers to occasions in which the model incorrectly predicts a negative result, when the true label is actually positive.

The $\mathrm{F1-score}$ is the harmonic mean of precision and recall and expresses a trade-off between false positives and false negatives. The $\mathrm{F1-score}$ is especially useful in class-imbalanced situations because it evaluates performance with equal weight on precision and recall. The $\mathrm{F1-score}$ is defined by the below formula (**Eq. 3**; 16):

$$\mathrm{F1-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3}$$

Using the Jaccard score (**Eq. 4**; 49) as an encompassing metric yields a balanced and sophisticated perspective on the segmentation model's strengths and weaknesses. More granular assessments provide better context for improvement and fairness in comparisons when different methodological approaches are evaluated.

$$\text{Jaccard Score} = \frac{|A \cap B|}{|A \cup B|} \tag{4}$$

In **Equation 4**, A is the set of predicted values and B is the set of true values, $A \cup B$ refers to everything that exists in both A and , while $A \cap B$ refers to the overlaps between the two classes. All metrics were calculated for both classes, and the scores were averaged for both class-specific and overall performance scores. The calculations were performed using standard, well-validated libraries by calculating pixel-wise comparisons between the predicted segmentation masks and the ground truth data within the test dataset.

Intersection over Union (IoU) (36) quantifies the degree to which there is overlap between the predicted segmentation and the ground truth mask. The IoU is defined as the ratio of the area of intersection to the area of the union between the predicted and true regions. IoU provides a simple and sufficient measure of segmentation accuracy (see **Eq. 5**),

$$\mathrm{IoU} = \frac{TP}{TP + FP + FN} = \frac{|A \cap B|}{A + B - |A \cap B|} = \frac{Intersection\ of\ Predicted\ and\ ground\ truth}{Union\ of\ predicted\ and\ ground\ truth} \tag{5}$$

where A is the set predicted values, and B is the set of True values**.**

The Dice metric (42) is a measure used to compare how similar two sets are and is often used in image segmentation to see how closely a model's predictions match the actual ground truth (**Eq. 6**).

$$\mathrm{Dice} = \frac{2 \times |A \cap B|}{|A + B|} \tag{6}$$

The outcome ranges between 0 and 1, implying that 1 indicates a perfect match and 0 implies no relationship. The Dice coefficient is especially helpful when it comes to multiphase segmentation tasks, as it allows one to ensure that the predicted areas and true areas match as closely as possible. Because the loss function provides well-calibrated gradients upon backpropagation, it is particularly useful for multi-label segmentation scenarios. The F1 and Dice metrics are highly correlated in binary segmentation tasks, as has also been demonstrated in existing studies in the medical image segmentation literature (23). Finally, the DiceLoss formula is defined using the Dice metric as follows (**Eq. 7**):

$$\text{DiceLoss} = 1 - Dice \tag{7}$$

In the context of evaluating image segmentation tasks, IoU and the Jaccard Index are indicators of the spatial overlap between predicted regions and ground truth, and both are mathematically identical representations of set similarity. However, IoU and the Jaccard Index can be computed using one of two aggregation strategies (i.e., global (micro-averaged) and per-class (macro-averaged)). The global IoU computes the intersection and union for all classes together, resulting in an overall score that reflects performance across the evaluated image dataset. This method is particularly appropriate when class distributions are significantly imbalanced. The per-class Jaccard Index computes IoU as a separate score for each class and averages the per-class Jaccard indices together (either uniformly or using class-specific weights). This method yields a more interpretable measure of multiclass segmentation performance, while ensuring equal contribution of rare classes to the overall score. Both aggregation strategies were considered in this study to achieve both consistency and a comparison of differences.

2.5. Workflow Overview

A summary of the workflow is provided in **Figure 3**. The study begins with collecting X-ray images of naturally fractured coal rock reservoirs, each capturing three phases of matrix, filled fractures, and open fractures. Next, every image is segmented manually into the three target categories to establish ground-truth masks. After separating the CT image into an 80/10/10 portion for training, validation, and testing, the 1,840 × 1,840-pixel images are cropped into smaller 128 × 128 pixels. Those raw and mask images are converted into PyTorch tensors, normalized by their mean and standard deviation, and after that the mask labels are zero-centered. To address class imbalance, PyTorch's augmentation tools are applied, and the training loss function is reweighted accordingly. Segmentation is performed using the Watershed method, multi-Otsu thresholding, and a U-Net network, and each result is compared against the manual labels. Performance is assessed via classification metrics, like confusion matrices and pixel-wise metrics. Finally, the porosity of the open-fracture phase is computed on the original, mask, and segmented images, demonstrating that the U-Net approach yields the most accurate porosity estimates.

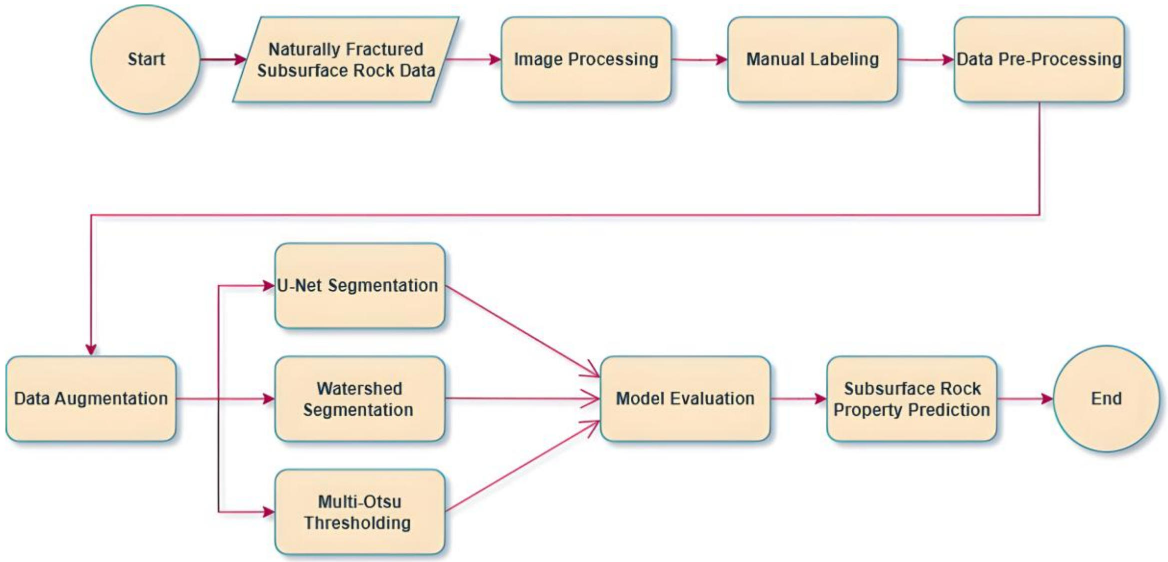

Figure 3: Workflow flowchart for this study.

