Machine-Learning-Assisted Segmentation of Focused Ion Beam-Scanning Electron Microscopy Images with Artifacts for Improved Void-Space Characterization of Tight Reservoir Rocks

Andrey Kazak and **Kirill Simonov,** Center for Hydrocarbon Recovery, Skolkovo Institute of Science and Technology; and **Victor Kulikov,** PicsArt Inc. and Skolkovo Institute of Science and Technology

Summary

The modern focused ion beam-scanning electron microscopy (FIB-SEM) allows imaging of nanoporous tight reservoir-rock samples in 3D at a resolution up to 3 nm/voxel. Correct porosity determination from FIB-SEM images requires fast and robust segmentation. However, the quality and efficient segmentation of FIB-SEM images is still a complicated and challenging task. Typically, a trained operator spends days or weeks in subjective and semimanual labeling of a single FIB-SEM data set. The presence of FIB-SEM artifacts, such as porebacks, requires developing a new methodology for efficient image segmentation. We have developed a method for simplification of multimodal segmentation of FIB-SEM data sets using machine-learning (ML)-based techniques.

We study a collection of rock samples formed according to the petrophysical interpretation of well logs from a complex tight gas reservoir rock of the Berezov Formation (West Siberia, Russia). The core samples were passed through a multiscale imaging workflow for pore-space-structure upscaling from nanometer to log scale. FIB-SEM imaging resolved the finest scale using a dual-beam analytical system. Image segmentation used an architecture derived from a convolutional neural network (CNN) in the DeepUNet (Ronneberger et al. 2015) configuration. We implemented the solution in the Pytorch® (Facebook, Inc., Menlo Park, California, USA) framework in a Linux environment. Computation exploited a high-performance computing system.

The acquired data included three 3D FIB-SEM data sets with a physical size of approximately $20 \times 15 \times 25\,\mu m$ with a voxel size of 5 nm. A professional geologist manually segmented (labeled) a fraction of slices. We split the labeled slices into training, validation, and test data. We then augmented the training data to increase its size. The developed CNN delivered promising results. The model performed automatic segmentation with the following average quality indicators according to test data: accuracy of 86.66%, precision of 54.93%, recall of 83.76%, and F_1 score of 55.10%. We achieved a significant boost in segmentation speed of 14.5 megapixel (MP)/min. Compared with 0.18 to 1.45 MP/min for manual labeling, this yielded an efficiency increase of at least 10 times.

The presented research work improves the quality of quantitative petrophysical characterization of complex reservoir rocks using digital rock imaging. The development allows the multiphase segmentation of 3D FIB-SEM data complicated with artifacts. It delivers correct and precise pore-space segmentation, resulting in little turn-around-time saving and increased porosity-data quality. Although image segmentation using CNNs is mainstream in the modern ML world, it is an emerging novel approach for reservoir-characterization tasks.

Introduction

To date, the worldwide petroleum industry considers low-permeability (tight) rocks, such as organic-rich mudstones (shale formations), as reservoirs containing substantial resources and reserves of hydrocarbons (oil, gas, and condensate; EIA 2017). Such unconventional reservoirs differ from conventional ones by the contribution of nanoscale pores (Nelson 2009) in terms of size, shape, connectivity to the total effective storage capacity (porosity), transport properties (permeability), spatial heterogeneity, and anisotropy (Kuila 2013). Thus, the reliable and comprehensive evaluation of reservoir quality remains one of the cornerstone questions for the final field-development decision.

Hydrocarbon-resource assessment followed by reserve estimation and evaluation of reservoir potential requires not only the application of routine (*API RP 40* 1998) but also special (Luffel and Guidry 1992; Durand et al. 2019; Kazak and Kazak 2019; Nikolaev and Kazak 2019) petrophysical core-analysis methods. However, the effective properties determination on the core scale is not enough to be applied to well-log-data interpretation, routine-core-analysis techniques, and classic fluid-flow modeling. A viable workaround is using novel techniques involving direct imaging of the rock's void-space structure. One of the emerging and promising approaches in this area is digital rock physics (DRP).

DRP, also referred to as digital core analysis and digital rock, is a technology permitted to run the numerical simulation and extract petrophysical parameters from a digital twin of real reservoir-rock samples. Throughout the last two decades, the need for understanding pore-scale features and processes and macroscopic petrophysical properties calculation contributed to the rapid development of the DRP technology (Andra et al. 2013b). The key paradigm of DRP is "image and compute." The approach involves imaging and digitizing the pore space and mineral-organic matrix of natural rocks and then numerically simulating various physical processes in the digital model to obtain macroscopic rock properties, including absolute permeability (Sun et al. 2017), electrical conductivity (Liu et al. 2009), and elastic moduli (Andra et al. 2013b). Globally, the DRP workflow consists of three-step flow (Dvorkin et al. 2011): imaging of rock plugs allowing us to resolve pore-scale features; processing the raw images to eliminate the noise and segment voids space and mineral/organic-matter/fluid-matrix phases; and numerical simulation of effective properties to obtain cross-property relations. To date, many

petroleum industry and academic research and development groups focus on multiscale (non)continuum fluid-flow and transport phenomena in shales. The toolbox includes particle-based molecular dynamics, dissipative particle dynamics, lattice Boltzmann method, or mesh-based computational-fluid-dynamics fluid-flow and transport simulations (Bultreys et al. 2016; Xiong et al. 2016; Xia et al. 2017; Wang et al. 2020). Nevertheless, the quality numerical simulation requires reliable knowledge of the structure and the interconnectivity of highly complex hydrocarbon-flow pathways.

The general physical trend suggests that higher imaging resolution requires smaller dimensions of the studied sample, and vice versa. Thus, a property determined on a fine (void) scale cannot be directly transferred to a coarse (core plug, whole core) scale. One way to scale up the petrophysical properties is to search thoroughly for representative elementary volume (REV; Bear 1988). In geoscience, many rock types spanning from highly porous and highly permeable conventional terrigenous and carbonate reservoir rocks to unconventional shales tend to possess REVs (Ringrose et al. 2008). It is worth noting that REV depends not only on the rock type but also on a property. For example, REV for porosity is not the same as REV for permeability. Searching for and mapping the REVs requires multiscale imaging.

Although heterogeneity of complex tight reservoir rocks can be observed and imaged over multiple scales by a variety of scientific digital-imaging techniques (i.e., light, X-ray, or electron microscopy), the dual-beam/cross-beam FIB-SEM serial sectioning provides the best 3D resolution for nanopore and micropore imaging (Goral et al. 2016). Within the DRP workflow, FIB-SEM became a widely applicable, although invasive, but high-resolution method for 3D imaging a sample's structure at microscale to nanoscale. The operation of milling and imaging by FIB runs ion-particle and solid-body interactions (Giannuzzi and Stevie 2005). Modern laboratory setups, so-called dual-beam instruments, include an ion column integrated with an electron column in a single device. The resulting FIB-SEM image stack reveals the 3D microstructure of the rock sample of interest. When applied to reservoir rocks holding a complex void-space structure, FIB-SEM delivers digital rock models with voxel size down to $2.5 \times 2.5 \times 5\,\mathrm{nm}$ (Goral et al. 2018, 2019, 2020).

An essential step within the DRP workflow is the building of 2D and 3D digital rock models that, in turn, heavily rely on image segmentation. Here we consider segmentation as a process of partitioning a digital image into multiple segments and assigning a label to every pixel (px) in an image such that pixels with the same label share specific characteristics (Frucci et al. 2008). Adequate characterization of the pore geometry leads to correct spatial void-space-structure characterization and the results of DRP simulations. Because the accuracy of petrophysical properties calculated using DRP heavily relies on the quality of reconstructed digital rock models, any performance gap in image processing or segmentation will lead to misleading reservoir-quality assessment. For example, incorrect segmentation can result in overestimating or underestimating porosity by a few percent, consequently introducing significant errors in reserve-estimation calculations. An inaccurate segmentation can also result in more significant errors in permeability simulations because permeability is highly sensitive to small changes in critical pore-throat diameter (Goral et al. 2018). For these reasons, segmentation is the crucial step while preparing a quality digital rock model.

FIB-SEM tomography has particular challenges and artifacts that are qualitatively different from those of, for example, X-ray computed tomography (CT; Kelly et al. 2016). In the presence of artifacts in general and porebacks in particular, the quality DRMs based on FIB-SEM images require applying dedicated segmentation techniques different from those for other imaging methods. Quality and reliable segmentation remain a challenging problem, however, to a large extent because of the shine-through effect. Our literature survey shows a shortage of feasible automated segmentation methods developed and successfully implemented in practice.

The presented research work aims at improving the quality of quantitative petrophysical characterization of complex reservoir rocks using digital rock imaging. The work has several tasks. First, we evaluate the modern ML methods for an objective (independent of the operator), accurate, and automated segmentation of FIB-SEM reservoir-rock images complicated with poreback artifacts. Second, we develop a feasible model with higher performance than currently used manual and semiautomated segmentation techniques in both segmentation quality and computing time.

Overall, high-quality FIB-SEM image segmentation is a complicated and challenging task. FIB-SEM artifacts, textural contrast, and the nature of imaging show the necessity for developing a new methodology for image segmentation. The conducted literature review indicates that ML is a promising approach to solve this kind of segmentation problem. Therefore, we dedicate this work to test the feasibility of ML-based segmentation of complex FIB-SEM 3D data sets.

State of the Art

Here we provide a concise summary of the most appropriate FIB-SEM and digital-rock-image-segmentation techniques in the image-processing/analysis domain. Although the described items are mostly general-purpose, their description provides a good starting point for the problem's current state. The review also provides references for more advanced segmentation methods.

Segmentation of FIB-SEM Images: Issues and Features. Application of the traditional image-segmentation methods to FIB-SEM images might induce significant errors (Andra et al. 2013a). In general, raw data sets include physical and nonphysical artifacts and noise, leading to low-quality segmentation results using conventional techniques. While reviewing digital images, the human visual-perception system automatically smooths noise and recognizes specific patterns (features) but ignores artifacts. However, the process is difficult to formalize and implement in a computational workflow. Naturally, there was one common way to handle this problem: laborious and time-consuming manual segmentation. However, over time, data sets become larger and digital images get complex and multidimensional (multimodal), while manual approaches consume much human and time resources.

Thus, FIB-SEM imaging requires applying dedicated segmentation techniques different from those for other imaging methods (data-acquisition types). Mostly, it is associated with the presence of image artifacts, which are intrinsic for FIB-SEM images. We here provide a brief overview of primary FIB-SEM image artifacts and describe modern conventional techniques for artifact-removal/mitigation strategies.

- The most common/prominent challenge during FIB-SEM imaging of porous materials is the so-called "poreback" artifact (also referred to as the "pore-shine" or "shine-through" effect; Kelly et al. 2016; Taillon et al. 2018). The artifact occurs if the resulting SEM images contain information from a planar cross section of the material and the underlying exposed subsurface pores. Consequently, FIB-SEM produces a stack of 2.5D rather than 2D images.
- During imaging of poorly or nearly nonconductive materials, electron-beam exposure can accumulate an electric charge on the sample surface. The phenomenon causes the appearance of anomalously bright regions in the images. The electron accumulations bring a net positive charge in areas because of the absence of the conduit path. The effect commonly appears on FIB-SEM images as anomalous bright spots, warpings, and nonphysical fissures. This effect, also known as a "sample-charging" artifact, can be reduced using low electron-beam energy and imaging with a backscattered electron-signal detector (Holzer et al. 2004).

- Inner-structure dissimilarity and local variations in hardness or thickness within the sample result in varying ion-milling rates locally and lead to a nonplanar cross section (ion channeling). The "curtaining" artifacts appear in SEM images as lines parallel to the ion beam (vertical striation). Any inhomogeneity can lead to such an artifact, but it is widespread when there is an open pore structure. Reservoir porous rocks are complicated heterogeneous aggregates with different mineral and element compositions and complex inner geometry of phases. The heterogeneity influences the milling because various materials have different milling rates, and thus, milled surfaces can be uneven throughout the sample. Lowering the milling rate or depositing a platinum coating on the sample surface reduces the effect (Giannuzzi and Stevie 2005).
- The redeposition effect occurs if the milled material that still presents in the chamber is deposited back onto the cross section. A lower milling rate typically reduced the artifact (Giannuzzi and Stevie 2005). Moreover, possible pore destruction and generation caused by ion-beam milling are still in question.
- Shadowing effects occur if the surrounding material blocks the electron-beam path (shadows the cross sections). Milling trenches on both sides of the cross section before imaging typically address this issue (Holzer et al. 2004).
- The electron-signal-detection efficiency from the bottom of the slices is mainly lower because of the longer path of electrons from the beam to the detector (Joos et al. 2011). The feature appears as the artificial-intensity gradient in the FIB-SEM images. At the experimental stage, changing the sample geometry mitigates the effect (Schaffer and Wagner 2008). An alternative way to compensate for the artifact is special image processing (Taillon 2016; Taillon et al. 2018).

Fager et al. (2020), Taillon et al. (2018), and Reuteler (2017) provide a more comprehensive review of the FIB-SEM experimental challenges, as well as a protocol for 3D imaging of soft, porous, and poorly conducting materials.

Image artifacts lead to difficulties in FIB-SEM data analysis. The poreback (topography-induced contrast) artifacts lead to misinterpretation of the pore-adjacent region. FIB-SEM slicing brings void space into the image. However, the depth of the poreback can be low, and electrons are not trapped to illustrate the black pores that we are supposed to see in this region, and alternately show this region as bright. Otherwise, the redeposited material can fill up the imaged voids while milling. In this case, the gray intensity value represents the matrix behind the void space instead of the void itself. Thus, these regions obtain the image-body characteristics that further lead to misclassification and wrong image properties. During manual segmentation, a human interpreter typically distinguishes pore borders. However, this is a nontrivial challenge for automated segmentation because pixel-intensity criteria are insufficient for proper interpretation.

Besides, one of the critical issues in shale imaging is the correct interpretation of pores and kerogen. Relatively smaller nanoscale pores, which can be inside of kerogen, correspond to darker regions. At the same time, relatively large microscale pores illustrate a 3D structure depicted in the 2D surface with different shades of gray that simultaneously are complicated with porebacks. Thus, a differentiation of pores according to grayscale values using either global or local thresholding does not deliver adequate results.

There are several specialized image-processing methods to analyze and segment FIB-SEM data. The toolbox includes adaptive thresholding (Blayvas et al. 2006); level sets (Jørgensen et al. 2010); morphological image processing (Prill et al. 2013); local threshold backpropagation (Salzer et al. 2012, 2014); watershed segmentation (Taillon et al. 2018); workflow consisting of watershed, variance-filtering, and morphological operations and the optical flow-based method (Moroni and Thiele 2020). Besides, Salzer et al. (2015) provides a comparison of some approaches. However, FIB-SEM image segmentation remains a challenging problem, largely because of the shine-through effect that leads to two complications. The first is uncertainty in the positioning of microstructural features along the axis perpendicular to the cross sections. The second is an overlap of grayscale intensities between solid and pore regions (Röding et al. 2020). We here summarize the several most common algorithms to segment FIB-SEM images of porous materials in the presence of porebacks.

The FIB-SEM imaging technique and the porous-media presence imply a vision of structure on the 2D image before it appears in a section. For this reason, it is challenging even for the human interpreter to distinguish the prolongation of the void structure in the z-direction. Typically, an experienced interpreter performs segmentation in such complicated cases. Salzer et al. (2012) presented an original local-threshold-backpropagation (LTB) approach. Analysis of the FIB-SEM image stack identifies a void-space structure's last occurrence according to the dramatic decrease of gray values. The corresponding fall marks the first appearance of the target object in the z-direction. Afterward, the algorithm estimates the feasible threshold for the last appearance. On the last step, the algorithm propagates the local threshold back to the previous slices, giving the name of the method. The threshold estimation and consequent backpropagation work as designed if the void structure's last appearance is inside the stack. Thus, the process fails for void structures penetrating an image stack in the z-direction. The technique compares intensity values in the z-direction and detects the appearance and disappearance of structures according to the comparison results rather than absolute-intensity values. The method produces excellent results when it is up to discovering structures in the z-direction. Still, additional image filtering is required. In particular, the foreground's dilation adds small structure bounds and results in expanding the foreground into the background. The LTB has the advantage to account for the nature of FIB-SEM images concerning the structure's visibility located on different layers. However, although the global threshold removes foreground edges partially, it can lead to the full destruction of small-sized foreground objects.

