Key Points:
- Convolutional neural network (CNN) can provide consistent and reliable image segmentation with limited user biases
- CNN models trained on ground truth data from SEM can provide results comparable to current segmentation standards
- Accurate absolute permeability, porosity, and pore size distribution can be obtained from CNN-segmented data

Correspondence to:
R. T. Armstrong,
ryan.armstrong@unsw.edu.au

Digital Rock Segmentation for Petrophysical Analysis With Reduced User Bias Using Convolutional Neural Networks

Yufu Niu[1] [iD], Peyman Mostaghimi[1] [iD], Mehdi Shabaninejad[2], Pawel Swietojanski[3], and Ryan T. Armstrong[1] [iD]

[1]School of Minerals and Energy Resources Engineering, The University of New South Wales, Sydney, New South Wales, Australia, [2]School of Physics and Engineering, The Australian National University, Canberra, ACT, Australia, [3]School of Computer Science and Engineering, The University of New South Wales, Sydney, New South Wales, Australia

Abstract Pore-scale digital images are usually obtained from microcomputed tomography data that has been segmented into void and grain space. Image segmentation is a crucial step in the process of digital rock analysis that can influence pore-scale characterization studies and/or the numerical simulation of petrophysical properties. This is concerning since all segmentation methods have user-selected parameters that result in biases. Convolutional neural networks (CNNs) provide a way forward since once trained, CNN can provide consistent and reliable image segmentation with no user-defined inputs. In this paper, a CNN is used to segment digital sandstone data, and various ground truth data sets are tested. The ground truth images are created based on high-resolution microcomputed tomography data and corresponding scanning electron microscope data. The results are evaluated in terms of porosity, permeability, and pore size distribution computed from the segmented data. We find that watershed-based segmentation provides a wide range of possible petrophysical values depending on user-selected thresholds, whereas CNN provides a smaller variance when trained on scanning electron microscope data. It can be concluded that CNN offers a reliable and consistent way to segment digital sandstone data for petrophysical analyses.

1. Introduction

Digital rock analysis has become an increasingly significant technology for the oil and gas industry and fundamental porous media research (Andrä et al., 2013; Blunt et al., 2013; Iassonov et al., 2009). For digital rock analysis, a rock sample is initially scanned by microcomputed tomography (micro-CT) to generate a 3-D image from which single or multiphase flow simulations are performed to obtain petrophysical parameters (Gao et al., 2017; Lin et al., 2018; Mostaghimi et al., 2013). For any digital rock study, image segmentation is a critical step that can have a profound influence on the results. Image segmentation, however, relies on user-selected parameters that can provide subjective results (Schlüter et al., 2014). To circumvent this issue, we propose a convolutional neural network (CNN) that once trained can be used to reduce the user bias associated with image segmentation.

CNN as a state-of-the-art bioinspired deep learning algorithm (Fukushima, 1980; Hubel & Wiesel, 1968) has been applied to many fields, such as image classification, object recognition, and natural language processing, and its origin can be traced back to the 1980s (LeCun, 1989; Rumelhart et al., 1986; Waibel et al., 1990). In the 1960s, researchers started to work on how to synthesize a neural network that can recognize patterns akin to human perception (*Kunihiko* Fukushima, 1975; Fukushima, 1980; Giebel, 1971; Rosenblatt, 1962). However, when the position or distortion of the input are changed, their model could not provide robust performance (Fukushima, 1980). Fukushima (1975) presented a self-organized multi-layered neural network called Cognitron. The results showed that Cognitron is similar to an animal's brain in many aspects and results demonstrated that Cognitron was more efficient than previous models. However, Cognitron is also affected by the position and distortion of the input data. Fukushima (1980) extended Hubel and Wiesel's (1962, 1965) hierarchy model and generated Neocognitron, which was the first model to achieve position invariant recognition meaning that stimulus patterns could be recognized even when the position of the stimulus pattern was shifted or the shape of the stimulus pattern was distorted. Modern CNNs were inspired by Fukushima (1980).

LeCun and Bengio (1995) explained why CNNs can solve the problem of shift invariance, which lead to current-day advanced CNNs that provide invariant properties, such as translation, rotation, scaling and deformation based on local receptive fields, weight sharing, and spatial subsampling. Hinton and Salakhutdinov (2006) presented a seminal paper that promoted progress in the field of deep learning. With the inspiration from this seminal paper, deep learning algorithms started to gather momentum, and deep CNNs (DCNNs) emerged. Krizhevsky et al. (2012) demonstrated a DCNN structure called AlexNet to solve an image classification problem that achieved the best performance ever reported for the challenge. Afterward, many DCNN structures were built for image processing, such as GoogleNet (Szegedy et al., 2015), VggNet (Simonyan & Zisserman, 2014), U-net (Ronneberger et al., 2015), ResNet (He et al., 2016), DenseNet (Huang et al., 2017), and SegNet (Badrinarayanan et al., 2017). With the help of deep learning algorithms, CNN was demonstrated to provide accurate results for various image processing tasks (Krizhevsky et al., 2012).

In recent years, researchers have started to focus on micro-CT image analysis by CNN in the medical field (Cernazanu-Glavan & Holban, 2013; Kang et al., 2017; Litjens et al., 2017; Lo et al., 1995; Roth et al., 2015; Shen et al., 2017). From 2016, the number of publications for DCNN techniques applied to micro-CT data has increased rapidly (Litjens et al., 2017). The main research application for medical imaging was classification, and many works have covered this topic in detail (Antony et al., 2016; Çiçek et al., 2016; Hosseini-Asl et al., 2016; Kawahara et al., 2017; Korez et al., 2016; Miao et al., 2016; Milletari et al., 2016; Moeskops et al., 2016; Payan & Montana, 2015; Ronneberger et al., 2015; van Grinsven et al., 2016; Wolterink et al., 2016). For digital rock image processing, we are starting to investigate if results similar to that obtained in the medical field can be obtained for geomaterials (Alqahtani et al., 2018; Sudakov et al., 2018). The main limitation, however, is access to geomaterial training data. Other fields of research have generated massive training data sets along with challenges, such as the ImageNet Large-Scale Visual Recognition Challenge (http://www.image-net.org/). Similar challenges for geomaterial data would facilitate the development of CNNs for geomaterial analyses; however, they currently do not exist.