3. RESULTS AND DISCUSSION

3.1. Deep Learning Segmentation & Tuned Hyperparameters

Table 2 shows the tuned hyperparameters of the U-Net model. Extensive analyses were performed to find the optimal hyperparameters after evaluating the robustness of the model by retraining it multiple times with different seeds. In addition, to verify the stability of the optimized configuration, the previous step was conducted to ensure that the final results were not dependent on a specific random initialization. Other training parameters, including the loss function, learning rate, optimizer, and weight decay, were selected according to the best-performing configuration identified during the training phase repeated over multiple runs with different random initializations. The mean ± standard deviation of the evaluation metrics (IoU, F1-score, Precision, Recall, Jaccard Index, and Dice Coefficient) are presented in **Table 3** across the related runs. A minor variation (below 0.25%) among the related results indicates that the model is stable and robust against the effects of the stochastic initialization.

Table 2: Tuned parameters of the U-net model.

Parameters	**Tuned amount**
Learning rate	0.002
Loss Function	Dice loss
Optimizer	Adam
Weight decay	1e-4

Table 3: U-Net model metrics based on different seeds, with mean ± standard deviations.

Metrics /Seed	**IoU (%)**	**F1-score (%)**	**Precision (%)**	**Recall (%)**	**Jaccard Score (%)**	**Dice Metric (%)**
1st	94.7	96.9	95.5	97.1	95.5	96.8
2nd	94.8	97.0	95.7	97.3	95.6	96.9
3rd	94.6	96.8	95.4	97.0	95.4	96.7
4th	94.9	97.2	95.9	97.6	95.8	97.1
5th	95.0	97.3	96.0	97.5	95.9	97.2
Mean ± STD	97.04 ± 0.19	95.70 ± 0.23	97.30 ± 0.23	95.64 ± 0.18	96.94 ± 0.19	94.80 ± 0.14

Figure 4 shows the accuracies and losses of both training and test data and demonstrates convergence within 20 epochs. The use of both batch normalization layers and an early stopping mechanism contributed to smooth and reliable training behavior. The rapid convergence, alongside the superior accuracy of the U-net architecture, highlights its strengths, making it well-suited for assessing complex multiphase segmentation challenges as demonstrated by heterogeneous reservoir images.

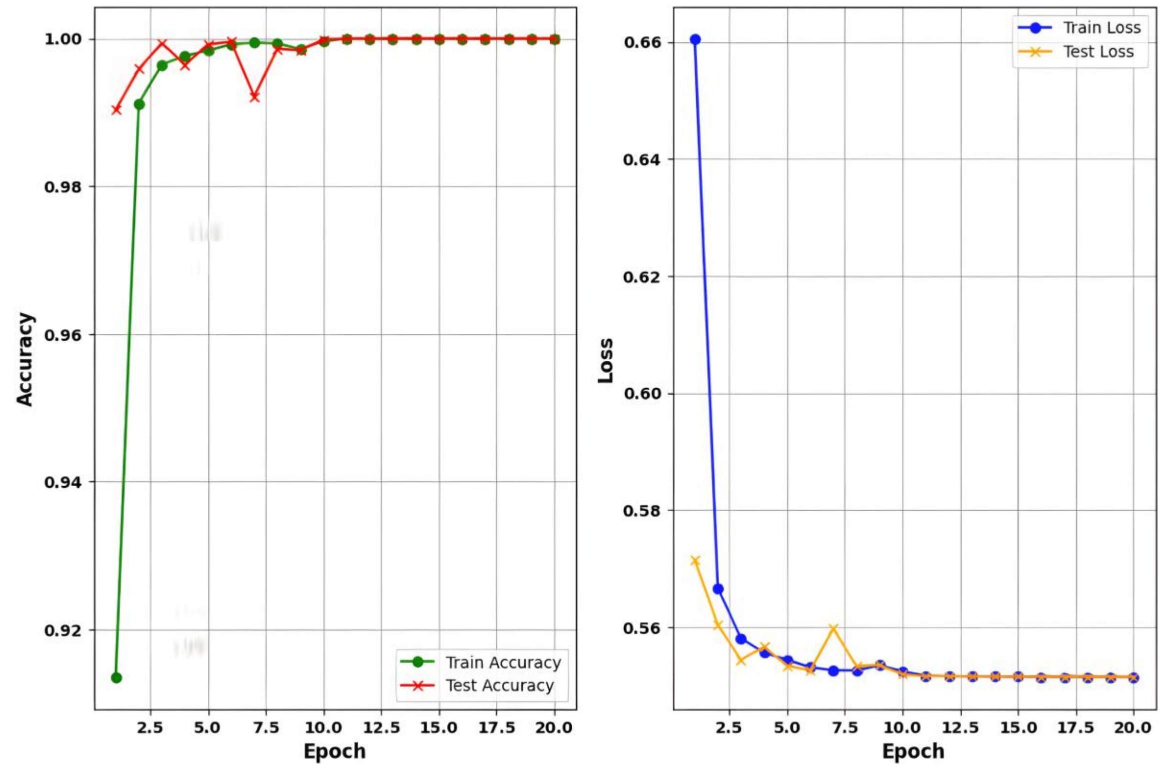

Figure 4: Train and test accuracy and loss of the U-net architecture per epoch.

3.2. Comparison of Segmentation Techniques

Figure 5 compares the segmentation results of all three techniques with the ground truth image, which represents a partial section of the core image. As can be seen, fractures with smaller apertures are not well segmented by the Watershed and multi-Otsu thresholding methods, resulting in partially connected or unsegmented fractures (the white-marked regions in **Figures 5d** and **5e**) compared to the raw image and ground truth. In certain instances, open fracture pixels are frequently misclassified as matrix, as exemplified by the yellow regions in **Figures 5d** and **5e**. Additionally, some segments identified by the Watershed method are incorrectly classified as open fractures instead of filled fractures, as indicated by the blue-circled areas in **Figure 5d**. In contrast, a class-wise analysis revealed that the U-Net model achieved the highest segmentation accuracy for filled fracture regions, particularly in the blue circled areas of **Figure 5c**, compared to the Watershed method. The U-Net model (**Fig. 5c**) demonstrated superior performance in distinguishing filled fractures, open fractures, and matrix classes. Its advanced localization capabilities enabled precise edge detection of narrow open fractures, which is critical for petrophysical modeling. The blue-, yellow-, and white-marked regions in **Figure 5c** highlight U-Net's superior results, particularly in the white-marked areas, which conventional methods failed to capture accurately. Visual inspection confirmed that U-Net produced clearer and more continuous fracture masks (compare **Fig. 5b** to **5e**), whereas other methods introduced artifacts that compromised segmentation reliability. These differences in segmentation fidelity across models are evident in **Figure 5**, underscoring U-Net's high-quality segmentation, especially when numerous fractures are present.