Prill et al. (2013) introduced a sophisticated multistage morphological segmentation algorithm (Morph) with five tuning parameters. The workflow includes contrast enhancement, detection of morphological features in the z-profiles of the images indicating matrix, void space, or artifacts followed by constrained watershed transformation. The authors report an overall error rate of 12.3% on the synthetic 3D FIB-SEM data set.

Salzer et al. (2014) proposed the valley detection algorithm that is somewhat an extension to the LTB approach. The LTB algorithm relies on the assumption that the gray intensity slightly fluctuates within the observed structure. However, not all data sets can meet this condition. Geometry and various material properties can lead to the sharp variations of gray intensity. For instance, oblique and rough surfaces have higher gray intensity, while roundish smoothed surfaces tend to have lower gray intensity values. Besides, porebacks influence thin structures greatly, and the latter tend to have lower gray intensity values. Valley detection attempts to mark the target structures' first and last occurrences by choosing local minima and maxima in the structure's scope. The corresponding intensity, "the valley" along z-direction, marks the phase separating two distinct structures. The method considers the existence of local minima at the last slice capturing the void structure.

The reviewed segmentation methods' common feature is that each technique is fit for purpose and developed for a specific material (ceramic, thin film, soft porous materials, and so forth). We found nearly no published reports on methods for complex reservoir-rock images, showing that the methods can potentially deal with rock-specific artifacts. Our experience in testing the methods, particularly LTB (Salzer et al. 2012) on the acquired FIB-SEM data sets, shows that the void microstructures are quite complicated and cannot be handled by the published technique. The issue turned out to be far from trivial and further development is required. In the general case, when dealing with heterogenous multiphase rock images, we need a less formalized (constrained) and more general adaptive approach.

ML Approach. The last 20 years show the gradual development of the ML approach in data science. The previous decade revealed a technological boom, and ML methods started to occupy the market. One of the main factors of ML's rise is the extensive growth of computational power. Mostly, it addresses advanced parallel computing using an array of general-purpose graphics processing units, with a representative example of the NVIDIA® Compute Unified Device Architecture (CUDA) programming model (NVIDIA, Santa Clara, California, USA).

One of the ML image-segmentation-framework's core technologies is artificial neural networks (ANNs). The ANN's working principle is more or less the same between all their types: to provide the ANN with learning similar to the human brain's cognitive process. The ANN can include a large number of connected layers that have their weight. In general, ANN-based image-segmentation methods form two main classes: supervised and unsupervised. Supervised methods require manually segmented data for training, so the quality of training directly depends on the input-data quality. Unsupervised methods are semiautomatic or fully automatic. User interaction might be required at some stage to assess or improve performance, but mainly the result should be more or less human independent. Zurada (1992) and Gurney (1997) provide both an introductory and advanced-level description of ANN mechanics.

The efficiency of similar algorithms applied to microscopy image segmentation implies generating random classifiers that work in high-dimensional space compared with those handling only 2D image output from the microscope's detector. Thus, so-called multimodal images result from data integration of different imaging and computational sources. The data sources include image acquisition using other detectors, voltages, energy levels, and angles. The additional dimensions can hold the original images after additional filtering (e.g., gradient smoothing, textural and compositional transformations) to extract different local and nonlocal features. The extracted features form a single vector to characterize every pixel according to local, nonlocal, or textural data.

Application of ML-based segmentation took off from the biomedical domain. Kreshuk et al. (2011) used automated detection and segmentation methods to study synaptic contacts in a nearly isotropic FIB-SEM data set. The software stack included an interactive learning and segmentation toolkit ilastik (Sommer et al. 2011) featuring a family of generic (nonlinear) image features and a robust nonlinear classifier (Breiman 2001) to estimate the probability of belonging to the classes individually for each voxel. Quantitative validation with manual annotations showed that the algorithm's error rate is comparable with three independent experts in recall (0.92) and precision (0.89). Caffrey et al. (2019) used the open-source ML-based Trainable Weka Segmentation (TWS) plugin (Arganda-Carreras et al. 2017) in ImageJ (Schindelin et al. 2015) to perform semiautomated 3D segmentation of human skeletal muscle using FIB-SEM images. TWS features a large variety of supervised classification, regression, and clustering algorithms. The features available in TWS include edge detectors (indicating object boundaries in an image), texture filters (extracting texture information), noise-reduction filters, membrane detectors (localizing membrane-like structures of a certain size and thickness), and user-customizable features. The research results showed a dramatic increase in image-analysis speed and improved speed of up to three orders of magnitude than manual approaches for data segmentation. Khadangi et al. (2019) automated the segmentation of cardiomyocyte Z-disks from high-throughput SEM data, including large serial block-face SEM and FIB-SEM. The proposed fully automated segmentation scheme includes three main modules (preprocessing, segmentation, and refinement). Model validation involved comparing the segmentation results with ground-truth annotated Z-disks in terms of pixelwise accuracy. The results showed that the model correctly detects Z-disks with 90.56% accuracy. The model outperformed ilastik in segmenting of the Z-disks accurately within the data set from serial block-face SEM. Maitin-Shepard et al. (2016) introduced an ML approach for image segmentation that uses an ANN to model the conditional energy of a segmentation given an image. The combinatorial energy learning for image segmentation approach relies on modeling the inherent combinatorial nature of dense image-segmentation problems. The application of this approach on a publicly available 3D microscopy neuroanatomic data set with ground-truth data showed that the method improves volumetric reconstruction accuracy by more than 20% compared with two state-of-the-art baseline methods (graph-based segmentation of the output of a 3D CNN trained to predict boundaries, and random forest classifier trained to agglomerate supervoxels that were generated by a 3D CNN). Koebernick et al. (2017) report on successful TWS-driven segmentation of high-resolution synchrotron images of root hairs. The results allowed image-based modeling and studying the effects of pore structure on the diffusion and permeability of roots. Schweizer et al. (2018) used multichannel ML segmentation to study the microspatial arrangement of soil-organic-matter coatings on intact soil-microaggregate structures during various glacier-ecosystem-development stages. The ilastik toolkit helped the supervised pixel classification of secondary ion distributions for SEM images of soil-clay-sized fractions on a silicon wafer.

ML-powered segmentation also plays a vital role in materials science. Converse and Fullwood (2013) proposed using crystallographic orientation data and ML-based methods to enhance the resolution of nanoscale gaps (1–20 nm) in composite materials. The authors used an ML algorithm to predict nanoscale-grain connectivity in a nickel nanostrand/epoxy composite on an FIB-SEM data set. TWS software distinguished between adjacent grain pairs and foreign grain pairs. The results show 71.9% accuracy for a 2D algorithm and 62.4% accuracy in 3D. The work demonstrated how ML-based algorithms could be used to predict the location of gaps between distinct nanostrands—gaps that would otherwise not be detected with the sole use of a scanning electron microscope. Röding et al. (2020) developed a segmentation method for FIB-SEM data acquired on porous films made from ethylcellulose and hydroxypropyl cellulose polymer blends. The segmentation involved scale-space features and a random forest classifier (Breiman 2001). The work features data augmentation and hyperparameter optimization of random forest in terms of the number of trees, the number of features per tree, and the minimum leaf size. The results demonstrate good agreement with the manual segmentation performed by an expert, with accuracy reaching 0.88 on the test data set. The method enabled quantitative characterization and subsequent optimization of the material structure for controlled-release applications.

A separate area of research is the ML-based separation of rock grains and fragments. Bull et al. (2014, 2015) suggested a method to delineate rock fragments by classification of SEM image patches. The method overcame the disadvantages of existing techniques by using an ML approach combined with random compressed features to classify small image patches. Approximately 85% of rock fragments are classified correctly. The method compares favorably with the watershed algorithm and has the advantage of not needing markers. The compressed feature's small size improved the training time by a factor of 10 and classification times by more than 25, compared with using uncompressed image patches. Later, Iwaszenko and Nurzynska (2019) and Nurzynska and Iwaszenko (2020) successfully approached rock-grain segmentation on bulk material images using curvilinear structure-based features. The feature vector includes the pixel color, pixel's neighborhood average, standard deviation, vesselness, and vesselness-scale values. ML classifiers had k-nearest neighbors, support-vector machine (SVM), and ANN. The results helped to discover the best features and optimal parameters of the classification methods. The linear SVM delivered the best classification results, but the differences between classifiers were not significant. The obtained results demonstrated accuracy as high as 89% compared with the manually segmented images as a ground truth.

Although these technologies enrich many domains where data science shows robust performance, such as medical diagnosis and target advertisement, the analysis of digital rock images (CT, optical, or electron microscopy) within geoscience is still under

development. A small number of recent publications related to ML-based FIB-SEM image segmentation resulting from our literature survey confirm this statement.

Chauhan et al. (2016a, 2016b) evaluated the performance and accuracy of ML techniques to segment rock grains, matrix, and pore voxels from a 3D volume of micro-CT grayscale rock images. The target methods included the segmentation and classification capability of unsupervised (k-means, fuzzy c-means, self-organized maps), supervised (ANNs, least-squares SVMs), and ensemble classifiers (bragging and boosting). The results showed that the averaged porosity for various rock types (andesite, Berea and Rotliegend sandstones) agrees with the corresponding laboratory measurement data and varies by a factor of 0.2. The study's important conclusion is that, in general, the accuracy largely depends on the feature-vector-selection scheme. Chauhan et al. (2020) introduced a general-purpose, automated, and explicitly tailored for accurate grayscale (multiphase) image segmentation using unsupervised- and supervised-ML techniques: CobWeb 1.0. The work demonstrates image segmentation using unsupervised-ML techniques on Berea Sandstone and Grosmont Carbonate digital rock images. Multiphase image-segmentation results enabled a reliable estimation of the relative porosity and pore-size distributions of the target digital models.

Khan et al. (2016) proposed a beam-hardening artifact-removal method by applying a best-fit quadratic surface algorithm to a given image data set (reconstructed slices). The method's validation involved receiver operating characteristic analysis on both corrected and uncorrected samples and showed that the correction is an essential prerequisite for accurate multiphase classification. Besides, the authors approached micro-CT image segmentation with a least-squares SVM (LS-SVM), an algorithm for pixel-based multiphase classification. Combining the two approaches delivered the quality classification results for three different complex multiphase rock-core samples.

Lormand et al. (2018) used TWS to derive the crystal size distributions of microlites in volcanic rocks. The software implements ML-based image segmentation using the FastRandomForest algorithm, a parallel implementation of the random forest classification technique (Breiman 2001). The data included backscattered-electron SEM images of volcanic-rock samples with different crystallinity with variable image resolutions. The study compares crystal measurements obtained from the automatically segmented images with those of the manual segmentation; samples with crystallinity up to 50 vol% had successful segmentation. In contrast, segmentation at significantly higher crystallinities failed because of the unreliable distinction of crystal boundaries. Accuracy-performance tests for the TWS classifiers yielded high F-scores (>0.930). Hence, the authors concluded that TWS is a successful and fast computing tool for outlining crystals from backscattered-electron images of glassy rocks.

Shu et al. (2018) avoided the need for laborious segmentation while analyzing the sorting level of rock particles by using an off-the-shelf CNN model that has been pretrained on a large-scale image data set to extract feature representations for their rock images. The authors trained an SVM classifier with the feature representations as input to obtain the sorting-level grade. The experiments showed that the CNN features significantly improved the classification results compared with handcrafted features and low-level k-means features.

Jobe et al. (2018) presented two ML-based approaches to classify petrographic features in thin sections. The first case study of carbonate reservoirs demonstrated a capability to predict and classify rock properties, such as lithofacies, reservoir zone, porosity, and permeability. The second approach focused on training and using CNNs to classify and predict Dunham textures (mudstone, wackestone, packstone, and grainstone; Dunham 1962; Embry and Klovan 1971) from thin carbonate sections.

Andrew (2018) addressed the quantitative benchmarking of ML-based classifiers against traditional processing of image workflow for different rock samples from petroleum geoscience. Assessment of the effect of different segmentation techniques involved three 3D (data that are not obtained by direct imaging) data sets synthetically created using real rock-sample images, including the FIB-SEM data set of the Vaca Muerta Shale and micro-CT of Bentheimer Sandstone. The paper compared Otsu thresholding, seeded watershed-region growing, and ML-based segmentation. The comparison highlighted several features. First, the misclassification rate for conventional techniques could be high even at low noise. Second, the ML-based segmentation used textural information forming a vector that allowed differentiating in classification. This technique kept the misclassification rate less than 15% within a 30% noise range. Goral et al. (2018, 2019) used ML segmentation of ultrahigh-resolution (voxel size of $2.5 \times 2.5 \times 5$ nm) FIB-SEM data sets to reconstruct 3D digital rock models for organic-rich shale-rock samples. Segmentation used the "forest of randomized trees" approach (Breiman 2001) and segmentation labels (identified and marked by a user with a paint tool), mask(s), and/or (filtered) data set(s) (Piche et al. 2017). The results included 3D labels for pores, organic matter, carbonate, silicate, and sulfide minerals. Later, (Goral et al. 2020) used the Dragonfly Prosoftware with ML image-segmentation platform for correlative core- to pore-scale digital rock modeling of Mancos Shale rock samples. A set of length scales spun from a few centimeters to a few nanometers. Digital rock analysis used correlative micro-CT, micro-3D X-ray microscopy, light microscopy, SEM, and FIB-SEM image data sets. The obtained multiscale/modal 2D/3D image data sets allowed reconstructing digital rock 2D/3D models and quantifying petrophysical properties (porosity and mineralogy).

Byrnes et al. (2017) and Howard et al. (2019) quantified image-segmentation uncertainty for image-based DRP for a shaly sandstone with a significant amount of authigenic chlorite/smectite that lined the larger pores. The authors studied two strategies, including a binary segmentation with a linear threshold and an ML approach to two-phase segmentation. The target data set included micro-CT 16-bit images collected on the cleaned miniplugs in the dry state with a voxel size of 4.0 and 2.0 µm. The ML engine optimized for a proprietary image-management cloud involved a random forest algorithm used to combine many statistical results extracted from a set of filters, edge detection, and other feature-extraction algorithms that evaluated nearest-neighbor and next-nearest-neighbor relationships for each voxel (Zhang et al. 2019). The research featured cross validation by independent nuclear-magnetic-resonance and mercury-injection capillary-pressure measurements and numerical simulations for basic petrophysical properties. The results showed that the threshold segmentation could misleadingly reach porosity matching the core-analysis data; however, the petrophysical properties derived therein will be completely erroneous with more than 100% uncertainty. ML segmentation limits the uncertainty to less than 30%, offers more flexibility, and does a better job of dealing with pore/grain interfaces.

Wu et al. (2019) reported on the successful experience of applying ML to locate organic matter and pores in SEM images of organic-rich shales. Segmentation related an individual pixel in the target SEM grayscale image to one of the four components: pores/cracks, kerogen/organic matrix, and pyrite. The ML solution used 16 features (including wavelet transform, Gaussian blur, and difference of Gaussians) and a random forest classifier implemented in the Scikit-Learn package (Pedregosa et al. 2011), which uses an optimized classification and regression tree algorithm for building a decision tree. The proposed segmentation methodology could segment a single $2,000 \times 2,600$-px image in approximately 30 seconds. The trained classifier's performance in terms of overall F_1 score on the validation data set was higher than 0.9 (Wu and Misra 2020).