Conventional image segmentation methods depend on user-selected parameters, such methods are Otsu (1979) thresholding, region-based segmentation (Adams & Bischof, 1994), watershed (Beucher, 1979), edge detection segmentation (Marr & Hildreth, 1980), and level-set method (Osher & Paragios, 2003). The user-selected parameters result in a range of possible segmentation outcomes. This subsequently results in discrepancies when determining petrophysical parameters from segmented digital images (Leu et al., 2014). In this paper, we propose that CNN-based image segmentation could solve this inconsistency problem since CNN provides a single result that has been optimized based on a collection of ground truth (GT) images. In recent years, CNN has been applied to predict porous media properties (Alqahtani et al., 2018; Sudakov et al., 2018), classify rock images (Cheng & Guo, 2017), and segment images (Wang et al., 2018); however, the resulting petrophysical properties predicted from CNN-segmented data have yet to be evaluated. To address this gap, this paper focuses on the petrophysical properties obtained from CNN-segmented data. We evaluate the importance of high-fidelity training data and the variance of petrophysical parameters obtained by CNN-based segmentation versus watershed-based segmentation.

2. Methodology

2.1. Workflow

Various GT data sets are generated to train the proposed CNN. The aim is to evaluate the overall influence of the training data on the network output. 3-D low-resolution (2.3 mm voxel length) micro-CT images of a North Sea sandstone provided by the Australian National University are interpolated (bicubic method) to 0.5 mm resolution (Shabaninejad, 2017). Then the corresponding scanning electron microscope (SEM) image (0.5 mm resolution) is registered to the related 3-D high-resolution micro-CT image. The image registration method used and parameters are explained in Latham et al. (2008). Afterward, the equivalent 2-D slice is extracted from the 3-D high-resolution micro-CT image. We segment the 2-D high-resolution micro-CT image and corresponding SEM image into three phases—pore, low-density minerals, and high-density minerals based on conventional segmentation methods—Otsu (1979) thresholding, and watershed-based segmentation (Leu et al., 2014). These segmented images are then used as selected GT data
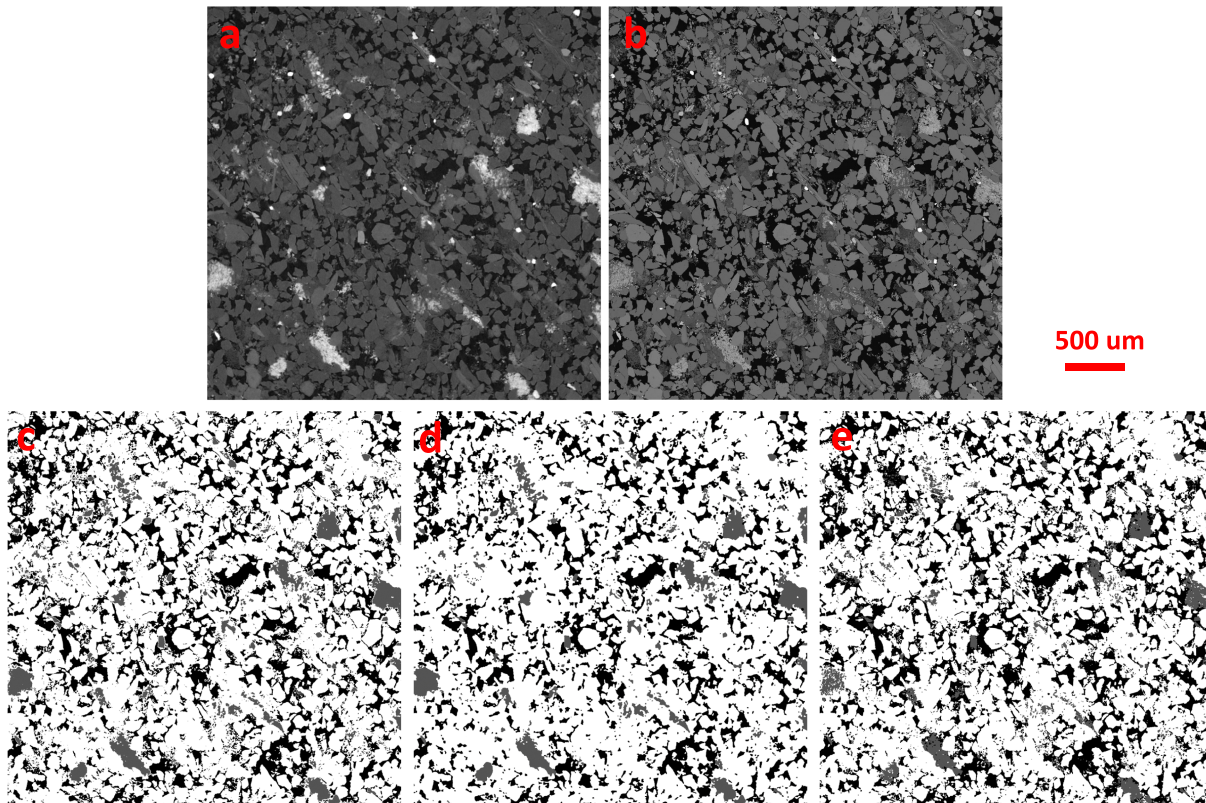
Figure 1. Image data sets used for CNN training and testing: (a) high-resolution micro-CT sandstone image; (b) SEM sandstone image; (c) Ground Truth 1 (GT1); (d) Ground Truth 2; (GT2); and (e) Ground Truth 3 (GT3).

sets. In Figure 1, we provide the micro-CT images of the sandstone and the corresponding SEM images along with the selected GT images, which are the following:

1. GT 1: Interpolated micro-CT data segmented to three phases by Otsu thresholding using ImageJ software;
2. GT 2: Interpolated micro-CT data segmented to three phases by watershed-based method using Avizo software; and
3. GT 3: SEM data segmented to three phases by watershed-based method using Avizo software.

In section 3, we demonstrate how the selected GT data influence the accuracy of the trained CNN.

Next, we prepare our GT data sets for CNN training. Figure 2 shows the workflow for this process, which can be explained in five steps.

Step 1. A series of grayscale image patches of size 81×81 pixels are extracted from the micro-CT sandstone image ($6,100 \times 6,100$ pixels). Moving an 81×81 window along the image row-by-row and column-by-column generates the image patches. The overlapping stride is defined by 20 pixels in both the X and Y directions. The total number of image patches used for CNN training and testing is 90,601. Selection of the image patch size is addressed in section 3.1.

Step 2. We repeat Step 1 using the GT data. Therefore, each grayscale image patch has a corresponding segmented image patch. The center pixel value of the segmented image patch ($0 =$ pore, $1 =$ low-density minerals, and $2 =$ high-density mineral) represents the label for the corresponding grayscale image patch.