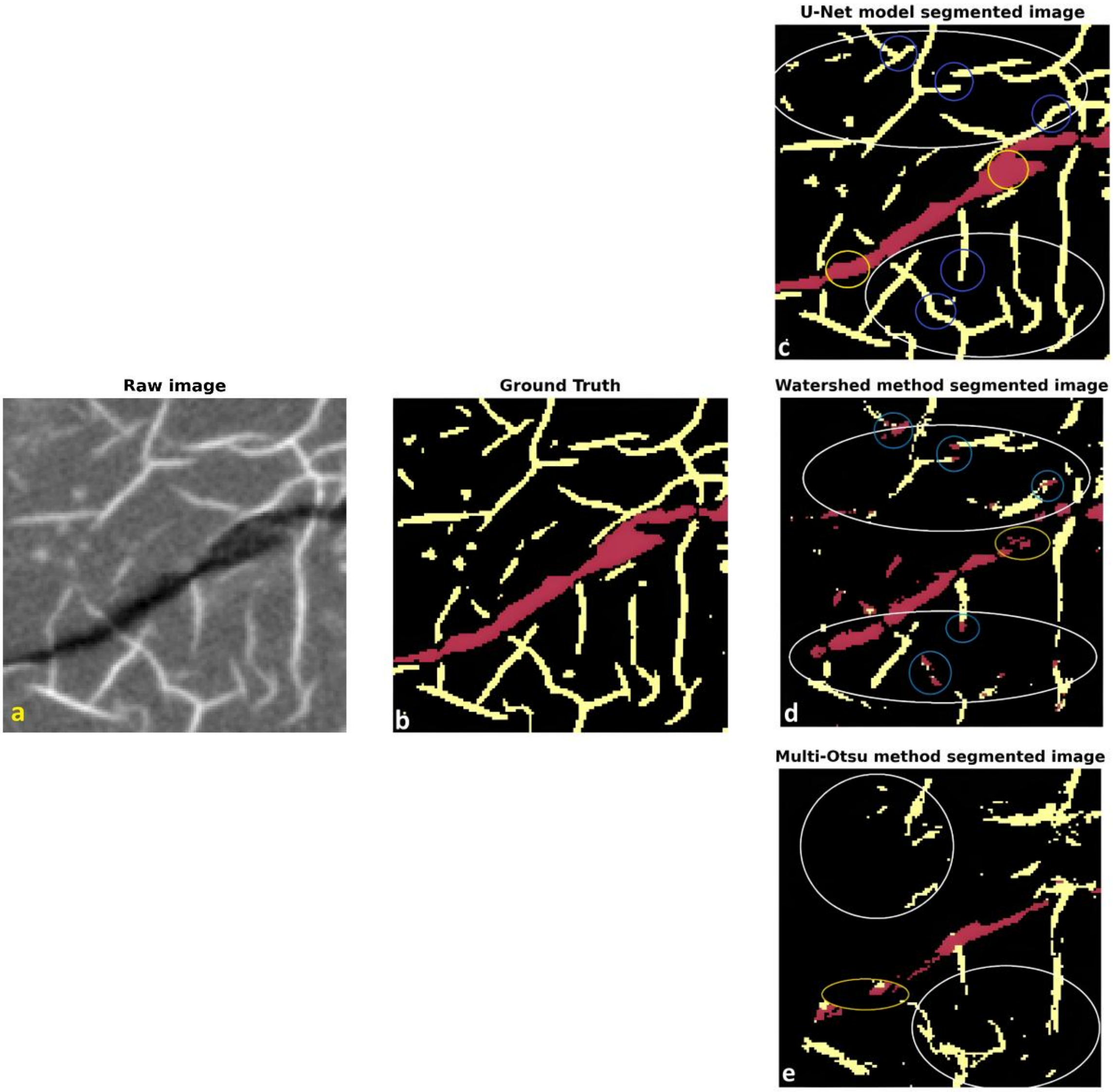

Figure 5: The differences between segmentation methods. **(a)** Grayscale raw image. **(b)** Ground truth (manually segmented). Segmentation results with **(c)** the U-Net model, **(d)** the Watershed method, and **(e)** the multi-Otsu thresholding method.

Quantitative findings emphasize the superiority of the U-Net, which outperforms the conventional methods across all metric scores. In particular, the U-Net model had an IoU of 94.9% and a recall of 97.5%, much higher than the Watershed method, which had an IoU of 66.4% and a recall of 71%, and the multi-thresholding method, with an IoU of 68.3% and a recall rate of 67.7% (see Table 5). Such a good performance is likely owed to the architecture of the U-Net, namely its skip connections, which allow high-resolution features to be maintained at the decoding stage, and hence these details can be identified more effectively in the spatial domain of fractures compared to classic segmentation methods.

Table 5: Differences between metrics of the implemented segmentation methods.

Model	IoU (%)	F1-score (%)	Precision (%)	Recall (%)	Jaccard Score (%)	Dice Metric (%)
U-Net	94.9	97.1	95.8	97.5	95.7	97
Watershed	66.4	75.6	80.1	71.0	72.4	59
Multi-Otsu	68.3	73.0	69.4	67.7	68.5	51

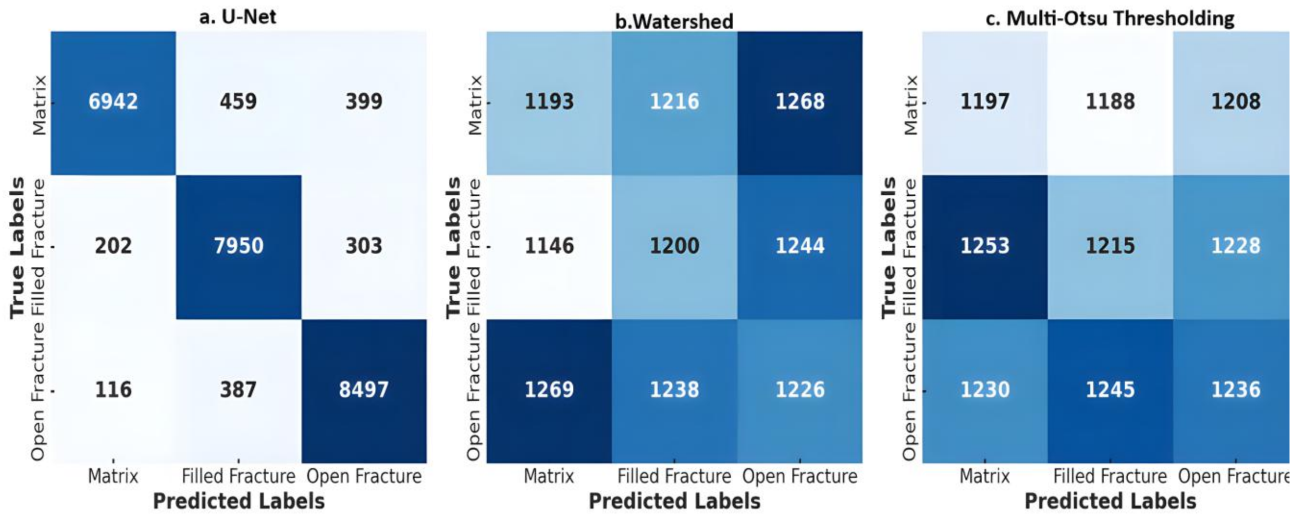

Figure 6: Confusion matrices for **(a)** U-net, **(b)** the Watershed method, and **(c)** the Multi-Otsu thresholding method.

