



Segmentation of digital rock images using deep convolutional autoencoder networks[☆]

Sadegh Karimpouli^a, Pejman Tahmasebi^{b,*}

^a Mining Engineering Group, Faculty of Engineering, University of Zanjan, Zanjan, Iran

^b Department of Petroleum Engineering, University of Wyoming, Laramie, WY, 82071, USA

ARTICLE INFO

Keywords:

Segmentation
Digital rock physics (DRP)
Mineral identification
Artificial intelligence

ABSTRACT

Segmentation is a critical step in Digital Rock Physics (DRP) as the original images are available in a gray-scale format. Conventional methods often use thresholding to delineate distinct phases and, consequently, watershed algorithm to identify the existing phases. Such methods are based on color contrast, which makes it difficult to automatically differentiate phases with similar colors and intensities. Recently, deep learning and machine learning algorithms have proposed several algorithms working with images, including Convolutional Neural Networks (CNN). Among them, convolutional autoencoder networks have produced accurate results in different applications when various images are available for the training. In this paper, thus, convolutional autoencoder algorithm is implemented to enhance segmentation of digital rock images. However, the bottleneck for applying the CNN algorithms in DRP is the limited available rock images. As an effective data augmentation method, a cross-correlation based simulation was used to increase the necessary dataset in this study. Therefore, using the originally available dataset, namely 20 images from Berea sandstone, a training seed comprising of the manually and semi-manually segmented images was used. Then, the dataset is divided into training, validation and testing groups with a fraction of 80, 10 and 10%, respectively. Next, the produced dataset is given to our stochastic image generator algorithm and 20000 realizations, along with their segmented images, are produced simultaneously. The implemented CNN algorithm was tested for two versions of basic and extended architectures. The results show that the extended network produces results with 96% of categorical accuracy using the designated images in the testing group. Finally, a qualitative comparison with the conventional multiphase segmentation (multi-thresholding) revealed that our results are more accurate and reliable even if very few rock images are available.

1. Introduction

Rock physics provides the relationships between the physical properties of the porous structure of rock and remotely-sensed geophysical measurements. Recently, emerging the high-resolution micro-computed tomography (μ CT) images of rock samples have led to a remarkable development in Digital Rock Physics (DRP). In the standard workflow of DRP, segmentation of pore and minerals to separate phases is a vital step. The segmentation methods developed for DRP are extensively reviewed in the previous publications (Iassonov et al., 2009; Sezgin, 2004). Among them, the procedure introduced by Visual Science Group (VSG), Stanford University (SU) and Kongju (KJ) segmentations are the most effective available frameworks (Andrä et al., 2013). These methods are summarized in the following steps: 1. Boundary

detection using gradient magnitude of the image and excluding them, 2. Thresholding the reminder pixels as pore and minerals phases, and 3. Expanding all phases to boundary pixels by a marker-based watershed algorithm (Beucher and Meyer, 1992). Marker detection in minerals with the complicated pattern is not straightforward and there may be no general method to achieve this (Beucher and Meyer, 1992). Although most of the segmentation methods are based on image processing algorithms occurring in an automatic framework, manually controlling is essential in each step. For instance, thresholding may fail when no color contrast is observed between two separate phases. In other words, there may be two different minerals in the rock image with similar colors. This issue happens due to either their close densities or limited detection power of the imaging instrument. For the sake of simplicity, researchers often consider the rock images as a two-phase

[☆] Dr. Sadegh Karimpouli and Dr. Pejman Tahmasebi together conceived the problem, developed the method, performed the computations and contributed to analyzing the results and writing the paper.

* Corresponding author.

E-mail addresses: s.karimpouli@znu.ac.ir (S. Karimpouli), ptahmase@uwyo.edu (P. Tahmasebi).

<https://doi.org/10.1016/j.cageo.2019.02.003>

Received 19 April 2018; Received in revised form 12 October 2018; Accepted 5 February 2019

Available online 06 February 2019

0098-3004/ © 2019 Elsevier Ltd. All rights reserved.

(porosity and mineral) sample (Karimpouli et al., 2018; Fattahi and Karimpouli, 2016; Karimpouli and Fattahi, 2016). This simplification is, however, not applicable in general and strongly affects the subsequent computations of rock physical parameters, particularly for P- and S-wave velocities (Andrä et al., 2013).