Step 3. For each GT data set, we train one CNN model. Seventy percent of the image patches (63,420 patches) are used for training, and 30% are used for testing (27,181 patches). The number of image patches is recorded from 0 to 90,600 based on the location of image patches from top left to bottom
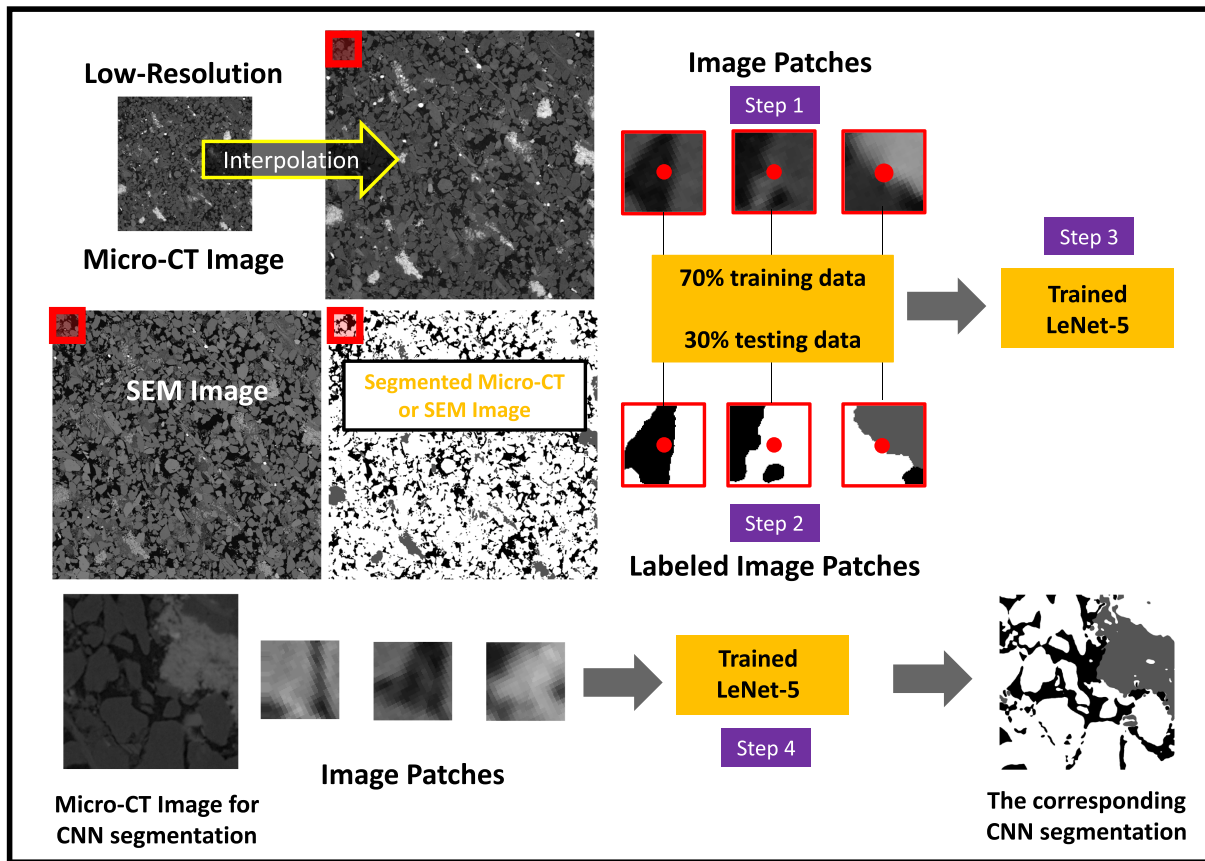
Figure 2. The overall workflow of CNN training and testing.

right on the micro/SEM image. The training and testing regions are separated by 70% from top left and 30% from bottom right.

Step 4. After 50 epochs of training, we utilize the trained CNN to segment data from a validation region that was not used for training. To achieve this, we supplement 40 pixels on the image boundaries of the validation region and extract 81×81 pixels (10^6 image patches) with an overlap of 1 pixel. Then all image patches are imported to the trained CNN to obtain the segmented data.

Step 5. According to the validation region results, we select the trained CNN that provided the best results. We then build a 3-D segmented image. We use the trained CNN model to segment 2-D slices one-by-one to generate the 3-D image. Absolute permeability, porosity, and pore size distribution of the CNN segmentation are measured and compared to values obtained by watershed-based segmentation.

2.2. CNN Training/Testing

The trained CNN is a modified version of LeNet-5 (LeCun et al., 1998). The network consists of two convolutional layers and two max-pooling layers, followed by a fully connected layer.The input is an 8-bit grayscale 81 x 81 pixels image patch. The output is a connected *softmax* layer with three possible labels corresponding to pore (0), low-density mineral (1), or high-density mineral (2). The convolutional layers have a kernel size of 5×5, and padding is 2 in order to preserve the original image size. Max-pooling layers have 2×2 kernel sizes, and the stride is equal to the kernel size. These two convolutional layers and the first fully connected layer are connected with a rectified linear unit activation function (Nair & Hinton, 2010) to generate nonlinear activations. We train the LeNet-5 parameters using the stochastic gradient descent method (Bishop, 2006) and error back propagation algorithm (Rumelhart et al., 1986). Cross entropy is applied as the loss function to calculate error
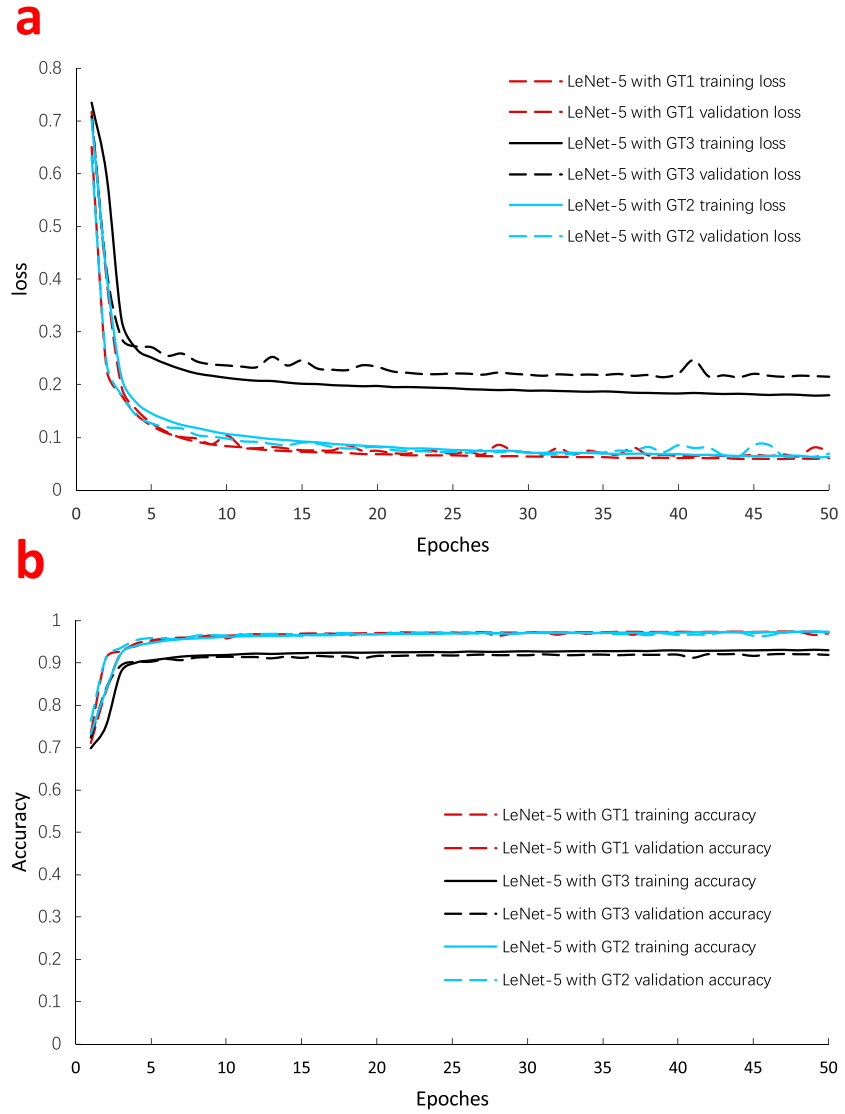
Figure 3. (a) LeNet-5 training and testing loss and (b) LeNet-5 training and testing accuracy.