The classification performance of the three segmentation approaches (i.e., U-Net, Watershed, and multi-Otsu) was evaluated using the confusion matrix method (Fig. 6). The results indicate that U-Net achieved the highest classification accuracy, with most samples correctly assigned to their respective fracture classes and only a small proportion misclassified. Watershed segmentation showed lower performance, with a higher number of open fractures being incorrectly labeled as matrix (11.5%, 1269 cases) and some filled fractures misidentified as open fractures (11.3 %, 1244 cases). Moreover, the multi-Otsu method also exhibited poor classification performance, particularly in distinguishing filled fractures from open fractures (11 %, 1128 cases). Overall, the confusion matrix analysis highlights that the deep learning–based U-Net is more robust in handling the subtle textural differences between open and filled fractures compared to conventional threshold- and morphology-based methods.

3.3. Porosity Measurement Assessment

Figure 7 illustrates the computed porosity values, showing a comparison of ground-truth open fracture porosity across the segmentation methods examined in this study. It emphasizes how accurately capturing phase porosity is crucial for understanding the material's structure and integrity, making these measurements valuable for further research and practical applications. The results demonstrate a strong match between the open fracture porosity values from U-Net–

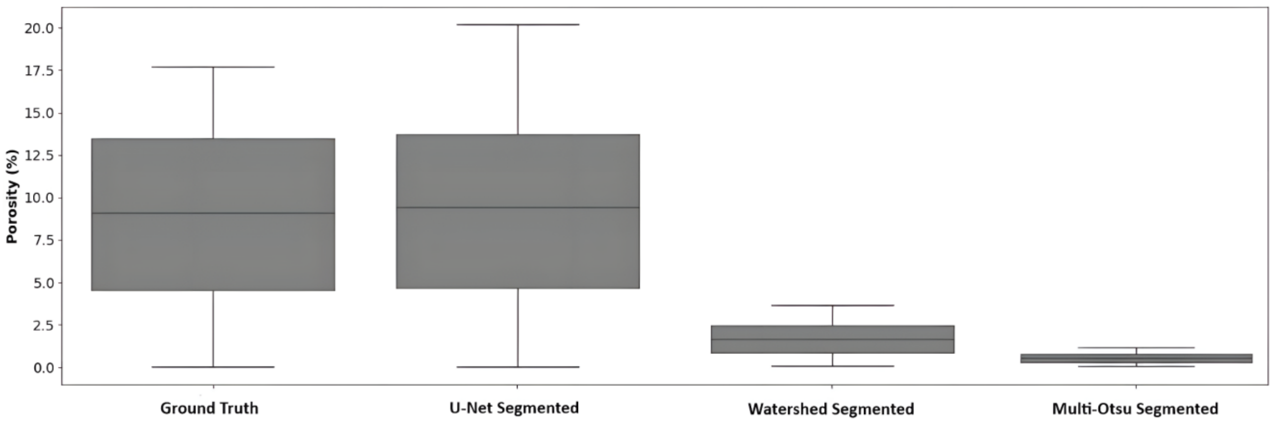

Figure 7: Comparison of ground-truth open fracture porosity with porosities derived from U-Net, Watershed, and Multi-Otsu thresholding segmentation methods.

segmented and ground-truth images. This close agreement emphasizes that U-Net-based segmentation is a reliable method for identifying porous regions, highlighting that this method enables more accurate property analysis and can be applied to a wide range of real-world scientific and industrial applications.

The method for quantifying open fracture porosity from CT images involved using two-dimensional pixel areas as a percentage of a given slice in order to calculate the amount of open fracture volume per core length (or per cross-section) based on the void pixel areas within each CT scan. In this process, only pixels belonging to the open fracture class were considered in the calculation of porosity. As mentioned previously, both types of fractures (open and filled) contribute to the total volume of open fracture space for each slice.

The image feature size used to measure the porosity of each slice has a resolution of approximately 25 µm, which is used solely as a measure of the size of pixels (in real-world units). The porosity is calculated solely on the open fractures that appear in the cross-section images. Therefore, the image porosity can simply be calculated by counting the open fracture pixels in each image and then dividing that count by the total number of pixels in the respective cross-section image.

The mean porosity (%) indicated by U-Net, as shown in **Table 6**, was very close to the actual values (mean ± SD of 9.5 ± 5.93%) and had the smallest absolute deviations (mean absolute ± SD of only 0.5 ± 0.30%). Conversely, Watershed and multi-Otsu methodologies had lower mean porosity values (6.95 and 8.90%, respectively) compared to those of U-Net's, as well as higher average error rates relative to U-Net's performance. Therefore, U-NET was the most reliable and consistent way to estimate porosity among the methods compared in this study.

Table 6: Quantitative porosity comparison: mean, standard deviation, and mean absolute error (MAE) of the absolute porosity difference relative to ground truth.

Method	Mean	Std	MAE vs Ground Truth
Ground Truth	9.0	6	–
U-Net	9.5	5.93	0.5
Watershed	2.0	1.48	6.95
Multi-Otsu	0.6	0.30	8.9

3.4. U-Net Robustness Evaluation

The publicly available multiphase CT datasets of fractured reservoir rocks are extremely limited, which constrains the possibility of testing on completely independent samples. Nevertheless, to evaluate the robustness and generalization potential of our model, we applied Gaussian blurring with different filter values to test dataset images to evaluate the effectiveness of the trained model. A Gaussian filter with low (σ ≈ 0.3–0.8), intermediate (σ ≈ 0.8–1.5), and high (σ ≈ 1.5–2.5) blurring levels was applied to each column in **Figure 8**. The second row presents the corresponding ground truth masks, and the third row illustrates the U-Net's segmentation outputs. When evaluated on low- and medium-blurred test images, the U-Net maintained accurate prediction of fractures and matrix regions, closely resembling the ground truth. In contrast, under the high-blurring effect, segmentation quality declined substantially, with blurred boundaries leading to matrix areas being misclassified as fractures (highlighted). These results indicate that while the U-Net remains robust to moderate image degradation, its accuracy declines notably when strong blur is introduced to the image.