A separate promising direction for image segmentation is called the deep neural network (DNN), which features a higher performance with increased depth or larger hidden-layer stacks (Szegedy et al. 2015; Urban et al. 2016). However, DNNs are not only computationally more intensive but are difficult to train because they have issues like vanishing and exploding gradients (Pascanu et al. 2012, 2013) and filter saturation by highly correlated features. He et al. (2016) improved the gradient propagation and enhanced DNN training

by introducing a residual network. The residual network concatenates an identity map to the output of a convolutional layer stack (residual unit). The improvement facilitates training by targeting this new referenced residual output, avoiding gradient vanishing or saturation. Ronneberger et al. (2015) introduced U-Net architecture, which concatenates feature maps from different layers of the encoding branch to the decoder, improving the segmentation accuracy significantly. The U-Net's structure retains high- (lines and edges) and low-level features (entire objects) to reconstruct the output. The direct pathways between the encoding and decoding branches (skip connections) provide a better flow of information both in the forward and backward computations and facilitate training with fewer parameters. Zhang et al. (2018) presented the deep residual U-Net (ResUnet), which uses residual units as building blocks and skips connections between them. The network proves to be easy to train (compared with the U-Net, which needs extensive data augmentation or a pretrained model), with an efficient number of parameters, and showed accurate results using a small training set.

Tracey et al. (2019) discussed integrating a random forest classifier with the deep-learning residual neural network U-Net (ResUnet) to segment porebacks from an FIB-SEM image stack for a membrane sample with 80% porosity. The training included five to 20 images for each case, resulting in a total of four trials. The results indicate that the training-sample distribution has a significant effect on prediction quality. The augmented ResUnet delivers a segmentation accuracy approaching 90% for 3D image stacks of polymer membranes if the training set is more than 20 images evenly distributed throughout the data set.

Karimpouli and Tahmasebi (2019) segmented digital rock images using deep convolutional autoencoder networks. The target data set included micro-CT images of Berea Sandstone and a training seed comprising manually and semimanually segmented images. The ML solution was an encoder/decoder convolutional network (Badrinarayanan et al. 2017). It is one of the currently best available networks used in different applications and demonstrated promising results. The implemented CNN algorithm produced 96% of categorical accuracy using the designated images in the testing group. A qualitative comparison with the conventional multiphase segmentation (multithresholding) revealed that our results are more accurate and reliable even if very few rock images are available.

In summary, quality FIB-SEM image segmentation is a complicated and challenging task. FIB-SEM artifacts, textural contrast, and the nature of imaging show the necessity for further developing the methodology for image segmentation. The conducted literature review indicates that ML is a promising approach to solve this kind of segmentation problem. To date, ML-based segmentation is a powerful tool for the transformation of challenging image data sets, which can carry a variety of modality-specific artifacts and noises, into segments (labels), representing different groups of features of the rock microstructure (e.g., pores or minerals) that were previously too difficult to label by threshold-based approaches (Andrew et al. 2017). The published results showed good coverage of the ground-truth delineation. However, applying the classic ML approach (using classifiers) prevails over deep learning, and further research is still required. Thus, in this work, we attempt to study the feasibility of DNN-based segmentation of complex FIB-SEM 3D data sets.

Materials

The presented study uses data from a research and development project dedicated to DRP characterization of one promising tight gas formation located in West Siberia, Russia: Lower Berezov Formation. Kazak et al. (2017a, 2017b) provide a comprehensive description of the geology, rock-sample collection, and supplementary data on the samples. For the sake of brevity, we outline only relevant information.

The reservoir rock is mostly represented by dark-gray gaze(opoka)-like clays, siliceous mudstones with rare interbeds of argillaceous siltstone, and fine-grained sandstones. Hydrocarbon accumulations of the Lower Berezov Formation are associated with siliceous remains of radiolarians, shells, frustules of diatoms, spicules of sponges, and so forth that dwelled in shallow sea areas. From the petrophysical perspective, the samples had porosity of 18 to 26%, with an average value of 22%, and gas permeability of 0.2 to 60 md, with an average value of 11 md. High porosity but relatively low gas permeability suggest a large number of microscale voids having interconnecting nanochannels.

The whole core pieces were split into several representative standard petrophysical plugs of Ø30 × 30–90 mm (Ø means diameter of the core plug). After core-plug petrophysical tests, the plugs, in turn, were separated by representative miniplugs of Ø3 × 10 mm. Then each miniplug was cut into two halves: the main and archive. We imaged a flat cut surface of the primary half, identified and marked the target regions, and performed FIB-SEM imaging. The FEI Versa 3D DualBeam (Thermo Fisher Scientific Inc., Hillsboro, Oregon, USA) analytical system acquired the target 3D FIB-SEM models (Kazak et al. 2018).

Raw Data. In this work, we handled three FIB-SEM data sets **(Table 1).** The exact locations of the FIB-SEM models correspond to rock-specific features, which shed light on the void-space nanostructure, and explained the transport and storage capacity of the studied reservoir rock. For example, one of the FIB-SEM models included a contact fracture between the clay-mineral grain and the surrounding siliceous matrix **(Fig. 1).** Another FIB-SEM model captures the void space of a stylolite seam **(Fig. 2).**

ID	SEM Mode/ Detector	Bit Depth	Image Dimension, X (px)	Image Dimension, Y (px)	Pixel Size (nm)	Number of Slices	Slice Thickness (nm)	Image Physical Dimensions, $L \times W \times H$ (μm)
3.1.1	SE	8	3,072	2,048	5.6	100	25	17.3 × 11.5 × 15
7.1.1	SE	8	4,096	3,536	4.5	553	25	18.3 × 15.8 × 15
7.1.2	SE	8	4,096	3,536	4.5	115	25	18.3 × 15.8 × 15

Table 1—Specifications of the acquired FIB-SEM data sets. $L \times W \times H$ = length × width × height; SE = secondary electrons.

It is worth noting that the resulting voxel size had anisotropy because of the difference in spatial resolution in the *x*-, *y*-, and *z*-direction (Table 1) and affected the connectivity of both voids and matrix in the resultant 3D digital rock models. The effect is an intrinsic feature of the FIB-SEM technology, but hardware-tuning and image-processing steps can mitigate it effectively.

Labeled Data. A professional digital rock analyst (petrographer) manually segmented (labeled) the filtered slices using the wand (tracing) tool of the Fiji 1.52 h software package (a package that includes the open-source ImageJ image-processing software). To ensure spatial data consistency, we registered (in terms of coordinates) each label to the corresponding image. We then selected the slices for labeling in such a way to the main specific void-space features and traced them throughout the whole data set **(Fig. 3).** Because of the image-processing features, some of the resulting labels had dimensions smaller than those of the original slices by 30 to 40%. For example, if the original slices had dimensions of 4,096 × 3,536 px, the corresponding labeled slices, on average, had dimensions of 3,700 × 1,600 px.
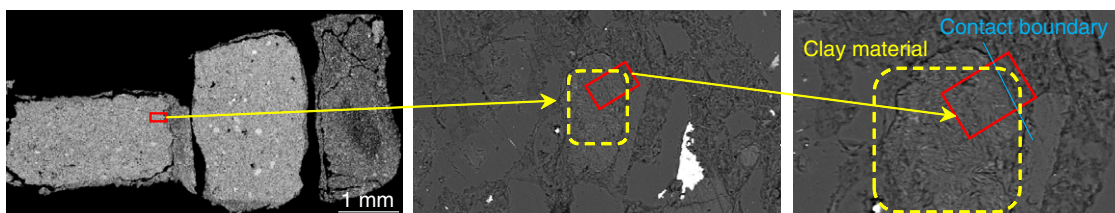
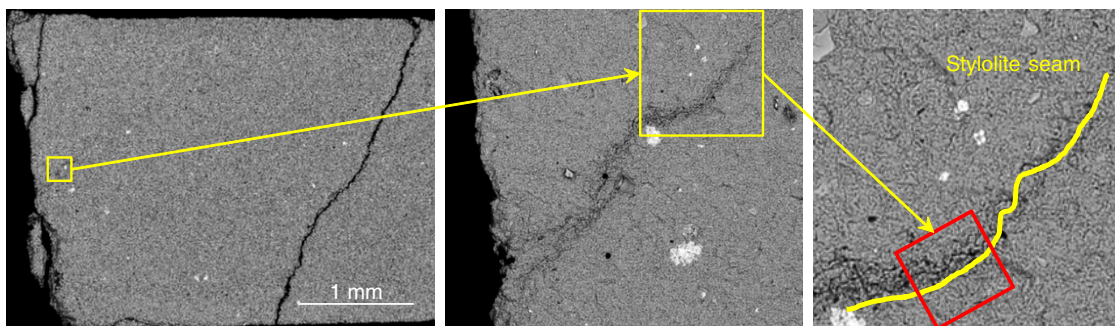**Fig. 1—Location of Data Set 3.1.1 captured the contact zone between rock matrix and clay material.**

Fig. 2—Location of Data Set 7.1.1 capturing the microstructure of a stylolite seam.
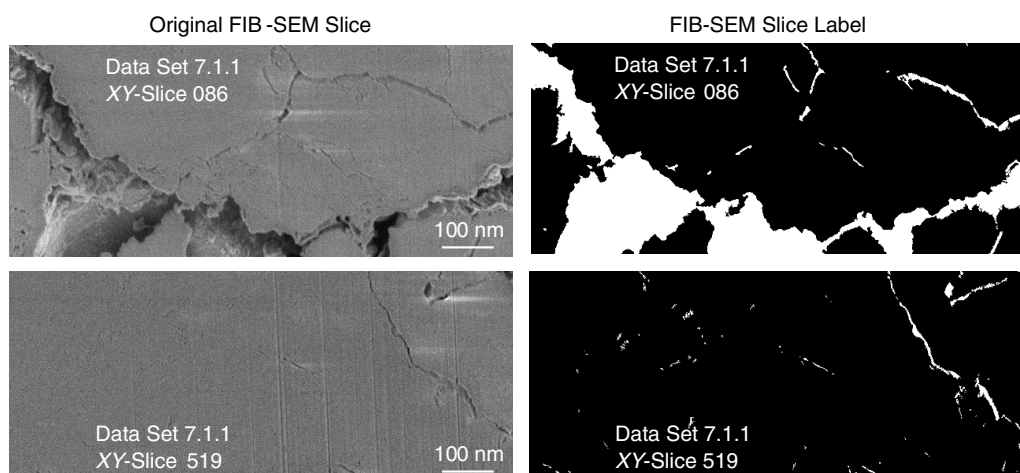
Original FIB -SEM Slice | FIB-SEM Slice Label

Fig. 3—Examples of original slices (left) and the corresponding manual-segmentation results (right).

Before labeling image slices from Data Set 7.1.1, we preprocessed them using FEI PerGeos 1.5.0 software (Thermo Fisher Scientific Inc., Hillsboro, Oregon, USA). The image-processing workflow included fast-Fourier-transform filtration to remove periodical artifacts in the *XY*-slice plane, individual slice alignment, image stack crop to the volume intersecting all slices, nonlocal means smoothing, and unsharp masking filtering to remove high-frequency noise and increase edge contrast. Eventually, we obtained a total of 21 manually labeled *XY* slices, including 19 slices from Data Set 7.1.1, one slice from Data Set 7.1.2, and one slice from Data Set 3.1.1.

Training Data Sets. The training data set (TD) is the sample of data used to fit the model and includes labeled slices. Each labeled slice is a pair of images consisting of an *XY* slice from the acquired and the corresponding label. For training, we used the original raw *XY* slice images without additional processing for two reasons. First, we know that in the general case, any data processing reduces the amount of valuable information (signal) from the input data set and, in the best case, keeps the quantity at its original level. Second, we realize that during training, a DNN captures poorly understood and weakly defined features of the SEM-image data, including that part of the image that a human observer considers to be noise.

The base TD included a total of 10 labeled slices. However, further testing of a single CNN configuration required expanding the base TD with a few extra labeled slices and areas (patches) from the same Data Set 7.1.1 **(Fig. 4).** The patches correspond to regions of the original slices outside of the base TD. Thus, the resulting patch dimensions (2,000–3,000 × 1,000–1,300 px) are smaller than those of the base TD slices.

Validation Data Set and TDs. The validation data set (VD) included a subset of labeled slices, and we used it to evaluate the DNN performance. The primary purpose of VD is model selection (Hastie et al. 2009). During training, DNN has not used the VD images; that is, the network did not know about their existence. If this essential condition fails, then a CNN performance assessment becomes complicated because the model has learned to derive from the TD or has just captured these observations. The base VD included a total of four labeled slices from Data Set 7.1.1.

The test data set (TSD) included three labeled slices from Data Sets 7.1.1, 7.1.2, and 3.1.1. The purpose of the TSD is to evaluate the prediction error of the final selected model (Hastie et al. 2009).
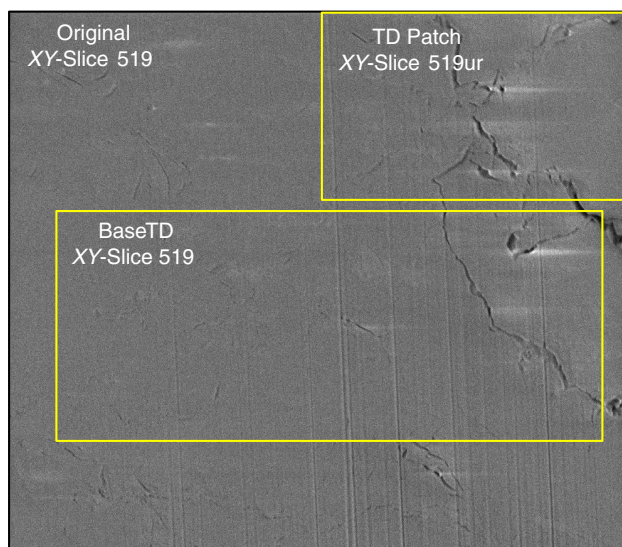
Fig. 4—Example of the original base FIB-SEM slice and extra TD slices for TD expansion.

Methods

Segmentation with Conventional Methods. Semantic segmentation aims to provide image information on a pixel level. The goal of semantic segmentation is to classify image pixels that belong to the same object class, such as voids, minerals of rock matrix, and organic matter.