$$L(\pounds) = -\frac{1}{N}\sum_{i=1}^{N}[T_i \ln P_i], \tag{1}$$

where L is the loss function, $\pounds$ denotes optimized LeNet-5 parameter weights and biases, N is the number of training image patches, T is the GT label, and P is the predicted label.

The weights are updated by the stochastic gradient descent with momentum

$$w_{i+1} = \lambda w_i - \eta \frac{\partial L}{\partial w}\bigg|_{w=w_i}, \tag{2}$$

$$b_{i+1} = \lambda b_i - \eta \frac{\partial L}{\partial b}\bigg|_{b=b_i}, \tag{3}$$

where w is weight, b is bias, L is the optimized loss, η is learning rate, and λ is learning momentum. In this work, we set $\eta = 0.0005$ and $\lambda = 0.9$. The LeNet-5 is implemented using PyTorch (https://pytorch.org/), and the initial weights and biases are sampled from a uniform distribution u, as follows:
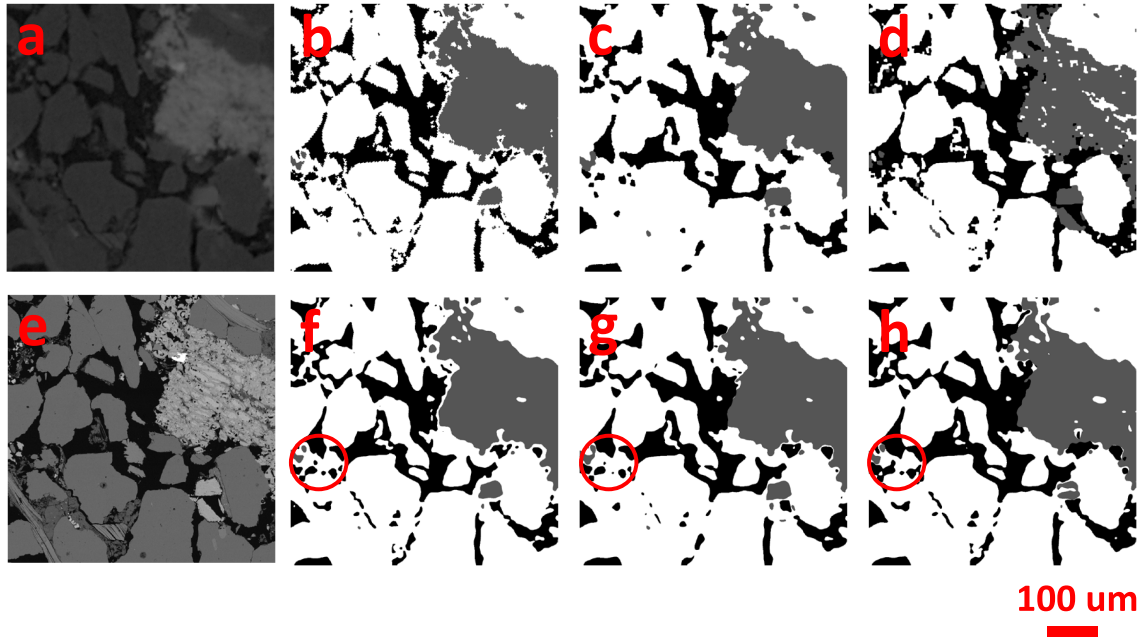
Figure 4. The 1,000² pixels image from validation region: (a) micro-CT image; (b) GT1 image; (c) GT2 image; (d) GT3 image; (e) SEM image; (f) LeNet-5 segmentation based on GT1; (g) LeNet-5 segmentation based on GT2; and (h) LeNet-5 segmentation based on GT3.

$$u\left(-\sqrt{a}, \sqrt{a}\right), \tag{4}$$

$$a = \frac{1}{C_{in}k}, \tag{5}$$

where C_{in} is the number of input image channels and k is the kernel size.

3. Result and Discussion

3.1. LeNet-5 Training

The first step is determination of the patch size for training. The patch size needs to represent enough local spatial information for accurate segmentation while not providing too much information resulting in overfitting. A simple exercise can be performed to measure the total loss based on the patch size used for the training and validation data sets. While the training loss consistently decreases with patch size, the validation loss decreases and then increases with increasing patch size. Based on this analysis, the optimal patch size was determined to be 81×81 pixels.