As summarized in **Table 7**, when the blur filter amount was low (σ ≈ 0.3–0.8) or medium (σ ≈ 0.8–1.5), all metrics remained close to the clean-image baseline, with IoU values above 93% and Dice coefficients around 96%. This indicates that the network's encoder–decoder architecture and skip connections preserved fine details and boundary continuity even when subtle texture information

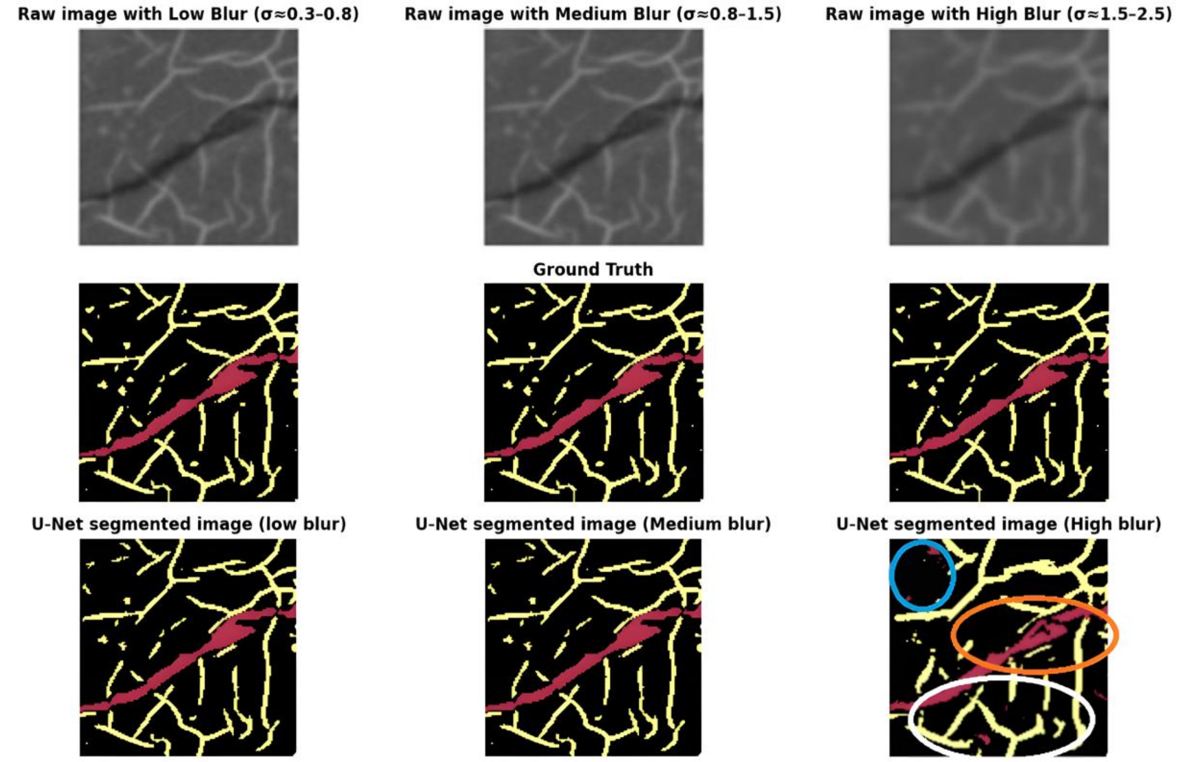

Figure 8: Segmentation results of the U-Net model tested on blurred images.

was lost. However, under the high blurring condition (σ ≈ 1.5–2.5), segmentation accuracy deteriorated sharply. The IoU dropped to 78.4%, and the Dice metric declined to 84.5%. Visual inspection of these results confirmed that the model frequently misclassified low-contrast matrix regions as fractures, particularly where pore edges and mineral boundaries became diffuse. Precision decreased substantially, suggesting over-segmentation of the fracture regions, while recall remained relatively high, indicating the network's bias toward labeling ambiguous blurred zones as fractures rather than matrix.

Table 7: Robustness of the U-Net under different Gaussian blur.

Model	Blur intensity	IoU (%)	F1-score (%)	Precision (%)	Recall (%)	Jaccard Score (%)	Dice Metric (%)
U-Net (low blur)	σ≈0.3–0.8	94.6	96.9	95.5	97.3	95.4	96.8
U-Net (medium blur)	σ≈0.8–1.5	93.8	96.3	94.9	96.9	94.6	96.2
U-Net (high blur)	σ≈1.5–2.5	78.4	84.6	81.3	87.9	80.1	84.5

Although the U-Net model performs reliably across the test set, several representative failure cases were identified, as shown in **Figure 9**. These include missed detection of very thin fractures whose widths approach the voxel resolution, reduced performance in low-contrast regions where the signal-to-noise ratio is insufficient to preserve fracture–matrix boundaries, and some test images, although it is important to note that these cases are exceptions. To better understand model robustness, we also introduced controlled noise levels into a subset of images, as explained in the previous section. The model maintained accurate segmentation under moderate and even high noise, indicating strong resilience to typical imaging variability. Performance degradation was observed only when noise intensity became large enough to obscure structural gradients, a well-known limitation in convolution-based segmentation. The failures presented here therefore arise from challenging image conditions, such as feature sizes near the resolution limit and regions with extremely weak contrast, combined with the inherent constraints of the standard U-Net architecture in capturing ultra-fine or noise-dominated boundaries.