First, we evaluated the performance of several standard image-segmentation methods. We used the Auto Threshold and Auto Local Threshold plugins of Fiji 1.52h and the Converging Active Contour segmentation of Mango v. 4.0.1 software (Australian National University, Research School of Physics, Canberra, Australia). Automatic threshold methods tend to optimize some objective criteria that can be statistical (e.g., maximization of interclass variance, entropy), probabilistic (e.g., minimization of pixel-classification error), or structural (e.g., circularity of detected objects; Arganda-Carreras 2016). The difference between "local" and "global" versions is that the former computes the threshold for each pixel within a window of chosen radius around it. In contrast, the latter decides the threshold for the whole image. The Converging Active Contour algorithm, part of the Level Sets method, is dependent on partial-differential equations. The algorithm finds object boundaries within the progressive evaluation among neighboring pixels and eventually converges at the object's boundary where the differences are the highest (Schindelin et al. 2015). In both cases, we observe erroneous segmentation in high-grayscale-intensity regions and only partial segmentation of voids **(Fig. 5).**
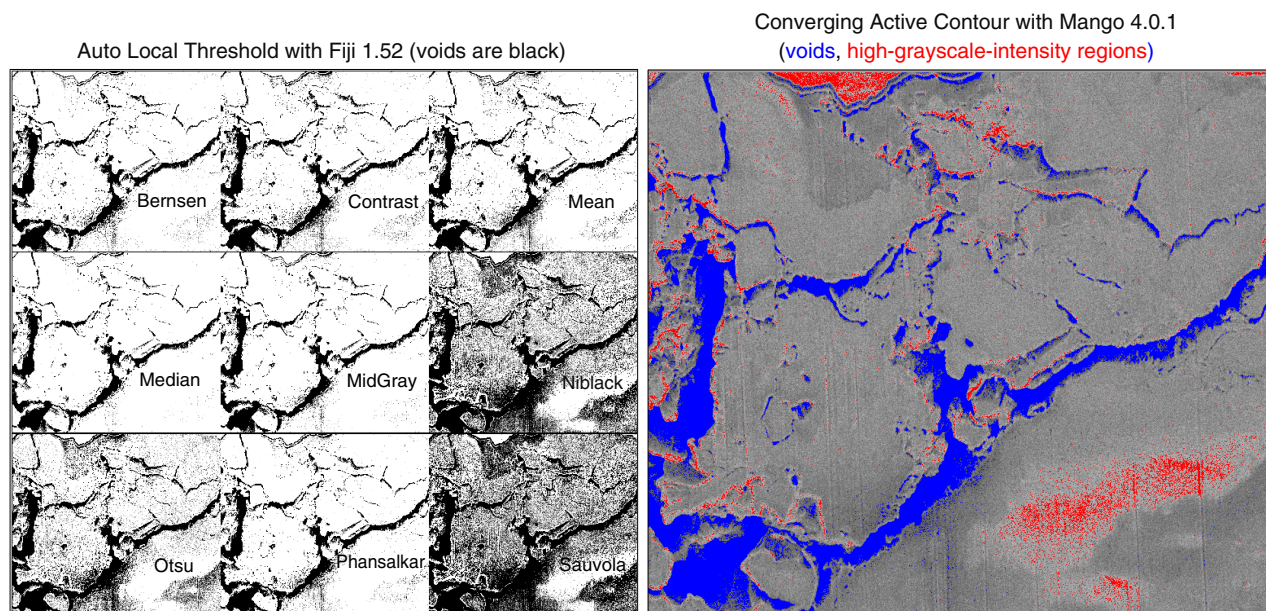

Fig. 5—Segmentation with conventional methods in Fiji and Mango software.

ML-Based Approach. Our next step was an attempt to evaluate commercially available ML-powered image-processing software. Software review showed not many available solutions, and we took the ZEISS® ZEN Intellesis 2.5 (blue edition) (Carl Zeiss Corporation, Oberkochen/Wuerttemberg, Germany). The software implements the "forest of randomized trees" approach (Breiman 2001) and segmentation labels (identified and marked by a user with a paint tool), mask(s), and/or (filtered) data set(s) (Piche et al. 2017). Unfortunately, the Intellesis Trainable Segmentation algorithm failed at the training stage with several TD images. After getting the negative results, we decided to develop a dedicated solution derived from deep-learning technology.

Currently, there are plenty of ANN types and architectures. To attack the problem of segmentation of FIB-SEM rock images, we chose the CNN called DeepUNet (Ronneberger et al. 2015) because its design can handle the segmentation of high-resolution images. DeepUNet (Ronneberger et al. 2015) is an end-to-end fully convolutional network with two other kinds of short connections called U-, and connections with an hourglass architecture to extract image features at different scales. The central concept of DeepUNet is a couple of layers in the contractive track with the extensive track. High-resolution features obtained during the contracting track are integrated into the upsampled output. Thus, a consecutive convolutional layer can learn to agglomerate more precise outputs, adopting the derived features. In addition, to decrease a loss while extracting high-level features, DeepUNet develops the contractive track and an extensive track by proposing the DownBlock, UpBlock, and Plus connection.

Both DownBlock and UpBlock (**Fig. 6**) complement the features before and after passing through the convolution layer. Such a structure yields a deeper and computationally efficient convolution neural architecture (Li et al. 2018). The DownBlock consists of two convolutional layers, sequentially linked to the Batch Normalization layer and concatenated with the rectified linear unit (ReLu) layer. Both convolutional layers apply a 3×3 convolution kernel with 1×1 step size. The input of the block is 64-dimension features. The second convolution-layer output has the same feature size as the input to the DownBlock. The Plus layer sums the data after the convolution.
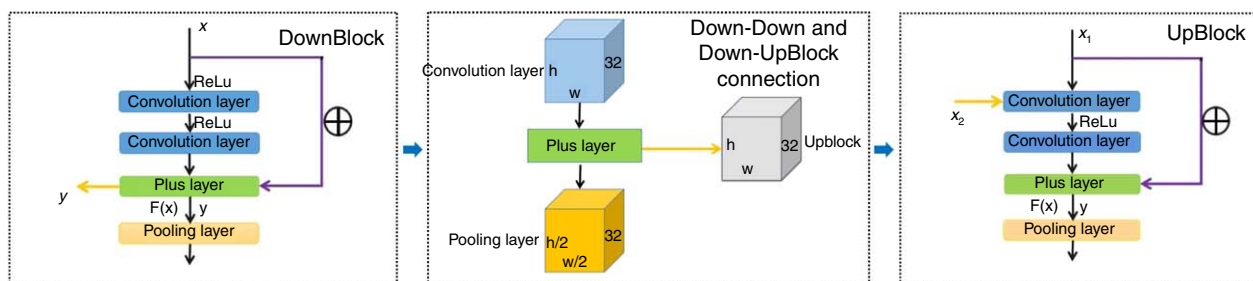

Fig. 6—Internal structure and interconnection of both DownBlock and UpBlock (Li et al. 2018).

Both input grayscale images illustrate reservoir rock with void space and matrix materials, and the output images are binary segmentation maps. CNN is not a fully connected network type and uses only a 1×1 convolution layer for dimension reduction. Finally, the Softmax activation layer is applied to transform the tensor values into maximum arguments and treat them as a two-class problem.

If the Plus layer produces y and x is the input of the DownBlock, the Plus layer passes x and the second convolution output into Eq. 1, and the filtering outputs keep the optimal value y further into the maximum-pooling layer. In Eq. 1, W_1 means the first convolution and W_2 means the second convolution, while σ describes the ReLu function.

$$y = W_{2\sigma}\sigma(W_1 x) + x. \quad \dotfill \quad (1)$$

The maximum-pooling layer in the DownBlock has a kernel size of 2×2 and a step size of 2×2. At this stage, y proceeds to the maximum-pooling layer, and integration of feature maps to the correspondent UpBlock occurs. The DeepUNet maintains a fancy symmetric structure. The UpBlock is filled with feature maps to provide more precise upsampling. The UpBlock mirrors the main components of the DownBlock because it is similarly composed of two convolutional layers and the Batch Normalization and ReLu layers. However, it is clear from the name that the upsampling layer replaced the maximum-pooling one. The integrated feature map x is the input for the convolution layer,

$$x = [\delta, x_1, x_2], \quad \dotfill \quad (2)$$

where x_1 is the feature map from the previous UpBlock, while x_2 is the map from the DownBlock through U-connection and is the upsampling operator. Because of the DeepUNet structure, an integrated feature map comes to the UpBlock before maximum pooling. The convolutional layers process it and update the data to obtain more accurate results during upsampling. To keep the original image dimensions, the resolution of DownBlocks' output should be the same as that of the UpBlocks' input.

Preprocessing of Training Data. The limited number of TD slices makes data augmentation a necessary step. We applied a set of image transforms, including vertical and horizontal flipping, rotation, blurring, histogram normalization, and gray-level expansion. The total number of TD images results from the number of input samples multiplied by the number of image transforms. The base number of TD images is 10, for which the augmentation increased the TD size to 150 images.

Computational Resources. The computation infrastructure relies on a modern high-performance-computing cluster (**Table 2**). The software stack included Python 2.7, CUDA 8.0, and Pytorch 1.1. For coding, we used the Jupyter Notebook.

Central processing unit	Intel® Core™ i7-6700 central processing unit, 3.40 GHz
	(Intel Corporation, Santa Clara, California, USA)
Graphics processing unit	NVIDIA® GTX 1080i TitanBlack (NVIDIA Corporation, Santa Clara, California, USA)
Random-access memory	32 GB
Operating system	Ubuntu® 16.04 (Canonical Limited, Douglas, Isle of Man)

Table 2—Technical specifications of the used high-performance-computing environment.

Tuning Parameters. Because we picked DeepUNet as the central CNN architecture, we focused on tuning its parameters rather than testing other CNN architectures. The latter, however, forms the scope of future research endeavors. There are several parameters that we studied to evaluate the CNN performance and conduct sensitivity analysis (**Table 3**).

Group	Tuning Parameter		Description
Loss function	Cross entropy		Cross-entropy loss or log loss measures a classification configuration's performance, the output of which is a probability value between zero and unity (Murphy 2012).
	Focal loss	$\gamma = 1$	DeepUNet with focal loss function smoothly adjusting the rate at which easy examples are downweighted (Lin et al. 2020).
		$\gamma = 2$	DeepUNet with focal loss function sharply adjusting the rate at which easy examples are downweighted (Lin et al. 2020).
Optimization algorithm	Adam		Adaptive Moment Estimation (Adam) (Kingma and Ba 2015).
	SGD		Stochastic gradient descent (SGD) is an iterative method for optimizing an objective function with suitable smoothness properties (Zhang et al. 2015).
	AMSGrad		AMSGrad optimizer is an alternative from Adam using lower learning rates and consolidating the instinct of slowly decaying past gradients' effect on the learning rate (Tran and Phuong 2019).
TD and VD order	K-folding		Cross validation delivers a more accurate estimate of configuration-prediction performance (Seni and Elder 2010).
Number of TD image pairs	Expansion		Expands TD by adding extra slices.

Table 3—List and description of the DeepUNet tuning parameters.

CNN Configurations. We analyzed each tuning parameter's effect on the CNN output and combined them in eight configurations (**Table 4**). The arrangements have different loss-function settings, optimizer algorithms, and variations of TD and VD. However, we realize that the number of possible configurations exceeds eight, but chose this number because of limited resources.

ID	Configuration	Description
1	CE	DeepUNet with the cross-entropy loss function and AMSGrad optimizer.
2	Fl_y1	DeepUNet with the focal loss function, $\gamma = 1$, and AMSGrad optimizer.
3	Adam	DeepUNet with the focal loss function, $\gamma = 2$, and Adam optimizer.
4	SGD	DeepUNet with the focal loss function and SGD optimizer.
5	K1	DeepUNet with the focal loss function, AMSGrad optimizer, and VD expanded by adding Slices 001, 086, 277, 553.
6	K2	DeepUNet with focal loss function and AMSGrad optimizer and VD expanded by adding Slices 035, 439, 477, 519.
7	K3	DeepUNet with focal loss function and AMSGrad optimizer and VD expanded by adding Slices 101, 439, 519, 553.
8	Retrained	DeepUNet with focal loss function, AMSGrad optimizer, and TD expanded by adding Slices 001ll, 402, 443, 519ur, and 553ur (Slices 001ll, 519ur, 553ur are the fragments of Slices 519 and 553, not overlapping with the original TD).

Table 4—Working CNN configurations.

Performance Metrics. We evaluated the CNN performance by calculating standard metrics by means of a confusion matrix (Powers 2011). A confusion matrix is a performance measurement giving combinations of predicted and actual values. The confusion-matrix elements enable calculating derivative parameters such as accuracy, recall, precision, and F_1 score. Accuracy is the ratio of correctly predicted examples to the total examples (Eq. 3, where TP is true positive, TN is true negative, FP is false positive, and FN is false negative). The recall is the ratio of correct positive predictions to the total positive examples (Eq. 4), while precision is the ratio of correct positive predictions to the total predicted positives (Eq. 5). The F_1 score is a measure to compare both recall and precision (Eq. 6).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (4)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (5)$$

$$F_1 \text{ score} = 2\frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}. \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (6)$$

Results

At the first step, we tested Configurations 1 through 4 (Table 4). To avoid graphics-processing-unit overload and memory shortage, we split the TD slices into 448×448-px tiles in the cross-entropy (CE) configuration. As a result, we have low segmentation quality because of patch stitching back after segmentation. Afterward, we studied the results for focal loss function $\gamma = 1$ (FL_y1), where the focal loss suppresses class imbalance, and γ smoothly adjusts the rate at which easy examples are downweighted. Setting $\gamma > 0$ reduces the relative loss for well-classified examples (probability > 0.5), focusing on challenging, misclassified cases. We used the $\gamma = 1$ value along with the AMSGrad optimizer. For the adaptive moment estimation (Adam) configuration, we replaced AMSGrad with the Adam optimizer and focal loss with $\gamma = 2$. This model showed consistent results, and we assumed that $\gamma = 2$ is the best setting for focal loss.

To test the error-rate convergence while selecting the optimizer, we switched to stochastic gradient descent (SGD). Although the training stage took as twice as much time as Adam did (5 hours of training for 36,000 epochs and 10 hours of training for the same number of epochs), the SGD provided reliable results. However, our visual inspection revealed mistakes in segmentation.

Having compared the FL_y1, Adam, and SGD configurations, we overlooked the significant advantages of SGD while the computation time doubled. We then decided to stop changing both optimizer and loss functions and arrived at AMSGrad and focal loss with $\gamma = 2$.

We analyzed the segmentation performance of Configurations 1 through 4 and revealed several slices (035, 101, 116, 339) with FN values higher than those for others **(Fig. 7).** The low FN values mean an underestimation of the pore space. FP values fall in the range of 0–3%, suggesting a minimal misclassification of the rock matrix. We consider that the rock matrix's correct classification is more important than an underestimation of pore space. Thus, the configurations show a positive trend. For evaluation, we used TD and VD.
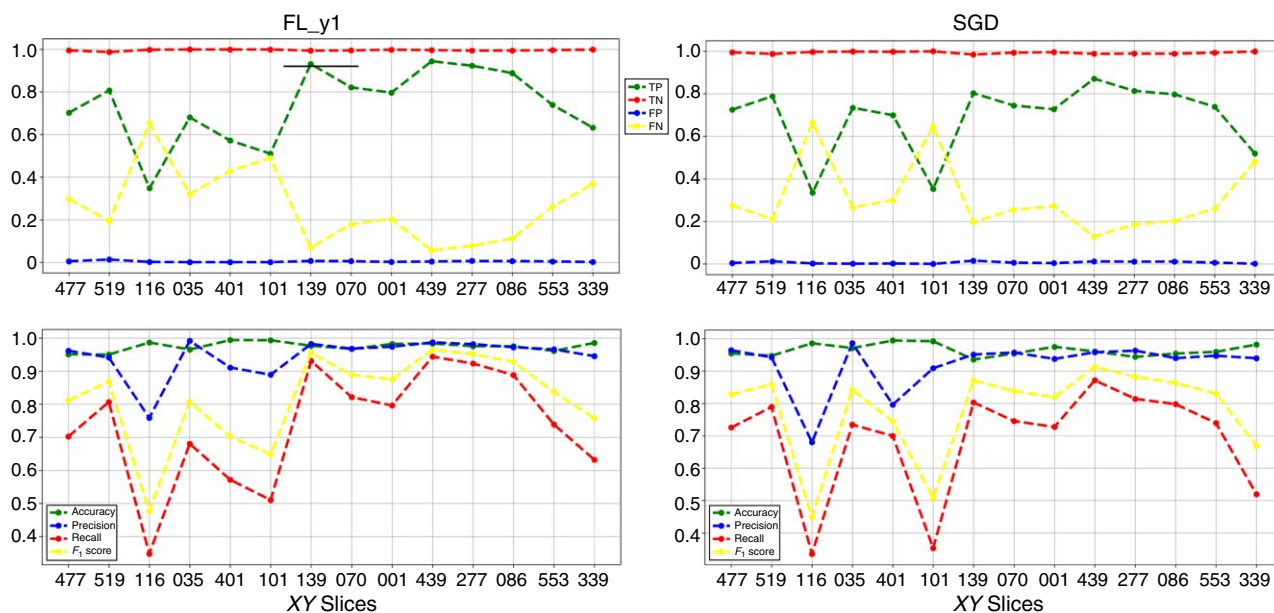

Fig. 7—Performance analysis of Configurations 0Fl_y1 and SGD. The top plots show the confusion-matrix components, while the bottom plots present the derivative parameters.