Figure 3 shows the losses and accuracies for both the model training and validation using the target data—GT1, GT2, and GT3—with an 81×81 patch size. As can be seen in Figure 3, the accuracy and loss functions are found to converge after ~20 epochs. The accuracy is computed by the number of predicted true labels divided by the total number of labels. The main reason for fast convergence is that (1) the output of the LeNet-5 is only three labels. So, the LeNet-5 learns the features easily without too many rounds of training, and (2) the architecture of the LeNet-5 is simple, such that the LeNet-5 can easily reveal the internal features of the image. Albeit the simplicity of the model, the results, as shown in the next section, are comparable to current standards for digital rock segmentation (Wildenschild & Sheppard, 2013).

3.2. 2-D Segmentation

Figure 4 shows the 2-D micro-CT/SEM data, GT data, and LeNet-5 segmentations from a validation region, which is not used for training. From the SEM image on Figure 4c, we can detect small submicrometer pores that provide more detailed features in the segmented image. From Figures 4d to 4f, we can see that the GT images have sawtoothed boundaries especially on Figure 4d since Otsu-thresholding does not account for
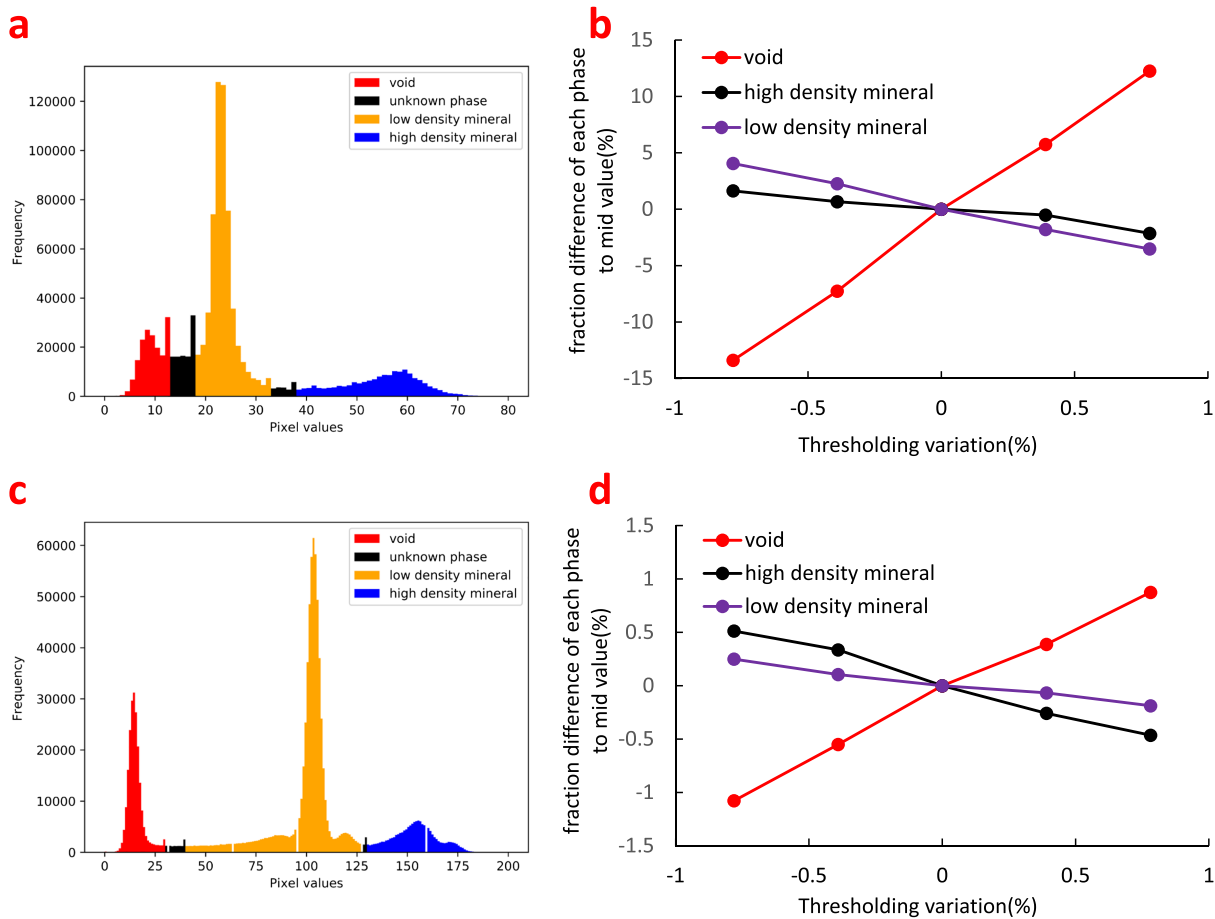
Figure 5. The sensitivity of micro-CT/SEM GT image on training regions. (a) Pixel distribution of micro-CT image; (b) sensitivity of GT2 image on various thresholding values; (c) pixel distribution of SEM image; and (d) sensitivity of GT3 image on various thresholding values.

local image gradients. Figures 4g to 4i show that the LeNet-5 segmentation trained with GT3 restores more information than the LeNet-5 segmentations trained with GT1 and GT2 even though the LeNet-5 trained with GT3 is less accurate (accuracy = 91.8%) than the LeNet-5 with GT1 and GT2 (accuracy = 97%).

There are two main reasons for the lower accuracy of the LeNet-5 trained with GT3. First, to obtain the SEM image, the prepared sample surface needs to be cut and polished. Consequently, the sample may be slightly damaged, and the SEM image may not entirely match its corresponding micro-CT image. Second, SEM data can detect submicrometer pore spaces. These submicrometer pores are too small to be present in the micro-CT image, which is interpolated to the SEM image resolution. Nevertheless, the LeNet-5 trained with GT3 does appear to restore more small pores and pore throats than the LeNet-5 trained with GT1 and GT2. Overall, the LeNet-5 trained with GT3 appears to provide smoother segmented data than the corresponding networks trained with the other GT data sets. The output of the LeNet-5 trained with GT3 is quantified in the following section once 3-D data are generated.