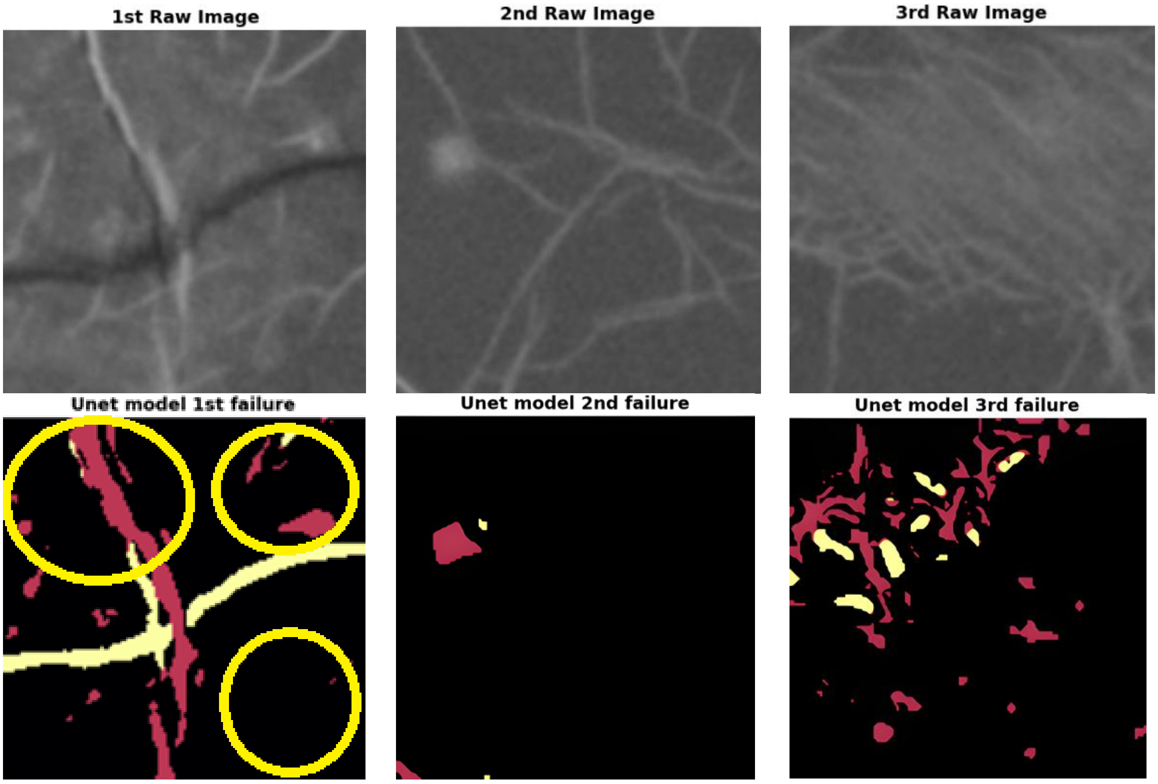

Figure 9: Representative qualitative failure cases from the U-Net segmentation model.

3.5. Limitations and Future Work

There are certain limitations associated with using the proposed U-Net model despite the strong overall performance in fracture detection and segmentation. The model occasionally struggled in low-contrast regions where the distinction between filled and open fractures was difficult. Detection of very thin fractures close to the image resolution limit also posed challenges, particularly in regions with dense fracture networks and near boundaries where fractures merge with the surrounding matrix. There is an additional limitation of this study related to the model's ability to accommodate varying patch sizes. Small patches such as 64 × 64 do not effectively represent the relatedness between fractures and tend to produce broken or incomplete fracture predictions. In contrast, larger patches, like 256 × 256, include more contextual information but come with several associated issues: **1)** They require much greater amounts of memory to store the models; **2)** Longer training times are required; **3)** High intra-patch class imbalance may decrease the stability of optimization. Moreover, there are some false positive detections associated with small noise-induced artifacts. These limitations can be related to the limited availability of data diversity, class imbalance between fracture types and matrix, densely crowded fractures, and low-contrast imaging processes. For future work, some improvements are recommended, such as including more data with different rock types and diverse examples, applying additional data augmentation or preprocessing steps, grouping attention mechanisms or multi-scale feature extraction modules, and computing or simulating images and fractures to enable calculation of additional fracture properties.

4. CONCLUSIONS

In this study, the applications of segmentation methods, including multi-Otsu thresholding, the Watershed method, and a deep CNN, specifically the U-Net architecture, are discussed and demonstrate improved performance of U-Net compared to the other methods for multiphase segmentation of naturally fractured subsurface coal reservoir images.

Based on the three methods tested (U-Net, multi-Otsu thresholding, and the Watershed method), the U-Net model consistently delivered the best overall results. It achieved high scores across the board, with an IoU of 94.9%, an F1-score of 97.1%, precision of 95.8%, recall of 97.5%, a Jaccard score of 95.7, and a Dice metric of 97%. These numbers indicate that U-Net was accurate and reliable in identifying fracture features. In comparison, Watershed and multi-Otsu did not perform as well: Watershed had an IoU of 66.4% and an F1-score of 75.6, while multi-Otsu had an IoU of 68.3% and an F1-score of 73.0%. U-Net's encoder-decoder design, augmented with skip connections, enabled better preservation of spatial details, resulting in more accurate fracture segmentation. Additionally, Gaussian blurs with different intensities were applied to the original images, and the U-Net model was tested on these to evaluate model generalizability. The results show that U-Net preserved strong performance under low and medium blur conditions, achieving IoU values of 94.6% and 93.8%, with corresponding recall scores of 97.3% and 96.9%, respectively. Under high blur conditions, performance decreased, with IoU and recall scores of 78.4% and 87.9%, respectively.

However, there are several challenges in segmenting extremely fine or low-contrast fractures, and this opens avenues for future research. Potential directions include the integration of multimodal imaging and the adoption of more complex network architectures.

Overall, this work shows the potential of deep learning methods as robust, automated tools for DRP workflows to advance reservoir characterization, offering promising avenues for more efficient and precise digital reservoir modeling, as well as improved strategies for enhancing hydrogen storage.

STATEMENTS AND DECLARATIONS

Author Contributions

B. Tabrizipour: Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft. **S. Sadeghnejad:** Conceptualization, Methodology, Project Administration, Resources, Supervision, Validation, Writing – Review & Editing. **M. Hajipour:** Supervision, Writing – Review & Editing; **T. Schäfer**: Resources, Writing – Review & Editing

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Data, Code & Protocol Availability

The datasets generated and the source code are available in the following repository: https://github.com/behradtp/Image-Segmentation_Interpore

Funding Received

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ORCID IDs

Behrad Tabrizipour	https://orcid.org/0009-0002-1747-0909
Saeid Sadeghnejad	https://orcid.org/0000-0001-7386-4158
Mastaneh Hajipour	https://orcid.org/0000-0002-3122-4934
Thorsten Schäfer	https://orcid.org/0000-0002-7133-8717