An analysis of the segmentation performance of Configurations 1 through 4 motivated the study of TD and VD composition. We used K-folding to create two more data sets identical to the host ones but with different TD and VD patches. The technique is a common approach to check model stability and its independence of choice for slices within a single TD. For this purpose, we created Configurations K1, K2, K3 with VD, including Slices 001, 086, 277, and 553; Slices 035, 439, 477, and 519; and Slices 101, 439, 519, 553, correspondingly. Eventually, Configuration K1 showed the best results, while Configurations K2 and K3 showed unstable performance **(Fig. 8).** Thus, we decided to expand TD rather than to exchange TD and VD.
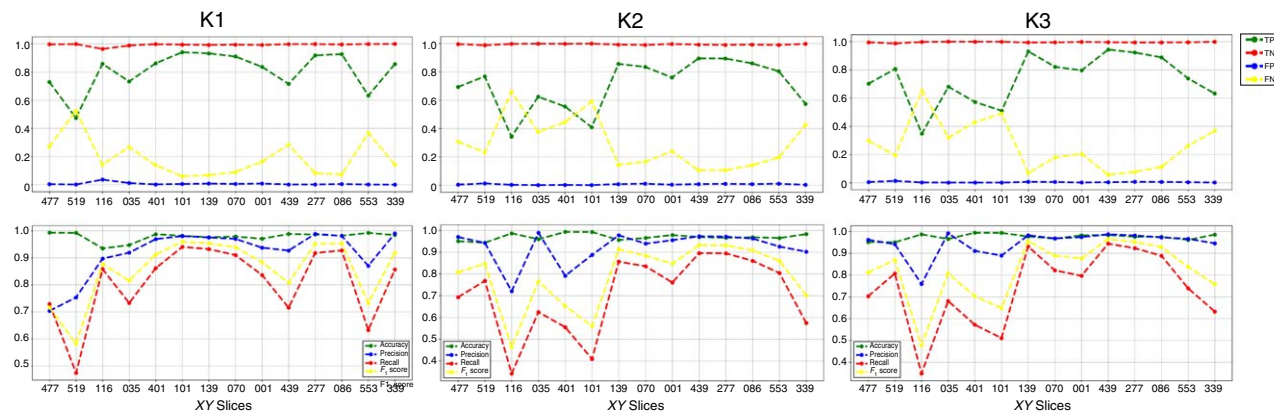

Fig. 8—Performance analysis of Configurations K1, K2, and K3. The top plots show the confusion-matrix components, while the bottom plots present the derivative parameters.

All the studied configurations show similar trends (Fig. 8). We noticed that a few slices had difficult-to-segment features, leading to a segmentation quality lower than that for most TD and VD slices. Slices 035, 101, 116, 339, and 401 became the most problematic, while segmentation of all the rest yielded lower error rates with all configurations. Moreover, the slices highly correlate with each other. Thus, we have tested an idea to expand TD with extra labeled slices and areas (patches). The TD expansion included five items: Slices 402 and 443 and Patches 001ll, 519ur, 553ur, where "ll" stands for lower left and "ur" stands for upper right. The additional data items reside close to the problematic slices. Because Configuration K1 performed the best, we used it as the basis for the Retrained configuration. The Retrained configuration originated from Configuration K1 by additional training on the expanded TD for 12,000 epochs. However, this effort has not delivered further growth in segmentation quality. **Table 5** summarizes the results of our study.

Configuration Name	Epochs	Training Time	Learning Rate	Segmentation Time (seconds)	Accuracy	Precision	Recall	F_1 Score
CE	36,000	05:30:00	0.0010	45–50	0.80	0.52	0.85	0.53
Fl_y1	36,000	05:30:00	0.0010	45–50	0.79	0.53	0.77	0.51
Adam	36,000	05:30:00	0.0010	45–50	0.83	0.54	0.79	0.53
SGD	58,000	10:40:00	0.0010	45–50	0.84	0.46	0.57	0.38
K1	**36,000**	**06:00:00**	**0.0001**	**45–50**	**0.86**	**0.54**	**0.83**	**0.55**
K2	36,000	06:00:00	0.0001	45–50	0.81	0.49	0.85	0.52
K3	36,000	06:00:00	0.0001	45–50	0.84	0.31	0.33	0.23
Retrained	48,000	09:00:00	0.0001	45–50	0.77	0.23	0.24	0.24

Table 5—Summary of average segmentation-quality metrics for all studied configurations. The bolded entries highlight the best performer configuration.

Fig. 9 illustrates the segmentation performance for all studied configurations. For evaluation, we used the TSD. Interestingly, although the Adam configuration delivered the best metrics (Table 5), a visual inspection of the segmentation results clearly shows that Configuration K1 performs much better (Fig. 9). We did not expect this observation, and thus it is a promising direction for further research.
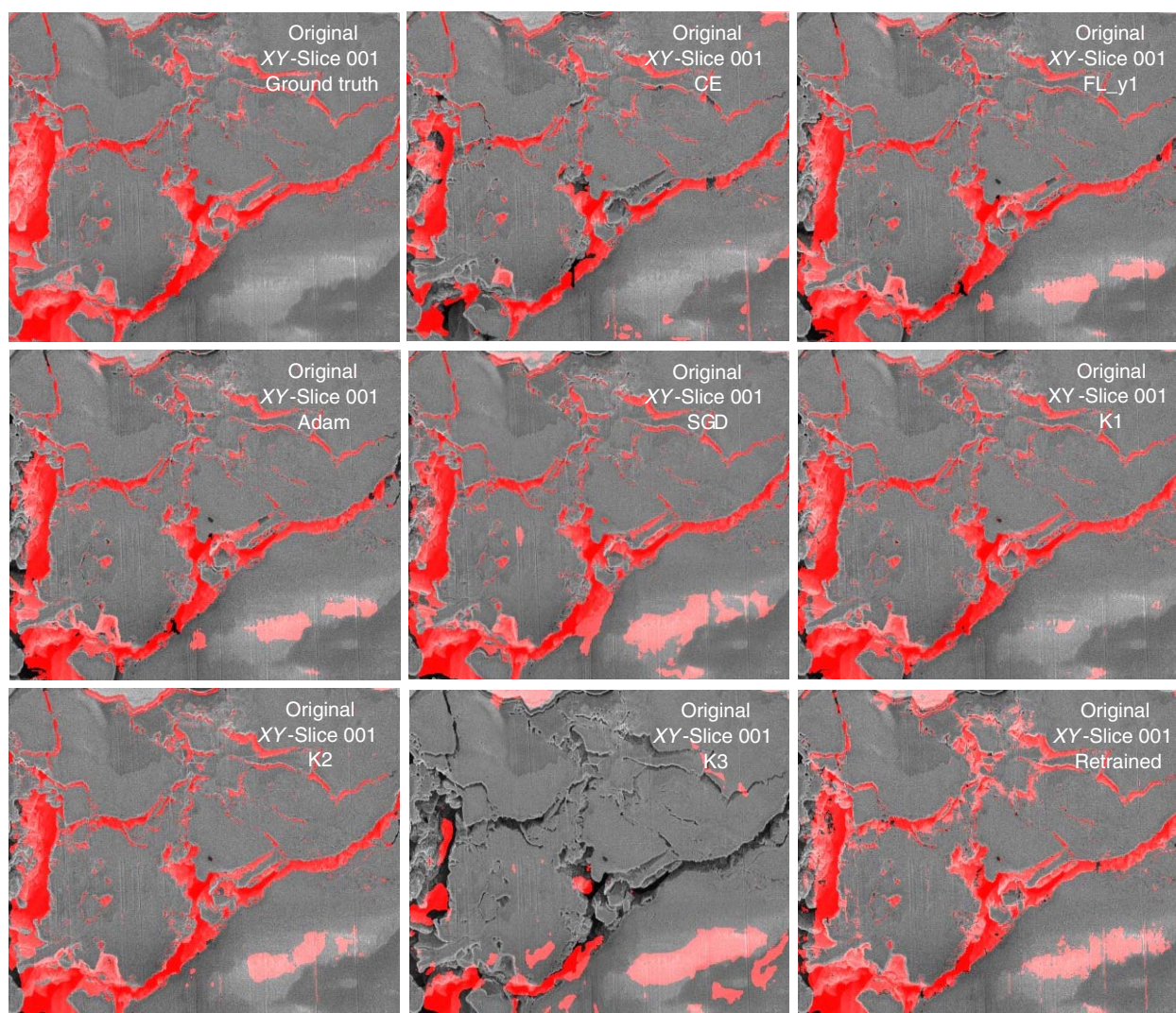

Fig. 9—Visual comparison of segmentation results by all studied configurations on the TSD, *XY*-Slice 001 from Data Set 7.1.1.
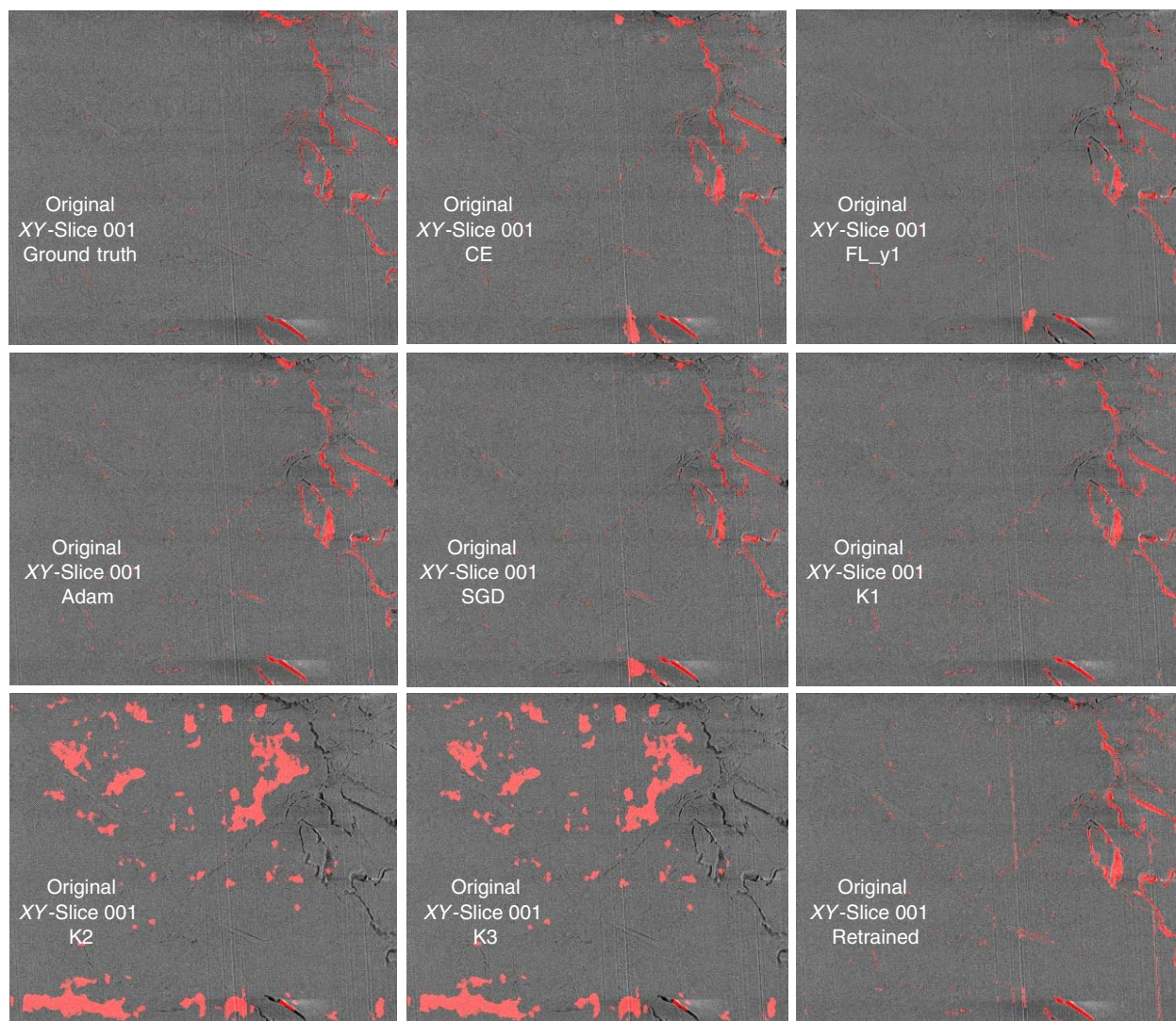
We also analyzed the performance on the TSD, which was not a part of TD and is entirely new to all configurations. We observed fair performance results for nearly all studied configurations. However, both Configurations K2 and K3 showed massive misinterpretation, while the Retrained configuration misinterpreted the image-warping artifact as a pore **(Fig. 10).**

Fig. 10—Visual comparison of segmentation results by all the studied configurations on the TSD, *XY*-Slice 001 from Data Set 7.1.2.

To provide the most stressful performance test, we ran all configurations on an entirely new slice from Data Set 3.1.1. **Fig. 11** compares the application results of the so-called transfer-learning technique. The idea is to overcome the isolated learning paradigm and use the knowledge acquired for one task to solve related ones (Weiss et al. 2016). All configurations failed to accomplish the exercise. From this experience, we conclude that different FIB-SEM data sets (image stacks) obtained with other parameters and characterizing different rocks differ dramatically, disabling a straightforward application of transfer learning.

Integrated evaluation of performance metrics and visual-inspection results shows that Configuration K1 delivers the best segmentation results.

Fig. 12 compares the derivative performance indicators for Configurations K1, K2, and K3 for arbitrary slices from all acquired data sets and highlights the sensitivity to selected slices. For all three configurations, the accuracy gradually decreases from 100% for the TD and VD slices to nearly 80% for TSD slices previously unseen by the CNN. Simultaneously, the results show a class imbalance toward a rock matrix of 70 to 90%.

Configuration K1 shows both precision and F_1 score performance indicators greater than 80% for the slices from TD, VD, and Data Sets 7.1.1 and 7.1.2, part of TSD. We consider these results to be fairly successful. Configuration K2 demonstrated all the performance indicators greater than 60% for the TD, VD, and TSD slices. At the same time, Configuration K3 shows fair results only for TD and VD classification with a strong recommendation for further results validation by visual inspection. The comparison indicates that the CNN configurations strongly depend on the choice of TD for the correct segmentation of FIB-SEM images, even for the same sets of tuning parameters.