We further examine the thresholding sensitivity (Leu et al., 2014) for GT2 and GT3 images that are generated from the watershed-based segmentation method. The thresholding sensitivity can be used to determine how thresholding values influence the segmentation. Figure 5a provides the pixel intensity histogram of the micro-CT image in the training region. The histogram displays no clear boundaries between pore space (void) and high-density and low-density minerals. Therefore, the exact configuration of how to set the upper and lower bounds for initialization of the watershed-based segmentation is unclear. The relationship between thresholding variation and the fraction of each phase are provided in Figure 5b. The zero point of the X axis represents the optimum micro-CT segmentation as defined by visual inspection. One percent
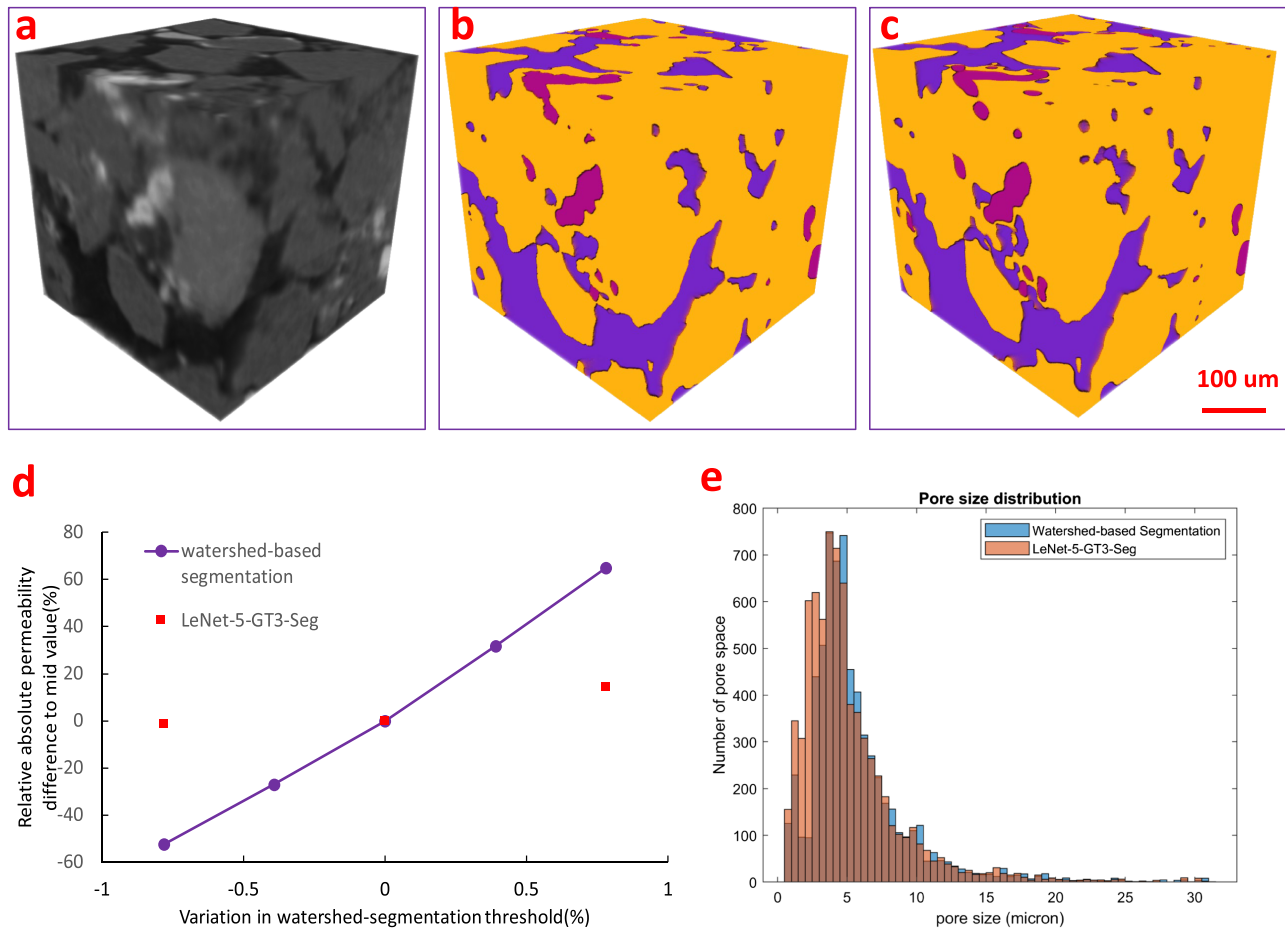
Figure 6. (a) The 500^3 voxels micro-CT sandstone image; (b) 500^3 voxels watershed-based segmentation image; and (c) 500^3 voxels LeNet-5-GT3-Seg image. (d) The difference of absolute permeability which are from 3-D watershed-based segmentation and LeNet-5-GT3-Seg with various thresholding values. (e) Pore size distribution of 3-D watershed-based segmentation and LeNet-5-GT3-Seg.

variation from the optimum set point is also a potential solution for other users. As shown in Figure 5b, the thresholding variation can significantly influence the measured fraction of the phases obtained from the segmented image. For example, a 1% change in threshold resulted in ~15% change for total void space.

The same sensitivity analysis is also implemented on the GT3 data, and the results are presented in Figures 5c and 5d. The histogram in Figure 5c demonstrates clear boundaries between pore space (void) and high-density and low-density minerals. As a result, there is only a slight variation in the determined phase fractions when different thresholding values are applied for segmentation. Overall, the GT3 data are less sensitive to thresholding variations than the GT2 data. In Figure 5d, we observe that a 1% change in threshold resulted in ~1.5% change for total void space, which is 10 times less than what the micro-CT data provided in Figure 5b. This makes the high contrast and low noise SEM data (GT3) suitable for training the LeNet-5 model.

3.3. 3-D Segmentation

To verify the sensitivity of petrophysical parameters measured from the CNN-based segmentation, we built 3-D images (500^3 voxels) using LeNet-5 trained with GT3 data. For simplicity, this model will be called LeNet-5-GT3-Seg from hereon. A qualitative comparison between the CNN- and watershed-based segmented data sets are displayed in Figure 6a. Overall, both methods capture similar features of the pore space with a few smaller pores appearing in the CNN-based segmentation.

We estimate the absolute permeability for the 3-D segmented data sets by using an elliptic flow equation as presented by Chung et al. (2019) and Da Wang et al. (2019), and porosity is determined by pixel counting. The results show that the absolute permeability and porosity for the watershed-based segmentation are 1.11 D and 24.2%, while the LeNet-5-GT3-Seg are 1.23 D and 24.1%, respectively. It can be concluded that the results from the LeNet-5-GT3-Seg are comparable to the result of the optimum watershed-based segmentation. However, the comparison is subjective to user-selected thresholding values. Therefore, we also provide the thresholding sensitivities for both the watershed-based segmentation and LeNet-5-GT3-Seg (Figure 6b). When selecting various thresholding values for the GT3 data and then re-training the network, the maximum absolute permeability variation of LeNet-5-GT3-Seg is 14.2%, while the maximum absolute permeability variation of the watershed-based segmentation is 64.9%. Overall, the LeNet-5-GT3-Seg provides less variation in the predicted absolute permeability than the watershed-based segmentation. This suggests that LeNet-5 can provide digital rock data for petrophysical analysis that is more consistent than that obtained by watershed-based segmentation.