REFERENCES

1. Alqahtani, N. J., Niu, Y., Wang, Y. D., Chung, T., Lanetc, Z., Zhuravljov, A., Armstrong, R. T., & Mostaghimi, P. (2022). Super-resolved segmentation of X-ray images of carbonate rocks using deep learning. *Transport in Porous Media*, *143*(2), 497–525. https://doi.org/10.1007/s11242-022-01781-9
2. Alzubaidi, F., Makuluni, P., Clark, S. R., Lie, J. E., Mostaghimi, P. et al. (2021). Automated fracture detection and characterization from unwrapped drill-core images using Mask R-CNN. *3rd AEGC: Geosciences for Sustainable World* (pp. 1–4) [Online]. Retrieved September 13–17, https://doi.org/10.1053/j.petrol.2021.109471
3. Angelina, C. L., & Rospawan, A. (2025). A benchmark study of DeepLabV3+, U-Net++, and Attention U-Net for blood cell segmentation. *Green Intelligent Systems and Applications*, *5*(1), 61–73. https://doi.org/10.53623/gisa.v5i1.607
4. Azad, R., Aghdam, E. K., Rauland, A., Jia, Y., Avval, A. H., Bozorgpour, A., Karimijafarbigloo, S., Cohen, J. P., Adeli, E., & Merhof, D. (2024) [Advance online publication]. Medical image segmentation review: The success of U-Net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *46*(12), 10076–10095. https://doi.org/10.1109/TPAMI.2024.3435571
5. Beucher, S., & Lantuéjoul, C. (1979). Use of watersheds in contour detection. Paper presented at the International Workshop on Image Processing, Real-Time Edge and Motion Detection/Estimation. Rennes, France.
6. Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011
7. Byun, H., Kim, J., Yoon, D., Kang, I.-S., & Song, J.-J. (2021). A deep convolutional neural network for rock fracture image segmentation. *Earth Science Informatics*, *14*(4), 1937–1951. https://doi.org/10.1007/s12145-021-00650-1
8. Cao, K., Wei, C., Gaidon, A., Arechiga, N., & Ma, T. (2019). Learning imbalanced datasets with label-distribution-aware margin loss. In *Advances in Neural Information Processing Systems (NeurIPS)*. https://arxiv.org/abs/1909.07413
9. Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104
10. Denney, D. (2010, July). Modeling well performance in shale-gas reservoirs. *Journal of Petroleum Technology*, *62*(7), 47–48. https://doi.org/10.2118/0710-0047-JPT
11. Esfandi, T., Sadeghnejad, S., & Jafari, A. (2024). Effect of reservoir heterogeneity on well placement prediction in CO2-EOR projects using machine learning surrogate models: Benchmarking of boosting-based algorithms. *Geoenergy Science and Engineering*, *233*, Article 212564. https://doi.org/10.1016/j.geoen.2023.212564
12. Espinoza, D. N., Shovkun, I., Makni, O., & Lenoir, N. (2016). Natural and induced fractures in coal cores imaged through X-ray computed microtomography: Impact on desorption time. *International Journal of Coal Geology*, *154–155*, 165–175. https://doi.org/10.1016/j.coal.2015.12.012
13. Feng, H., Wang, Y., Li, Z., Zhang, N., Zhang, Y., & Gao, Y. (2023). Information leakage in deep learning-based hyperspectral image classification: A survey. *Remote Sensing*, *15*(15), 3793. https://doi.org/10.3390/rs15153793
14. Feng, Y., Jia, L., Zhang, J., & Chen, J. (2024). FFSwinNet: CNN-transformer combined network with FFT for shale core SEM image segmentation. *IEEE Access*, *12*, 73021–73032. https://doi.org/10.1109/ACCESS.2024.3392421
15. Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*(5), 378–382. https://doi.org/10.1037/h0031619
16. Foody, G. M. (2023). Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient. *PLOS One*, *18*(10), Article e0291908. https://doi.org/10.1371/journal.pone.0291908
17. Gong, L., Liu, K., & Ju, W. (2023) [Editorial]. Editorial: Advances in the study of natural fractures in deep and unconventional reservoirs. *Frontiers in Earth Science*, *10*, Article 1096643. https://doi.org/10.3389/feart.2022.1096643
18. He, C., Sadeghpour, H., Shi, Y., Mishra, B., & Roshankhah, S. (2024). Mapping distribution of fractures and minerals in rock samples using Res-VGG-UNet and threshold segmentation methods. *Computers and Geotechnics*, *175*, Article 106675, Article 106675. https://doi.org/10.1016/j.compgeo.2024.106675
19. He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

20. Idowu, N. A. A., Nardi, C., Long, H., Varslot, T., & Øren, P. E. E. (2014). Effects of segmentation and skeletonization algorithms on pore networks and predicted multiphase-transport properties of reservoir-rock samples. *SPE Reservoir Evaluation and Engineering*, *17*(4), 473–483. https://doi.org/10.2118/166030-PA
21. Jamshidi Gohari, M. S., Emami Niri, M., Sadeghnejad, S., & Ghiasi-Freez, J. (2023). An ensemble-based machine learning solution for imbalanced multiclass dataset during lithology log generation. *Scientific Reports*, *13*(1), Article 21622. https://doi.org/10.1038/s41598-023-49080-7
22. Jamshidi Gohari, M. S., Niri, M. E., Sadeghnejad, S., & Ghiasi-Freez, J. (2023). Synthetic graphic well log generation using an enhanced deep learning workflow: Imbalanced multiclass data, sample size, and scalability challenges. *SPE Journal*, *29*(1), 1–20. https://doi.org/10.2118/217466-PA
23. Khan, M. S., Ali, H., Villazan Garcia, L., Badshah, N., Trattnig, S., Schwarzhans, F., Woitek, R., & Zaric, O. (2025). Variational U-Net with local alignment for joint tumor extraction and registration (VALOR-Net) of breast MRI data acquired at two different field strengths. *arXiv*. arXiv:2501.13690. https://doi.org/10.48550/arXiv.2501.13690
24. Li, C., Zou, C., Peng, C., Lan, X., & Zhang, Y. (2024). Intelligent identification and segmentation of fractures in images of ultrasonic image logging based on transfer learning. *Fuel*, *369*, Article 131694, Article 131694. https://doi.org/10.1016/j.fuel.2024.131694
25. Li, Z., Zhang, G., Zhang, X., Zhang, X., Long, Y., Sun, Y., & Lin, C. (2025). EILnet: An intelligent model for the segmentation of multiple fracture types in karst carbonate reservoirs using electrical image logs. *Natural Gas Industry B*, *12*(2), 158–173. https://doi.org/10.1016/j.ngib.2025.03.002
26. Lin, W., Li, X., Yang, Z., Lin, L., Xiong, S., Wang, Z., Wang, X., & Xiao, Q. (2018). A new improved threshold segmentation method for scanning images of reservoir rocks considering pore fractal characteristics. *Fractals*, *26*(2), Article 1840003. https://doi.org/10.1142/S0218348X18400030
27. Ma, Z., He, X., Sun, S., Yan, B., Kwak, H. et al. (2023). Zero-shot digital rock image segmentation with a fine-tuned segment anything model [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2311.10865
28. Marques, A., Racolte, G., de Souza, E. M., Domingos, H. V., Horota, R. K., Motta, J. G., Zanotta, D. C., Cazarin, C. L., Gonzaga, L., & Veronez, M. R. (2021). Deep learning application for fracture segmentation over outcrop images from UAV-based digital photogrammetry. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 4692–4695). IEEE. https://doi.org/10.1109/IGARSS47720.2021.9553232
29. Meyer, F. (1994). Topographic distance and watershed lines. In *Signal Processing*, *38*(1), (113–125). https://doi.org/10.1016/0165-1684(94)90060-4
30. Mousavi, S. M., Bakhtiarimanesh, P., Enzmann, F., Kersten, M., & Sadeghnejad, S. (2024, January 10). Machine-learned surrogate models for efficient oil well placement under operational reservoir constraints. *SPE Journal*, *29*(1), 518–537. https://doi.org/10.2118/217467-PA
31. Nande, S. B., & Patwardhan, S. D. (2024). Automated reservoir characterization of carbonate rocks using deep learning image segmentation approach. *SPE Journal*, *29*(8), 4356–4375. https://doi.org/10.2118/219769-PA
32. Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076
33. Phan, J., Ruspini, L. C., & Lindseth, F. (2021). Automatic segmentation tool for 3D digital rocks by deep learning. *Scientific Reports*, *11*(1), Article 19123. https://doi.org/10.1038/s41598-021-98697-z
34. Pun, T. (1980). A new method for grey-level picture thresholding using the entropy of the histogram. *Signal Processing*, *2*(3), 223–237. https://doi.org/10.1016/0165-1684(80)90020-1
35. Punn, N. S., & Agarwal, S. (2022). Modality specific U-Net variants for biomedical image segmentation: *A survey*. *Artificial Intelligence Review*, *55*(7), 5845–5889. https://doi.org/10.1007/s10462-022-10152-1
36. Rahman, M. A., & Wang, Y. (2016). Optimizing intersection-over-union in deep neural networks for image segmentation. In G. Bebis (Eds.), *Advances in visual computing* (pp. 234–244). Springer. https://doi.org/10.1007/978-3-319-50835-1_22
37. Reinhardt, M., Jacob, A., Sadeghnejad, S., Cappuccio, F., Arnold, P., Frank, S., Enzmann, F., & Kersten, M. (2022). Benchmarking conventional and machine learning segmentation techniques for digital rock physics analysis of fractured rocks. *Environmental Earth Sciences*, *81*(3), 71. https://doi.org/10.1007/s12665-021-10133-7
38. Sadeghnejad, S., Hupfer, S., Pingel, J., Lanyon, B., Schneeberger, R., Blechschmidt, I., Alonso, U., Hauser, W., Kraft, S., Geckeis, H., & Schäfer, T. (2025). Bentonite mass loss in fractured crystalline rock quantified from CT scans using digital rock physics and machine learning: Case study from the Grimsel Test Site (Switzerland). *Applied Clay Science*, *276*, Article 107915. https://doi.org/10.1016/j.clay.2025.107915