Discussion

Choosing the Optimal Configuration. Among all tested configurations, we chose K1 because it has the best performance according to visual inspection. Configuration K1 correctly segmented the maximum number of regions, including complicated ones, such as the bottom-right corner of Fig. 9. All other configurations segmented this region erroneously as the pore space. With TD of only 10 slices,

the configuration successfully segmented the slices of both Data Sets 7.1.1 and 7.1.2, part of TSD, and showed a dramatic time reduction. We timed the manual segmentation of a single slice from Data Set 7.1.1 as 10 to 80 minutes. The CNN segmentation took only 45 to 60 seconds. Our estimate of the time required for training and segmentation of Data Set 7.1.1 ranges within 13 to 14 hours, while the manual segmentation can take 92 to 700 hours. We consider this to be an excellent achievement.
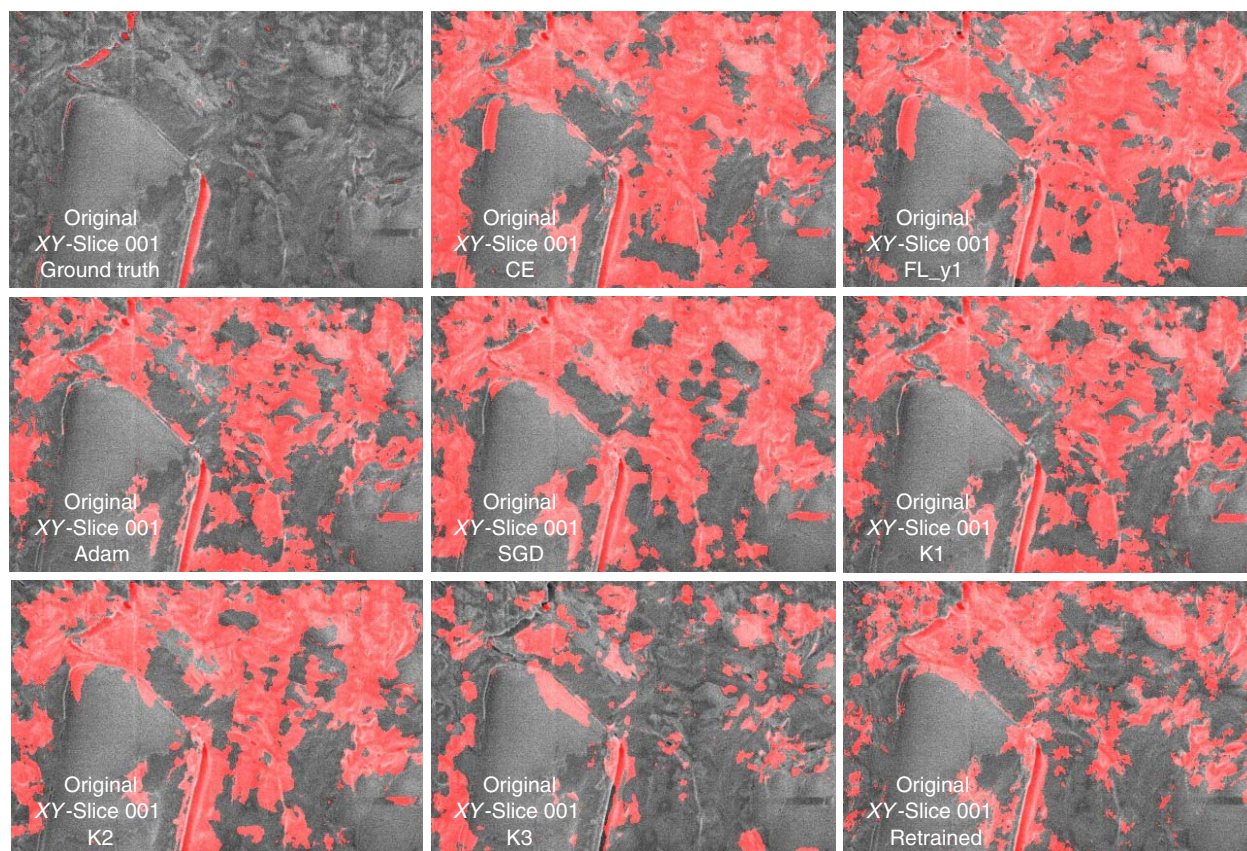

Fig. 11—Visual comparison of segmentation results by all studied configurations on TSD, *XY*-Slice 001 from Data Set 3.1.1.
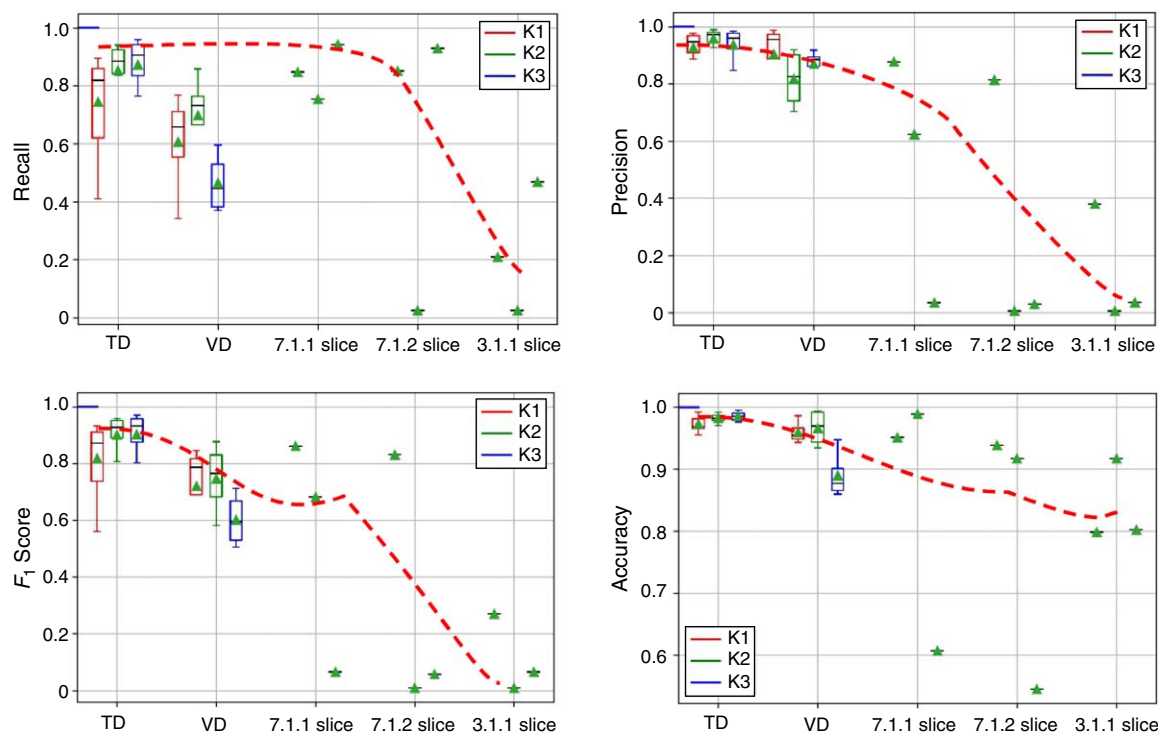
Fig. 12—Diagnostic plots compare segmentation-quality metrics for the best CNN configurations for different VD selections, including TD, VD, and TSD. The red dashed line shows the averaged results.

However, the segmentation of slices from Data Set 3.1.1, a part of TSD, showed that even the best-performing configuration currently does not provide satisfactory results. The observation suggests that the transfer-learning technique is likely not applicable in this case. Expansion of existing TD by adding slices from the unseen data set might potentially mitigate the issue. However, unlike other application areas where CNN training involves hundreds to millions of images (Girshick et al. 2016; Krizhevsky et al. 2017), the geoscience domain in general, and our research in particular, deal with a limited amount of labeled rock-image data. Moreover, publicly available FIB-SEM data are even less. For example, as of the writing date of this paper, the Digital Rocks Portal (Prodanovic et al. 2015) held only one FIB-SEM data set in total.

There are several research tools for automatic model selection and CNN's hyper-parameter optimization/tunning. One of the best-known approaches is automated ML (Wang and Wu 2019). Another solution is automatic WEKA (Kotthoff et al. 2017, 2019). Besides, major cloud-service providers, including Amazon® (Amazon Technologies, Inc., Seattle, Washington, USA), Microsoft® (Microsoft Corporation, Redmond, Washington, USA), and Google® (Google, LLC, Mountain View, California, USA), offer similar services. During the survey, we considered using automated ML but discovered several critical constraints. First, the current automated-ML products do not support the PyTorch framework and limit research opportunities. Second, although it is relatively easy to train models with different optimizers, the data-sampling-hyperparameter evaluation is getting trickier. Third, our experience shows that selecting uncorrelated data is more important than choosing an optimizer and its parameters.

Segmentation Features. Another critical feature would be a systematic void-space underestimation, even if CNN performed adequately and successfully suppressed the FIB-SEM artifacts. However, during image segmentation, we prefer to underestimate the pore-space volume. An overestimation of porosity leads to an optimistic evaluation of both the storage (porosity) and transport (permeability) properties of the target reservoir rock.

K-folding manipulation showed that the proper choice of TD is crucial and might lead to poor performance. We consider the selection of TD contents by using algorithms for image-statistics evaluation to provide the most representative slices from the whole target data set.

Although we did not greatly vary the image dimensions, the CE configuration demonstrated that the CNN does not capture enough features to segment the image correctly with a narrow field of view.

In this study, we focused on testing the DeepUNet architecture without considering others. However, to date, there are many promising CNN candidates (Sultana et al. 2020) for a feasibility study in the scope of FIB-SEM image segmentation. We consider selecting the optimal architecture and configurations as an important direction for further research on the topic.

Segmentation-Quality Evaluation. We calculated the error rates for all studied configurations (except for the Retrained configuration) to evaluate the segmentation quality in terms of surface porosity. The Retrained configuration showed the worst performance on TSD demonstrated in Table 5, and we decided not to use it for segmentation-quality evaluation. We calculated the so-called surface porosity (areal fraction held by segmented pores) by the output of each configuration and compared those with ground-truth-porosity values for each VD slice. **Fig. 13** (left) shows the range and mean porosity values. The behaviors of Configurations SGD and K3 are noticeably different from the others. Namely, the highest ground-truth-porosity value is 24.3%, while Configurations K3 and SGD showed 15.7% and 14.2%, respectively. To rank the configurations by segmentation quality, we analyzed the error rate vs. ground-truth-porosity values. Fig. 13 (right) shows that Configurations K1 and K2 are the best two performers, with average error rates of 10.9 and 11.8%, respectively.
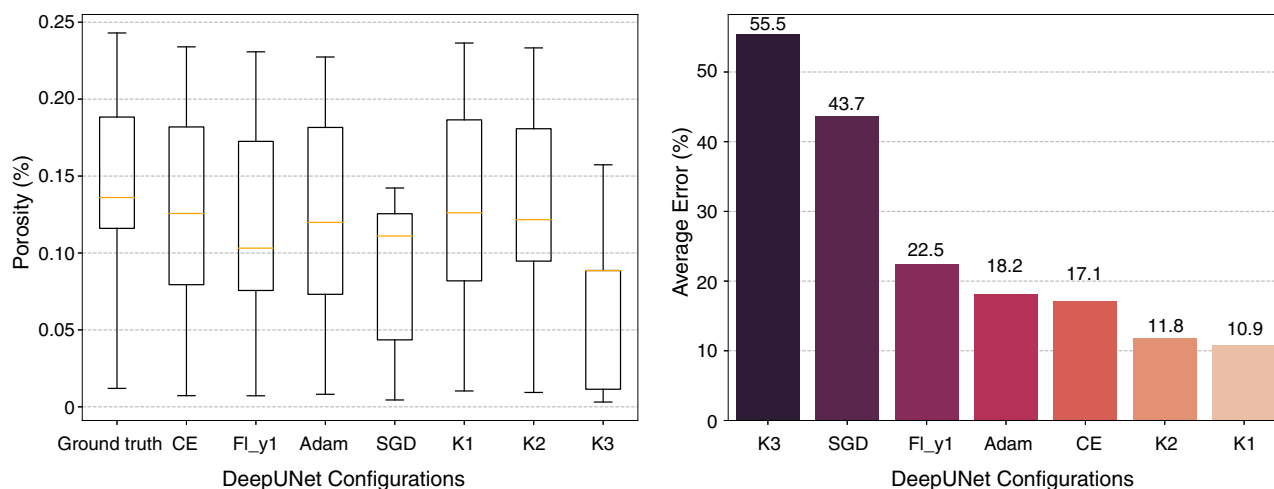

Fig. 13—(left) Porosity-evaluation and (right) error-rate analysis by configuration vs. ground-truth-porosity values.

Validation of Segmentation Results. The creation of digital 3D representations of pore microstructures for image-based rock-physics simulations remains a highly subjective enterprise. The decisions on identifying both macropores and micropores and various mineral components remain dependent on specialist choices and biases (Howard et al. 2019). One of the general issues in introducing segmentation approaches is that validation of the segmentation results requires the ground truth to be available (Berg et al. 2018). The latter, in the case of microscale or nanoscale rock samples, is rarely known.

Historically, digital rock analysts validated the segmentation by comparison of resulting properties such as porosity and permeability. However, following this approach can introduce a significant mismatch in porosity because imaging methods have different volumes of investigation and resolution limits (Leu et al. 2014; Saxena et al. 2019a, 2019b). Howard et al. (2019) introduced an uncertainty-quantification framework during image segmentation for image-based DRP to handle the issue with ground-truth

availability. Uncertainty quantification relies on two core components. First is the iterative study of the contribution of resolvable macropores, and their spatial distributions regarding pore-lining clay minerals with unresolvable microporosity. Second is cross validation of segmentation results by independent nuclear magnetic resonance and mercury-injection capillary-pressure measurements. Andrew (2018) approached the problem by creating synthetic volumes and comparing various segmentation techniques. However, the research results showed that synthetic data sets do not capture real data features, such as subresolution porosity. Thus, the question about the ground truth is still open and needs to be further researched.

Moreover, the porosity seems to be the easiest property to validate. The digital rock images resulting from SEM, FIB-SEM, nano-CT, and micro-CT allow purely visual description by a professional lithologist/petrographer. The experienced specialist can deliver manual labeling, which we believe is the closest to the ground truth. However, validation of absolute permeability would require a real microfluidic-flow cell holding a rock sample with physical dimensions of tens of microns in one direction (typical size of FIB-SEM volume). The latter is still not available because modern microfluidic-flow cells take rock samples larger than 500 µm (Singh et al. 2017). Thus, to date, permeability validation by means of standard flood tests seems to be impossible.

In the case of the target rock samples, we implicitly validated the FIB-SEM segmentation results by comparing the porosity determined at the smallest scale (FIB-SEM) with the effective porosity of the rock matrix calculated from plug-gas porosity and micro-CT porosity. The calculation of FIB-SEM porosity takes 3D segmentation of the void space to the input and is pretty straightforward (Kelly et al. 2016). We then calculated the effective porosity of the rock matrix by taking several reasonable assumptions. First, the gas porosity determined on the core plugs is true. Second, we consider that the digital rock model dependent on micro-CT data represents the core plug. Third, the micro-CT-data segmentation provides underestimated pore volume because of limited resolution (Kazak et al. 2017a). Fourth, we consider that the closed (isolated) porosity is negligible. In this case, the resulting core-plug porosity derived from micro-CT data is less than that from the corresponding core-plug-gas porosimetry test. Thus, we assume that the matrix in the micro-CT digital rock model has implicit (unresolved) open porosity. Taking this for granted, the following relationship describes the interplay between resolved and matrix porosity,

$$\phi_t = \phi_r + (1 - \phi_r) \cdot \phi_{ur}, \quad\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \quad (7)$$

where ϕ_t is the total open porosity equal to gas porosity from core plugs; ϕ_r is the (void) porosity resolved in the micro-CT digital rock model; and ϕ_{ur} is the (matrix) porosity unresolved in the micro-CT digital rock model.