To estimate the pore size distribution between the optimum 3-D watershed-based segmentation and LeNet-5-GT3-Seg data, we first apply a Euclidean distance transformation to the segmented domains (Breu et al., 1995). Then each pore region is identified by the local maximum inscribed sphere. Figure 6c depicts the resulting pore size distributions for the optimum segmentations. The result shows that the main difference between watershed-based segmentation and LeNet-5-GT3-Seg is that smaller radius pores are identified by LeNet-5-GT3-Seg, which is aligned with the previous visual observations. The main reason is that LeNet-5-GT3-Seg is trained with high-resolution SEM data, which provides more details on the spatial features in the image than that obtained by direct watershed-based segmentation of the micro-CT data.

4. Conclusion

We studied the segmentation of a North Sea sandstone by using LeNet-5. Otsu thresholding and watershed-based methods were used to segment micro-CT and corresponding SEM images to generate GT images from which LeNet-5 models were trained. The LeNet-5 models were initially used to segment 2-D micro-CT data, which demonstrated that the LeNet-5 trained on the SEM data (GT3) provided more detailed features of the pore structures than that captured by the other models. In addition, according to the thresholding sensitivity study conducted on the micro-CT and SEM data, the porosity of the SEM data was 10 times less sensitive than the micro-CT image to threshold variation. We then construct a segmented 3-D image based on LeNet-5 trained with SEM data and compared the results to that obtained by watershed-based segmentation. Petrophysical parameters including absolute permeability, porosity, and pore size distribution resulting from the segmented images are shown to be similar. However, the absolute permeability of the LeNet-5 model (LeNet-5-GT3-Seg) provided 4 times less sensitivity to threshold variation and identified more smaller pores than the watershed-based segmentation.

Overall, the presented LeNet-5-based method provided more consistent petrophysical analysis than that provided by watershed-based segmentation. Image segmentation is indeed a complex problem with various caveats depending on image quality and sample complexities, and thus, image processing workflows are often determined on a case-by-case basis. This is concerning given that the subsequent outputs, such as petrophysical parameters, are highly sensitive to user-selected parameters. While the presented results are from a single sandstone sample, we demonstrate that LeNet-5 can provide a consistent means to segment data for petrophysical analyses. As more SEM and micro-CT data become readily available for training, the LeNet-5 could be updated and potentially generalized for sandstone images. In our case, we applied 2-D subimages for training, which allowed for incorporation of high-quality SEM data. Despite the 2-D nature of the training data, the network provided results comparable to the watershed-based segmentation. However, 3-D subimages are also possible if the computation resources are not limited. Future works should focus on the robustness of LeNet-5 for segmenting a range of different sandstone rocks and identification of more complex mineralogical content. In parallel, the application of LeNet-5 to segment dual-porosity carbonate systems should be considered, where the application of superresolution data (Wang et al., 2019) would be highly applicable.

References

Adams, R., & Bischof, L. (1994). Seeded region growing. *IEEE Transactions on Pattern Analysis Machine Intelligence*, *16*(6), 641–647.

Alqahtani, N., R. T. Armstrong, and P. Mostaghimi (2018), Deep learning convolutional neural networks to predict porous media properties, paper presented at SPE Asia Pacific Oil and Gas Conference and Exhibition, Society of Petroleum Engineers.

Andrä, H., Combaret, N., Dvorkin, J., Glatt, E., Han, J., Kabel, M., et al. (2013). Digital rock physics benchmarks—Part I: Imaging and segmentation. *Computers Geosciences*, *50*, 25–32.

Antony, J., K. McGuinness, N. E. O'Connor, and K. Moran (2016). Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks, paper presented at Pattern Recognition (ICPR), 2016 23rd International Conference on, IEEE.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis Machine Intelligence*, *39*(12), 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615

Beucher, S. (1979). Use of watersheds in contour detection, paper presented at Proceedings of the International Workshop on Image Processing, CCETT.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Blunt, M. J., Bijeljic, B., Dong, H., Gharbi, O., Iglauer, S., Mostaghimi, P., et al. (2013). Pore-scale imaging and modelling. *Advances in Water Resources*, *51*, 197–216.

Breu, H., Gil, J., Kirkpatrick, D., & Werman, M. (1995). Linear time Euclidean distance transform algorithms. *IEEE Transactions on Pattern Analysis Machine Intelligence*, *17*(5), 529–533.

Cernazanu-Glavan, C., & Holban, S. (2013). Segmentation of bone structure in X-ray images using convolutional neural network. *Advances in Electrical and Computer Engineering*, *13*(1), 87–94.

Cheng, G., & Guo, W. (2017). *Rock images classification by using deep convolution neural network, paper presented at Journal of Physics*, *Conference Series*. Xi'an Shiyou University Xi'an Shanxi 710065 China: IOP Publishing.

Chung, T., Wang, Y. D., Armstrong, R. T., & Mostaghimi, P. (2019). Approximating permeability of microcomputed-tomography images using elliptic flow equations. *SPE Journal*, *24*(03), 1154–1163. https://doi.org/10.2118/191379-PA

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). *3D U-Net: learning dense volumetric segmentation from sparse annotation, paper presented at International Conference on Medical Image Computing and Computer-Assisted Intervention*. Athens, Greece: Springer.

Da Wang, Y., Chung, T., Armstrong, R. T., McClure, J. E., & Mostaghimi, P. (2019). Computations of permeability of large rock images by dual grid domain decomposition. *Advances in Water Resources*, *126*, 1–14.

Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics*, *20*(3-4), 121–136. https://doi.org/10.1007/bf00342633

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193–202. https://doi.org/10.1007/bf00344251

Gao, Y., Lin, Q., Bijeljic, B., & Blunt, M. J. (2017). X-ray microtomography of intermittency in multiphase flow at steady state using a differential imaging method. *Water Resources Research*, *53*, 10,274–10,292. https://doi.org/10.1002/2017WR021736

Giebel, H. (1971). *Feature extraction and recognition of handwritten characters by homogeneous layers, in Zeichenerkennung durch biologische und technische Systeme/Pattern Recognition in Biological and Technical Systems*, (pp. 162–169). Berlin, Technical University: Springer.

He, K., X. Zhang, S. Ren, and J. Sun (2016), Deep residual learning for image recognition, paper presented at Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507.

Hosseini-Asl, E., G. Gimel'farb, and A. El-Baz (2016), Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional network, *arXiv preprint arXiv:1607.00556*.