39. Sadeghnejad, S., Masihi, M., & King, P. R. (2013). Dependency of percolation critical exponents on the exponent of power law size distribution. *Physica. Part A*, *392*(24), 6189–6197. https://doi.org/10.1016/j.physa.2013.08.022
40. Satapathy, S. C., Sri Madhava Raja, N., Rajinikanth, V., Ashour, A. S., & Dey, N. (2018). Multi-level image thresholding using Otsu and chaotic bat algorithm. *Neural Computing and Applications*, *29*(12), 1285–1307. https://doi.org/10.1007/s00521-016-2645-5
41. Shafiabadi, M., & Kamkar-Rouhani, A. (2025). Comparison of edge detection algorithms for automatic identification of fractures in hydrocarbon reservoirs with image logs. *Journal of Mining and Environment (JME)*, *16*(2), 689–704. https://doi.org/10.22044/jme.2024.14190.2642
42. Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Cardoso, M. J. (2017). Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. In M. J. Cardoso et al. (Eds.), *DLMIA/ML-CDS 2017* (LNCS, vol. 10553, pp. 240–248). Springer. https://doi.org/10.1007/978-3-319-67558-9_28
43. Tatbul, N., Lee, T. J., Zdonik, S., Alam, M., & Gottschlich, J. (2018). In. Precision And Recall For Time Series. *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, Montréal, Canada.
44. Teimouri, A., Sadeghnejad, S., & Dehaghani, A. H. S. (2020). Investigation of acid pre-flushing and pH-sensitive microgel injection in fractured carbonate rocks for conformance control purposes. *Oil & Gas Science and Technology—Revue d'IFP Energies Nouvelles*, *75*, Article 52. https://doi.org/10.2516/ogst/2020048
45. Truong, Q. B., & Lee, B. R. (2013). Automatic multi-thresholds selection for image segmentation based on evolutionary approach. *International Journal of Control, Automation and Systems*, *11*(4), 834–844. https://doi.org/10.1007/s12555-011-0055-0
46. Wang, F., & Zai, Y. (2023). Image segmentation and flow prediction of digital rock with U-net network. *Advances in Water Resources*, *172*, Article 104384, Article 104384. https://doi.org/10.1016/j.advwatres.2023.104384
47. Wang, H., Guo, R., Leng, J., Hosseini, S. A., & Fan, M. (2023). A comparative study of deep learning models for fracture and pore space segmentation in synthetic fractured digital rocks. In (S. P. E. Paper (Ed.) (pp. 1–12). San Antonio, TX. https://doi.org/10.2118/215117-MS
48. Wang, Y., Jia, B., & Xian, C. (2023). Machine learning and UNet++ based microfracture evaluation from CT images. *Geoenergy Science and Engineering*, *226*, Article 211726. https://doi.org/10.1016/j.geoen.2023.211726
49. Wang, Z., Ning, X., & Blaschko, M. B. (2023). Jaccard metric losses: Optimizing the Jaccard index with soft labels. In *Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS 2023)*.
50. Wang, Z., Zhang, Z., Bai, L., Yang, Y., & Ma, Q. (2022). Application of an improved U-Net neural network on fracture segmentation from outcrop images. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, *2022*, (3512–3515). https://doi.org/10.1109/IGARSS46834.2022.9883208
51. Wu, B.-F., Chen, Y.-L., & Chiu, C.-C. (2005). A discriminant analysis based recursive automatic thresholding approach for image segmentation. *IEICE Transactions on Information and Systems*, *E88–D*(7), 1716–1723. https://doi.org/10.1093/ietisy/e88-d.7.1716
52. Xu, Y., Cavalcante Filho, J. S. A., Yu, W., & Sepehrnoori, K. (2017). Discrete-fracture modeling of complex hydraulic-fracture geometries in reservoir simulators. *SPE Reservoir Evaluation and Engineering*, *20*(2), 403–422. https://doi.org/10.2118/183647-PA
53. Yin, X.-X., Sun, L., Fu, Y., Lu, R., & Zhang, Y. (2022). U-Net-based medical image segmentation. *Journal of Healthcare Engineering*, *2022*, Article 4189781, 16 pages. https://doi.org/10.1155/2022/4189781
54. Zhang, Z., Wu, C., Coleman, S., & Kerr, D. (2020). DENSE-Inception U-net for medical image segmentation. *Computer Methods and Programs in Biomedicine*, *192*, Article 105395. https://doi.org/10.1016/j.cmpb.2020.105395
55. Zhao, L., Zhang, H., Sun, X., Ouyang, Z., Xu, C., & Qin, X. (2024). Application of ResUNet-CBAM in thin-section image segmentation of rocks. *Information*, *15*(12), Article 788. https://doi.org/10.3390/info15120788
56. Zheng, J., Qi, Z., Liu, J., Ma, D., & Ju, Y. (2022). Segmentation of micro-cracks in fractured coal based on convolutional neural network. *Mining Science and Technology*, *7*(6), 680–688. https://doi.org/10.19606/j.cnki.jmst.2022.06.005
57. Zhou, Z., Zheng, C., Liu, X., Tian, Y., Chen, X., Chen, X., & Dong, Z. (2023). A dynamic effective class balanced approach for remote sensing imagery semantic segmentation of imbalanced data. *Remote Sensing*, *15*(7), 1768. https://doi.org/10.3390/rs15071768