One can express ϕ_{ur} from Eq. 7 as

$$\phi_{ur} = \frac{\phi_t - \phi_r}{1 - \phi_r}. \quad\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \quad (8)$$

A graphical representation of Eq. 7 shows the possible range of the matrix porosity calculated using Eq. 8 as 8.8 to 17.6% **(Fig. 14).** Indeed, the porosity calculated from the acquired FIB-SEM data sets' segmentation ranges within 1 to 10%. Interestingly, both ranges overlap. The latter indicates that there can physically be a pore volume at the nanoscale sufficient to provide the rock matrix's effective porosity. The effective void space, in turn, is enough to provide (according to the accepted porosity model) the total porosity equal to that of the core plug measured by gas. The conclusion suggests that the resulting segmentation results are valid.
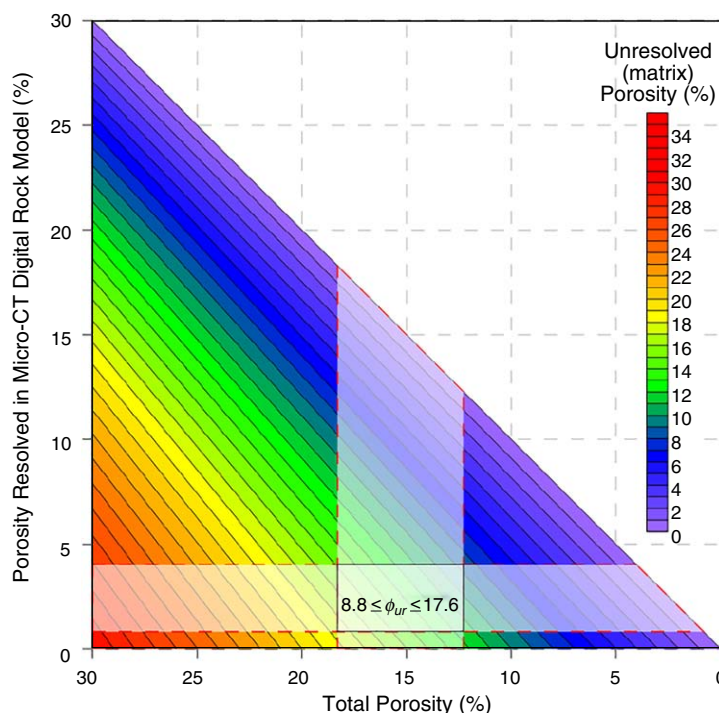

Fig. 14—Contour map representing the relationship between total, resolved, and unresolved (matrix) porosity as per Eq. 7. The vertical white band shows the core-plug-gas-porosity range, while the horizontal one corresponds to the range of open porosity of the digital rock model for the same rock sample. The intersection of the bands provides a domain of unresolved (matrix) porosity.

Conclusions

The correct determination of storage and transport properties from FIB-SEM images requires fast and robust segmentation. However, the quality and efficient segmentation of FIB-SEM images is still a complicated and challenging task. The presence of FIB-SEM artifacts, such as porebacks, requires developing a new methodology for efficient image segmentation.

The conducted literature review demonstrated that the conventional segmentation techniques, including global and watershed thresholding, show poor performance for the segmentation of FIB-SEM images complicated with artifacts. Typically, a trained operator spends days or weeks for subjective and semimanual labeling of one FIB-SEM data set.

We addressed the segmentation-quality problem in the presence of imaging artifacts using DeepUNet CNN. The architecture has a unique design suitable for high-resolution images and features extraction from small local areas on images and training these features' weights to classify objects.

The acquired data sets came from FIB-SEM imaging of a representative collection of tight gas reservoir-rock samples (Berezov Formation, West Siberia, Russia). The studied rock's complex void-space structure complicated the digital rock images with acquisition artifacts, including porebacks, curtaining, bright pore edges, and complicated 3D pores. We used a subset of a limited number of cropped and manually labeled slices to form training data. We then augmented the TD by a set of image transforms, including vertical and horizontal flipping, rotation, blurring, histogram normalization, and gray-level expansion. Eventually, we developed and implemented the configuration, which can automatically segment the acquired data sets.

The application of the DeepUNet to the FIB images delivered promising results. First, based on the TD, the CNN misclassification rate is 2%. Second, CNN pore-space segmentation to two classes showed higher quality than that to a single class. Third, CNN segmentation speed is 14.5 MP/min, which is more than 10 times higher than the average manual segmentation speed, which is 0.18 to 1.45 MP/min.

In summary, we developed a robust approach for an automated, highly efficient, multimodal segmentation of FIB-SEM data sets using ML-based methods. The research results significantly improve the quantitative petrophysical characterization of complex reservoir rocks using digital rock FIB-SEM imaging. Directions for further development include the intelligent choice of training data and exploring different CNN architectures besides DeepUNet.

Acknowledgments

References

Andra, H., Combaret, N., Dvorkin, J. et al. 2013a. Digital Rock Physics Benchmarks—Part I: Imaging and Segmentation. *Comput & Geosci* **50** (January): 25–32. https://doi.org/10.1016/j.cageo.2012.09.005.

Andra, H., Combaret, N., Dvorkin, J. et al. 2013b. Digital Rock Physics Benchmarks—Part II: Computing Effective Properties. *Comput & Geosci* **50** (January): 33–43. https://doi.org/10.1016/j.cageo.2012.09.008.

Andrew, M. 2018. A Quantified Study of Segmentation Techniques on Synthetic Geological XRM and FIB-SEM Images. *Computat Geosci* **22** (6): 1503–1512. https://doi.org/10.1007/s10596-018-9768-y.

Andrew, M., Bhattiprolu, S., Butnaru, D. et al. 2017. The Usage of Modern Data Science in Segmentation and Classification. *Microsc Microanal* **23** (S1): 156–157. https://doi.org/10.1017/s1431927617001465.

API RP 40, Recommended Practices for Core Analysis. 1998. Washington, DC, USA: API.

Arganda-Carreras, I. 2016. Introduction to Image Segmentation Using ImageJ/Fiji. Oral presentation given at Biological Image Reconstruction and Analysis, Berlin, Germany, 17 November.

Arganda-Carreras, I., Kaynig, V., Rueden, C. et al. 2017. Trainable Weka Segmentation: A Machine Learning Tool for Microscopy Pixel Classification. *Bioinformatics* **33** (15): 2424–2426. https://doi.org/10.1093/bioinformatics/btx180.

Badrinarayanan, V., Kendall, A., and Cipolla, R. 2017. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans Pattern Anal Mech Intell* **39** (12): 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615.

Bear, J. 1988. *Dynamics of Fluids in Porous Media.* New York, New York, USA: Environmental Science Series, American Elsevier Publishing Company.

Berg, S., Saxena, N., Shaik, M. et al. 2018. Generation of Ground Truth Images To Validate Micro-CT Image-Processing Pipelines. *Lead Edge* **37** (6): 412–420. https://doi.org/10.1190/tle37060412.1.

Blayvas, I., Bruckstein, A., and Kimmel, R. 2006. Efficient Computation of Adaptive Threshold Surfaces for Image Binarization. *Pattern Recognit* **39** (1): 89–101. https://doi.org/10.1016/j.patcog.2005.08.011.

Breiman, L. 2001. Random Forests. *Machine Learning* **45** (1): 5–32. https://doi.org/10.1023/a:1010933404324.

Bull, G., Gao, J., and Antolovich, M. 2014. Delineation of Rock Fragments by Classification of Image Patches Using Compressed Random Features. Oral presentation given at the International Conference on Computer Vision Theory and Applications (VISAPP), Lisbon, Portugal, 5–8 January.

Bull, G., Gao, J., and Antolovich, M. 2015. Rock Fragment Boundary Detection Using Compressed Random Features. In *Computer Vision, Imaging and Computer Graphics–Theory and Applications*, ed. S. Battiato, S. Coquillart, J. Pettré, et al., 273–286. Cham, Switzerland: Communications in Computer and Information Science Series, Springer International Publishing.

Bultreys, T., De Boever, W., and Cnudde, V. 2016. Imaging and Image-Based Fluid Transport Modeling at the Pore Scale in Geological Materials: A Practical Introduction to the Current State-of-the-Art. *Earth Sci Rev* **155** (April): 93–128. https://doi.org/10.1016/j.earscirev.2016.02.001.

Byrnes, A. P., Zhang, S., Canter, L. et al. 2017. Application of Integrated Core and 3D Image Rock Physics To Characterize Niobrara Chalk Properties Including Relative Permeability with Bound Water Effect. Paper presented at the SPE/AAPG/SEG Unconventional Resources Technology Conference, Austin, Texas, USA, 24–26 July. URTEC-2670963-MS. https://doi.org/10.15530/URTEC-2017-2670963.

Caffrey, B. J., Maltsev, A. V., Gonzalez-Freire, M. et al. 2019. Semi-Automated 3D Segmentation of Human Skeletal Muscle Using Focused Ion Beam-Scanning Electron Microscopic Images. *J Struct Biol* **207** (1): 1–11. https://doi.org/10.1016/j.jsb.2019.03.008.

Chauhan, S., Rühaak, W., Anbergen, H. et al. 2016a. Phase Segmentation of X-Ray Computer Tomography Rock Images Using Machine Learning Techniques: An Accuracy and Performance Study. *Solid Earth* **7** (4): 1125–1139. https://doi.org/10.5194/se-7-1125-2016.

Chauhan, S., Rühaak, W., Khan, F. et al. 2016b. Processing of Rock Core Microtomography Images: Using Seven Different Machine Learning Algorithms. *Comput Geosci* **86** (January): 120–128. https://doi.org/10.1016/j.cageo.2015.10.013.

Chauhan, S., Sell, K., Rühaak, W. et al. 2020. CobWeb 1.0: Machine Learning Toolbox for Tomographic Imaging. *Geosci. Model Dev*. **13** (1): 315–334. https://doi.org/10.5194/gmd-13-315-2020.

Converse, M. I. and Fullwood, D. T. 2013. Enhancing Nanoscale SEM Image Segmentation and Reconstruction with Crystallographic Orientation Data and Machine Learning. *Mater Charact* **83** (September): 109–122. https://doi.org/10.1016/j.matchar.2013.06.011.

Dunham, R. J. 1962. Classification of Carbonate Rocks According to Depositional Texture. In *Classification of Carbonate Rocks*, ed. W. E. Ham, AAPG Memoir 1, 108–121.

Durand, M., Nikitin, A., McMullen, A. et al. 2019. Crushed Rock Analysis Workflow Based on Advanced Fluid Characterization for Improved Interpretation of Core Data. Paper presented at the SPWLA 60th Annual Logging Symposium, The Woodlands, Texas, USA, 15–19 June. SPWLA-2019-AAAA. https://doi.org/10.30632/T60ALS-2019_AAAA.

Dvorkin, J., Derzhi, N., Diaz, E. et al. 2011. Relevance of Computational Rock Physics. *Geophysics* **76** (5): E141–E153. https://doi.org/10.1190/geo2010-0352.1.

Embry, A. F. and Klovan, J. E. 1971. A Late Devonian Reef Tract on Northeastern Banks Island, N.W.T. *Bull Can Petr Geol*. **19** (4): 730–781.

Fager, C., Röding, M., Olsson, A. et al. 2020. Optimization of FIB-SEM Tomography and Reconstruction for Soft, Porous, and Poorly Conducting Materials. *Microsc Microanal* **26** (4): 837–845. https://doi.org/10.1017/S1431927620001592.

Frucci, M., Sanniti, G., and Sanniti di Baja, G. 2008. From Segmentation to Binarization of Gray-Level Images. *JPRR* **3** (1): 1–13. https://doi.org/10.13176/11.54.

Giannuzzi, L. and Stevie, F. ed. 2005. *Introduction to Focused Ion Beams: Instrumentation, Theory, Techniques and Practice*. New York, New York, USA: Springer Science+Business Media.

Girshick, R., Donahue, J., Darrell, T. et al. 2016. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Trans Pattern Anal Mach Intell* **38** (1): 142–58. https://doi.org/10.1109/TPAMI.2015.2437384.

Goral, J., Andrew, M., Olson, T. et al. 2020. Correlative Core- to Pore-Scale Imaging of Shales. *Mar Pet Geol* **111** (January): 886–904. https://doi.org/10.1016/j.marpetgeo.2019.08.009.

Goral, J., Deo, M., and Andrew, M. 2018. Pore Network Modeling of Marcellus Shale Using Digital Rock Analysis with Machine Learning Image Segmentation. *ZEISS Solutions for Shale Characterization, ZEISS Application Note*, Carl Zeiss Microscopy GmbH, Jena, Germany.

Goral, J., Miskovic, I., Gelb, J. et al. 2016. Correlative X-Ray and Electron Microscopy for Multi-Scale Characterization of Heterogeneous Shale Reservoir Pore Systems. In *Imaging Unconventional Reservoir Pore Systems*, ed. T. Olson, 77–88. Tulsa, Oklahoma, USA: American Association of Petroleum Geologists.

Goral, J., Walton, I., Andrew, M. et al. 2019. Pore System Characterization of Organic-Rich Shales Using Nanoscale-Resolution 3D Imaging. *Fuel* **258** (15 December): 116049. https://doi.org/10.1016/j.fuel.2019.116049.

Gurney, K. 1997. *An Introduction to Neural Networks*, first edition. Boca Raton, Florida, USA: CRC Press.

Hastie, T., Tibshirani, R., and Friedman, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, New York, USA: Springer Series in Statistics, Springer Science+Business Media.

He, K., Zhang, X., Ren, S. et al. 2016. Deep Residual Learning for Image Recognition. Oral presentation given at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, Nevada, USA, 27–30 June.

Holzer, L., Indutnyi, F., Gasser, P. et al. 2004. Three-Dimensional Analysis of Porous BaTiO$_3$ Ceramics Using FIB Nanotomography. *J. Microsc.* **216** (1): 84–95. https://doi.org/10.1111/j.0022-2720.2004.01397.x.

Howard, J., Lin, S., and Zhang, S. 2019. Uncertainty Quantification in Image Segmentation for Image-Based Rock Physics in a Shaly Sandstone. *Petrophysics* **60** (2): 240–254. SPWLA-2019-v60n2a2. https://doi.org/10.30632/PJV60N2-2019a2.

Iwaszenko, S. and Nurzynska, K. 2019. Rock Grains Segmentation Using Curvilinear Structures Based Features. Oral presentation given at SPIE Defense + Commercial Sensing 2019, Baltimore, Maryland, USA, 14–18 April.

Jobe, T. D., Vital-Brazil, E., and Khait, M. 2018. Geological Feature Prediction Using Image-Based Machine Learning. *Petrophysics* **59** (6): 750–760. SPWLA-2018-v59n6a1. https://doi.org/10.30632/PJV59N6-2018a1.

Joos, J., Carraro, T., Weber, A. et al. 2011. Reconstruction of Porous Electrodes by FIB/SEM for Detailed Microstructure Modeling. *J Power Sources* **196** (17): 7302–7307. https://doi.org/10.1016/j.jpowsour.2010.10.006.

Jørgensen, P. S., Hansen, K. V., Larsen, R. et al. 2010. A Framework for Automatic Segmentation in Three Dimensions of Microstructural Tomography Data. *Ultramicroscopy* **110** (3): 216–228. https://doi.org/10.1016/j.ultramic.2009.11.013.

Karimpouli, S. and Tahmasebi, P. 2019. Segmentation of Digital Rock Images Using Deep Convolutional Autoencoder Networks. *Comput & Geosci* **126** (May): 142–150. https://doi.org/10.1016/j.cageo.2019.02.003.

Kazak, A., Chugunov, S., and Chashkov, A. 2017a. Quantitative Integration of Modern Automated Mineralogy, and Large-Area SEM Imaging Techniques To Select Representative Regions of Interest for Microstructural Reservoir Rock Analysis at Micro- and Nano-Scale. Oral presentation given at the 17th International Scientific Conference on Earth and Geosciences, Albena, Bulgaria, 29 June–5 July.

Kazak, A., Chugunov, S., and Chashkov, A. 2018. Integration of Large-Area SEM Imaging and Automated Mineralogy-Petrography Data for Selection of Nano-Scale Pore-Space Characterization Sites. *SPE Res Eval & Eng* **21** (4): 821–836. SPE-191369-PA. https://doi.org/10.2118/191369-PA.

Kazak, A., Chugunov, S., Nachev, V. et al. 2017b. Integration of Large-Area SEM Imaging and Automated Mineralogy-Petrography Data for Justified Decision on Nano-Scale Pore-Space Characterization Sites, as a Part of Multiscale Digital Rock Modeling Workflow. Paper presented at the SPE/AAPG/SEG Unconventional Resources Technology Conference, Austin, Texas, USA, 24–26 July 2017. URTEC-2697437-MS. https://doi.org/10.15530/URTEC-2017-2697437.

Kazak, E. S. and Kazak, A. V. 2019. A Novel Laboratory Method for Reliable Water Content Determination of Shale Reservoir Rocks. *J Pet Sci Eng* **183** (December): 106301. https://doi.org/10.1016/j.petrol.2019.106301.