Huang, G., Z. Liu, L. Van Der Maaten, and K. Q. Weinberger (2017), Densely connected convolutional networks, paper presented at CVPR.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106–154.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, *28*(2), 229–289.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455

Iassonov, P., Gebrenegus, T., & Tuller, M. (2009). Segmentation of X-ray computed tomography images of porous materials: A crucial step for characterization and quantitative analysis of pore structures. *Water Resources Research*, *45*, W09415. https://doi.org/10.1029/2009WR008087

Kang, E., Min, J., & Ye, J. C. (2017). A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Medical Physics*, *44*(10).

Kawahara, J., Brown, C. J., Miller, S. P., Booth, B. G., Chau, V., Grunau, R. E., et al. (2017). BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, *146*, 1038–1049. https://doi.org/10.1016/j.neuroimage.2016.09.046

Korez, R., Likar, B., Pernuš, F., & Vrtovec, T. (2016). *Model-based segmentation of vertebral bodies from MR images with 3D CNNs, paper presented at International Conference on Medical Image Computing and Computer-Assisted Intervention*. Athens, Greece: Springer.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012), Imagenet classification with deep convolutional neural networks, paper presented at Advances in Neural Information Processing Systems.

Latham, S., T. Varslot, and A. Sheppard, Abu Dhabi, UAE (2008), Image registration: Enhancing and calibrating X-ray micro-CT imaging, 1-12.

LeCun, Y. (1989). Generalization and network design strategies. *Connectionism in Perspective*, *19*, 143–155.

LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks*, *3361*(10), 1995.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Leu, L., Berg, S., Enzmann, F., Armstrong, R., & Kersten, M. (2014). Fast X-ray micro-tomography of multiphase flow in berea sandstone: A sensitivity study on image processing. *Transport in Porous Media*, *105*(2), 451–469.

Lin, Q., Bijeljic, B., Pini, R., Blunt, M. J., & Krevor, S. (2018). Imaging and measurement of pore-scale interfacial curvature to determine capillary pressure simultaneously with relative permeability. *Water Resources Research*, *54*, 7046–7060. https://doi.org/10.1029/2018WR023214

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Lo, S.-C. B., Chan, H.-P., Lin, J.-S., Li, H., Freedman, M. T., & Mun, S. K. (1995). Artificial convolution neural network for medical image pattern recognition. *Neural Networks*, *8*(7-8), 1201–1214.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *207*(1167), 187–217.

Miao, S., Wang, Z. J., & Liao, R. (2016). A CNN regression approach for real-time 2D/3D registration. *IEEE Transactions on Medical Imaging*, *35*(5), 1352–1363.

Milletari, F., N. Navab, and S.-A. Ahmadi (2016), V-net: Fully convolutional neural networks for volumetric medical image segmentation, paper presented at 3D Vision (3DV), 2016 Fourth International Conference, IEEE.

Moeskops, P., Wolterink, J. M., van der Velden, B. H., Gilhuijs, K. G., Leiner, T., Viergever, M. A., & Išgum, I. (2016). *Deep learning for multi-task medical image segmentation in multiple modalities, paper presented at International Conference on Medical Image Computing and Computer-Assisted Intervention*. Athens, Greece: Springer.

Mostaghimi, P., Blunt, M. J., & Bijeljic, B. (2013). Computations of absolute permeability on micro-CT images. *Mathematical Geosciences*, *45*(1), 103–125.

Nair, V., and G. E. Hinton (2010), Rectified linear units improve restricted Boltzmann machines, paper presented at Proceedings of the 27th international conference on machine learning (ICML-10).

Osher, S., & Paragios, N. (2003). *Geometric level set methods in imaging, vision, and graphics*. New York: Springer Science & Business Media.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, *9*(1), 62–66.

Payan, A., and G. Montana (2015), Predicting Alzheimer's disease: A neuroimaging study with 3D convolutional neural networks, *arXiv preprint arXiv:1502.02506*.

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-net: Convolutional networks for biomedical image segmentation, paper presented at International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich, Germany: Springer.

Rosenblatt, F. (1962), Principles of Neurodynamics.

Roth, H. R., A. Farag, L. Lu, E. B. Turkbey, and R. M. Summers (2015), Deep convolutional networks for pancreas segmentation in CT imaging, paper presented at Medical Imaging 2015: Image Processing, International Society for Optics and Photonics.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.

Schlüter, S., Sheppard, A., Brown, K., & Wildenschild, D. (2014). Image processing of multiphase images obtained via X-ray microtomography: A review. *Water Resources Research*, *50*, 3615–3639. https://doi.org/10.1002/2014WR015256

Shabaninejad, M. (2017), Pore-scale analysis of residual oil in sandstones and its dependence on waterflood salinity, analysed by tomography and microscopy.

Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, *19*, 221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442

Simonyan, K., and A. Zisserman (2014), Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*.

Sudakov, O., E. Burnaev, and D. Koroteev (2018), Driving digital rock towards machine learning: Predicting permeability with gradient boosting and deep neural networks, *arXiv preprint arXiv:1803.00758*.

Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al. (2015), Going deeper with convolutions, paper presented at Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

van Grinsven, M. J., van Ginneken, B., Hoyng, C. B., Theelen, T., & Sánchez, C. I. (2016). Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images. *IEEE Transactions on Medical Imaging*, *35*(5), 1273–1284. https://doi.org/10.1109/TMI.2016.2526689

Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., & Lang, K. J. (1990). *Phoneme recognition using time-delay neural networks, in Readings in Speech Recognition*, (pp. 393–404). New York City: Elsevier.

Wang, Y., Armstrong, R., & Mostaghimi, P. (2019). Enhancing resolution of digital rock images with super resolution convolutional neural networks. *Journal of Petroleum Science Engineering*, *182*. https://doi.org/10.1016/j.petrol.2019.106261

Wang, Y., Arns, C. H., Rahman, S. S., & Arns, J.-Y. (2018). Porous structure reconstruction using convolutional neural networks. *Mathematical Geosciences*, *50*(7), 781–799. https://doi.org/10.1007/s11004-018-9743-0

Wildenschild, D., & Sheppard, A. P. (2013). X-ray imaging and analysis techniques for quantifying pore-scale structure and processes in subsurface porous medium systems. *Advances in Water Resources*, *51*, 217–246.

Wolterink, J. M., Leiner, T., de Vos, B. D., van Hamersvelt, R. W., Viergever, M. A., & Išgum, I. (2016). Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. *Medical Image Analysis*, *34*, 123–136. https://doi.org/10.1016/j.media.2016.04.004