Kelly, S., El-Sobky, H., Torres-Verdín, C. et al. 2016. Assessing the Utility of FIB-SEM Images for Shale Digital Rock Physics. *Adv Water Resour* **95** (September): 302–316. https://doi.org/10.1016/j.advwatres.2015.06.010.

Khadangi, A., Hanssen, E., and Rajagopal, V. 2019. Automated Segmentation of Cardiomyocyte Z-Disks from High-Throughput Scanning Electron Microscopy Data. *BMC Med Inform Decis Mak* **19** (S6): 272. https://doi.org/10.1186/s12911-019-0962-1.

Khan, F., Enzmann, F., and Kersten, M. 2016. Multi-Phase Classification by a Least-Squares Support Vector Machine Approach in Tomography Images of Geological Samples. *Solid Earth* **7** (2): 481–492. https://doi.org/10.5194/se-7-481-2016.

Kingma, D. P. and Ba, J. 2015. Adam: A Method for Stochastic Optimization. http://arxiv.org/abs/1412.6980.

Koebernick, N., Daly, K. R., Keyes, S. D. et al. 2017. High-Resolution Synchrotron Imaging Shows that Root Hairs Influence Rhizosphere Soil Structure Formation. *New Phytol* **216** (1): 124–135. https://doi.org/10.1111/nph.14705.

Kotthoff, L., Thornton, C., Hoos, H. H. et al. 2017. Auto-WEKA: Automatic Model Selection and Hyperparameter Optimization in WEKA. *J Mach Learn Res* **18** (25): 1–5.

Kotthoff, L., Thornton, C., Hoos, H. H. et al. 2019. Auto-WEKA: Automatic Model Selection and Hyperparameter Optimization in WEKA. In *Automated Machine Learning: Methods, Systems, Challenges*, ed. F. Hutter, L. Kotthoff, and J. Vanschoren, 81–95. Cham, Switzerland: Springer Series on Challenges in Machine Learning Series, Springer International Publishing.

Kreshuk, A., Straehle, C. N., Sommer, C. et al. 2011. Automated Detection and Segmentation of Synaptic Contacts in Nearly Isotropic Serial Electron Microscopy Images. *PLoS One* **6** (10): e24899. https://doi.org/10.1371/journal.pone.0024899.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. 2017. ImageNet Classification with Deep Convolutional Neural Networks. *Commun ACM* **60** (6): 84–90. https://doi.org/10.1145/3065386.

Kuila, U. 2013. *Measurement and Interpretation of Porosity and Pore-Size Distribution in Mudrocks: The Hole Story of Shales*. PhD dissertation, Colorado School of Mines, Golden, Colorado, USA.

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 Med* **105** (2): 451–469. https://doi.org/10.1007/s11242-014-0378-4.

Li, R., Liu, W., Yang, L. et al. 2018. DeepUNet: A Deep Fully Convolutional Network for Pixel-Level Sea-Land Segmentation. *IEEE J Sel Top Appl Earth Obs Remote Sens* **11** (11): 3954–3962. https://doi.org/10.1109/JSTARS.2018.2833382.

Lin, T. Y., Goyal, P., Girshick, R. et al. 2020. Focal Loss for Dense Object Detection. *IEEE Trans Pattern Anal Mach Intell* **42** (2): 318–327. https://doi.org/10.1109/TPAMI.2018.2858826.

Liu, X., Sun, J., and Wang, H. 2009. Numerical Simulation of Rock Electrical Properties Based on Digital Cores. *Appl. Geophys.* **6** (1): 1–7. https://doi.org/10.1007/s11770-009-0001-6.

Lormand, C., Zellmer, G. F., Nemeth, K. et al. 2018. Weka Trainable Segmentation Plugin in ImageJ: A Semi-Automatic Tool Applied to Crystal Size Distributions of Microlites in Volcanic Rocks. *Microsc Microanal* **24** (6): 667–675. https://doi.org/10.1017/S1431927618015428.

Luffel, D. L. and Guidry, F. K. 1992. New Core Analysis Methods for Measuring Reservoir Rock Properties of Devonian Shale. *J Pet Technol* **44** (11): 1184–1190. SPE-20571-PA. https://doi.org/10.2118/20571-PA.

Maitin-Shepard, J., Jain, V., Januszewski, M. et al. 2016. Combinatorial Energy Learning for Image Segmentation. Oral presentation given at the 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, 5–10 December.

Moroni, R. and Thiele, S. 2020. FIB/SEM Tomography Segmentation by Optical Flow Estimation. *Ultramicroscopy* **219** (December): 113090. https://doi.org/10.1016/j.ultramic.2020.113090.

Murphy, K. 2012. *Machine Learning: A Probabilistic Perspective*. Cambridge, Massachusetts, USA: Adaptive Computation and Machine Learning Series, The MIT Press.

Nelson, P. H. 2009. Pore-Throat Sizes in Sandstones, Tight Sandstones, and Shales: Reply. *AAPG Bull.* **93** (3): 329–340. https://doi.org/10.1306/12141010159.

Nikolaev, M. Y. and Kazak, A. V. 2019. Liquid Saturation Evaluation in Organic-Rich Unconventional Reservoirs. *Earth Sci Rev* **194** (July): 327–349. https://doi.org/10.1016/j.earscirev.2019.05.012.

Nurzynska, K. and Iwaszenko, S. 2020. Application of Texture Features and Machine Learning Methods to Grain Segmentation in Rock Material Images. *Image Anal. Stereol.* **39** (2). https://doi.org/10.5566/ias.2186.

Pascanu, R., Mikolov, T., and Bengio, Y. 2012. Understanding the Exploding Gradient Problem. https://arxiv.org/abs/1211.5063.

Pascanu, R., Mikolov, T., and Bengio, Y. 2013. On the Difficulty of Training Recurrent Neural Networks. https://arxiv.org/abs/1211.5063.

Pedregosa, F., Varoquaux, G., Gramfort, A. et al. 2011. Scikit-Learn: Machine Learning in Python. *J Mach Learn Res* (12): 2825–2830.

Piche, N., Bouchard, I., and Marsh, M. 2017. Dragonfly Segmentation Trainer–A General and User-Friendly Machine Learning Image Segmentation Solution. *Microsc Microanal* **23** (S1): 132–133. https://doi.org/10.1017/s1431927617001349.

Powers, D. M. W. 2011. Evaluation: From Precision, Recall and F-Measure to ROC. *Int J Mach Learn Comput* **2** (1): 37–63. https://arxiv.org/abs/2010.16061.

Prill, T., Schladitz, K., Jeulin, D. et al. 2013. Morphological Segmentation of FIB-SEM Data of Highly Porous Media. *J. Microsc.* **250** (2): 77–87. https://doi.org/10.1111/jmi.12021.

Prodanovic, M., Esteva, M., Hanlon, M. et al. 2015. Digital Rocks Portal: A Repository for Porous Media Images. https://www.digitalrocksportal.org (accessed 13 December 2020).

Reuteler, J. 2017. FIB Artifacts and Tricks To Overcome Them. Oral presentation given at EUFN 2017, 1st European FIB Network Workshop, Graz, Austria, 4–5 July.

Ringrose, P. S., Martinius, A. W., and Alvestad, J. 2008. Multiscale Geological Reservoir Modelling in Practice. *Geol Soc Spec Publ* **309** (1): 123–134. https://doi.org/10.1144/sp309.9.

Röding, M., Fager, C., Olsson, A. et al. 2020. Three-Dimensional Reconstruction of Porous Polymer Films from FIB-SEM Nanotomography Data Using Random Forests. *J. Microsc.* **281** (1): 76–86. https://doi.org/10.1111/jmi.12950.

Ronneberger, O., Fischer, P., and Brox, T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Proc.,* Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October, ed. M. Navab, J. Hornegger, W. M. Wells et al., 234–241. Cham, Switzerland: Springer International Publishing.

Salzer, M., Prill, T., Spettl, A. et al. 2015. Quantitative Comparison of Segmentation Algorithms for FIB-SEM Images of Porous Media. *J. Microsc.* **257** (1): 23–30. https://doi.org/10.1111/jmi.12182.

Salzer, M., Spettl, A., Stenzel, O. et al. 2012. A Two-Stage Approach to the Segmentation of FIB-SEM Images of Highly Porous Materials. *Mater Charact* **69** (July): 115–126. https://doi.org/10.1016/j.matchar.2012.04.003.

Salzer, M., Thiele, S., Zengerle, R. et al. 2014. On the Importance of FIB-SEM Specific Segmentation Algorithms for Porous Media. *Mater Charact* **95** (September): 36–43. https://doi.org/10.1016/j.matchar.2014.05.014.

Saxena, N., Hows, A., Hofmann, R. et al. 2019a. Rock Properties from Micro-CT Images: Digital Rock Transforms for Resolution, Pore Volume, and Field of View. *Adv Water Resour* **134** (December): 103419. https://doi.org/10.1016/j.advwatres.2019.103419.

Saxena, N., Hows, A., Hofmann, R. et al. 2019b. Estimating Pore Volume of Rocks from Pore-Scale Imaging. *Transp Porous Med* **129** (1): 403–412. https://doi.org/10.1007/s11242-019-01295-x.

Schaffer, M. and Wagner, J. 2008. Block Lift-Out Sample Preparation for 3D Experiments in a Dual Beam Focused Ion Beam Microscope. *Microchim Acta* **161** (3–4): 421–425. https://doi.org/10.1007/s00604-007-0853-5.

Schindelin, J., Rueden, C. T., Hiner, M. C. et al. 2015. The ImageJ Ecosystem: An Open Platform for Biomedical Image Analysis. *Mol Reprod Dev* **82** (7–8): 518–29. https://doi.org/10.1002/mrd.22489.

Schweizer, S. A., Hoeschen, C., Schluter, S. et al. 2018. Rapid Soil Formation After Glacial Retreat Shaped by Spatial Patterns of Organic Matter Accrual in Microaggregates. *Glob Chang Biol* **24** (4): 1637–1650. https://doi.org/10.1111/gcb.14014.

Seni, G. and Elder, J. F. 2010. *Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions*. Williston, Vermont, USA: Synthesis Lectures on Data Mining and Knowledge Discovery Series, Morgan & Claypool Publishers.

Shu, L., Osinski, G. R., McIsaac, K. et al. 2018. An Automatic Methodology for Analyzing Sorting Level of Rock Particles. *Comput Geosci* **120** (November): 97–104. https://doi.org/10.1016/j.cageo.2018.08.001.

Singh, R., Sivaguru, M., Fried, G. A. et al. 2017. Real Rock-Microfluidic Flow Cell: A Test Bed for Real-Time In Situ Analysis of Flow, Transport, and Reaction in a Subsurface Reactive Transport Environment. *J Contam Hydrol* **204** (September): 28–39. https://doi.org/10.1016/j.jconhyd.2017.08.001.

Sommer, C., Straehle, C., Köthe, U. et al. 2011. Ilastik: Interactive Learning and Segmentation Toolkit. Oral presentation given at the 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Chicago, Illinois, USA, 30 March–2 April.

Sultana, F., Sufian, A., and Dutta, P. 2020. Evolution of Image Segmentation Using Deep Convolutional Neural Network: A Survey. *Knowl Based Syst* **201–202** (9 August): 106062. https://doi.org/10.1016/j.knosys.2020.106062.

Sun, H., Yao, J., Cao, Y-C. et al. 2017. Characterization of Gas Transport Behaviors in Shale Gas and Tight Gas Reservoirs by Digital Rock Analysis. *Int J Heat Mass Transf* **104** (January): 227–239. https://doi.org/10.1016/j.ijheatmasstransfer.2016.07.083.

Szegedy, C., Wei, L., Yangqing, J. et al. 2015. Going Deeper with Convolutions. Oral presentation given at the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, Massachusetts, USA, 7–12 June.

Taillon, J. A. 2016. *Advanced Analytical Microscopy at the Nanoscale: Applications in Wide Bandgap and Solid Oxide Fuel Cell Materials*. PhD dissertation, University of Maryland, College Park, College Park, Maryland, USA.

Taillon, J. A., Pellegrinelli, C., Huang, Y.-L. et al. 2018. Improving Microstructural Quantification in FIB/SEM Nanotomography. *Ultramicroscopy* **184A** (January): 24–38. https://doi.org/10.1016/j.ultramic.2017.07.017.

Tracey, J., Lin, S., Jankovic, J. et al. 2019. Iterative Machine Learning Method for Pore-Back Artifact Mitigation in High Porosity Membrane FIB-SEM Image Segmentation. *Microsc Microanal* **25** (S2): 186–187. https://doi.org/10.1017/S1431927619001661.

Tran, P. T. and Phong, L. T. 2019. On the Convergence Proof of AMSGrad and a New Version. *IEEE Access* **7** (1): 61706–61716. https://doi.org/10.1109/ACCESS.2019.2916341.

Urban, G., Geras, K. J., and Ebrahimi Kahou, S. 2016. Do Deep Convolutional Nets Really Need To Be Deep and Convolutional? https://arxiv.org/abs/1603.05691.

US Energy Information Administration (EIA). 2017. International Energy Outlook 2017 Overview. Report IEO2017, 14 September, https://www.eia.gov/outlooks/ieo/pdf/0484(2017).pdf (accessed 13 December 2020).

Wang, C. and Wu, Q. 2019. FLO: Fast and Lightweight Hyperparameter Optimization for AutoML.

Wang, H., Chen, L., Qu, Z. et al. 2020. Modeling of Multi-Scale Transport Phenomena in Shale Gas Production—A Critical Review. *Appl Energy* **262** (15 March): 114575. https://doi.org/10.1016/j.apenergy.2020.114575.

Weiss, K., Khoshgoftaar, T. M., and Wang, D. 2016. A Survey of Transfer Learning. *J Big Data* **3** (1): 9. https://doi.org/10.1186/s40537-016-0043-6.

Wu, Y. and Misra, S. 2020. Intelligent Image Segmentation for Organic-Rich Shales Using Random Forest, Wavelet Transform, and Hessian Matrix. *IEEE Geosci Remote Sens Lett* **17** (7): 1144–1147. https://doi.org/10.1109/lgrs.2019.2943849.

Wu, Y., Misra, S., Sondergeld, C. et al. 2019. Machine Learning for Locating Organic Matter and Pores in Scanning Electron Microscopy Images of Organic-Rich Shales. *Fuel* **253** (1 October): 662–676. https://doi.org/10.1016/j.fuel.2019.05.017.

Xia, Y., Goral, J., Huang, H. et al. 2017. Many-Body Dissipative Particle Dynamics Modeling of Fluid Flow in Fine-Grained Nanoporous Shales. *Phys Fluids* **29** (5): 056601. https://doi.org/10.1063/1.4981136.

Xiong, Q., Baychev, T. G., and Jivkov, A. P. 2016. Review of Pore Network Modelling of Porous Media: Experimental Characterisations, Network Constructions and Applications to Reactive Transport. *J Contam Hydrol* **192** (September): 101–117. https://doi.org/10.1016/j.jconhyd.2016.07.002.

Zhang, S., Byrnes, A. P., Jankovic, J. et al. 2019. Management, Analysis, and Simulation of Micrographs with Cloud Computing. *Micros Today* **27** (2): 26–33. https://doi.org/10.1017/s1551929519000026.

Zhang, S., Choromanska, A., and LeCun, Y. 2015. Deep Learning with Elastic Averaging SGD. http://arxiv.org/abs/1412.6651.

Zhang, Z., Liu, Q., and Wang, Y. 2018. Road Extraction by Deep Residual U-Net. *IEEE Geosci Remote Sens Lett* **15** (5): 749–753. https://doi.org/10.1109/LGRS.2018.2802944.

Zurada, J. 1992. *Introduction to Artificial Neural Systems*. St. Paul, Minnesota, USA: West Group.