Artificial Neural Network (ANN) is a class of Machine Learning algorithms inspired by the human brain. ANNs learn to perform tasks such as classification or prediction if they are trained by some examples. Recent developments in ANN and, in particular, deep learning has offered new possibilities, to tackle very complex problems in image analysis. One of such an algorithm is the Convolutional Neural Networks (CNN) (Krizhevsky et al., 2012; Lecun et al., 1998), which use convolution and pooling functions to extract new features for analyzing visual imagery. These networks, which are considered as one of the deep learning derivatives, have manifested significant differences in terms of accuracy and effectiveness compared to the conventional networks (Garcia-Garcia et al., 2017). CNN have been used for many applications, including face detection (Li et al., 2015), semantic segmentation (Garcia-Garcia et al., 2017), video analysis in autonomous driving (Badrinarayanan et al., 2015), speech recognition (Y. Zhang et al., 2017b) and medical image analysis and applications (Havaei et al., 2017; Litjens et al., 2017; Wallach et al., 2015). In geosciences, and particularly in rock physics, deep learning methods and CNN are used in different applications such as: lithology detection using borehole imaging (P. Y. Zhang et al., 2017a), rock type classification (Cheng and Guo, 2017; Ferreira and Giraldo, 2017), permeability prediction (Srisutthiyakorn, 2016) and reconstruction of rock porous media (Laloy et al., 2017; Mosser et al., 2017). In this work, we aim to use this powerful method for rock image segmentation. Thus, deep learning segmentation methods are first reviewed briefly.

As mentioned earlier, CNN is capable to be used for pixel labeling problems or segmentation. The utmost significant advantage that makes the CNN to surpass the conventional methods is the ability to learn from patterns, features, and textures rather than only relying on color variation. According to a recent research performed by Garcia-Garcia et al. (2017), the most successful state-of-the-art deep learning based segmentation methods are fully convolutional networks (Shelhamer et al., 2017). In the fully convolutional networks, a set of connected layers are replaced by the convolutional layers, which produce spatial maps instead of classification scores. In fact, such maps are deconvolved in an up-sampling procedure to obtain a per-pixel labeled image (i.e. segmented image). SegNet (Badrinarayanan et al., 2015), an encoder-decoder convolutional network, is one of the currently best available networks of this kind which has been used in different applications and has demonstrated promising results (Garcia-Garcia et al., 2017; Kendall et al., 2015; Nanfack et al., 2017).

One of the major issues in using deep learning methods for DRP is the limited dataset that is available for the training step. Acquiring μ CT images is both expensive and time-consuming. Moreover, due to either long-term procedure of sample preparation or limited core samples, preparing hundreds (or thousands) polished thin-section samples for microscope imaging may not be plausible in subsurface applications. Therefore, data augmentation methods should be applied to increase the existing dataset. Using such big data would lead to more effective training and, thus, avoiding overfitting by either a fast convergence or better regularizing. The common methods for data augmentation are the available image transformation operators, such as rotation, translation, scaling, crops, etc. For example, Mosser et al. (2017) cropped several small size images from a large image and used an overlap of 12–64% between small images to increase the dataset. In this paper, however, a more efficient reconstruction method, namely Hybrid pattern- and pixel-based Simulation (HYPPS) introduced by Tahmasebi (2017), is used for data augmentation. HYPPS uses an input image and generates any number of realizations with different structures and any sizes, but similar statistics. HYPPS is a powerful tool which has been used in several applications to reconstruct new scenarios of

heterogeneous media such as sandstone and carbonate samples (Karimpouli and Tahmasebi, 2016), coal samples (Karimpouli et al., 2017), small-scale porous media modeling (Tahmasebi, 2018a; Tahmasebi, 2018c; Tahmasebi and Kamrava, 2018) and unconventional plays (Tahmasebi et al., 2017; Tahmasebi et al., 2018; Tahmasebi, 2018b).

In this paper, first, we tackle the problem of limited data in DRP by the HYPPS algorithm as an effective augmentation approach. Then, the SegNet is used to overcome the difficulties and drawbacks of the conventional methods for segmentation of digital rock images. This workflow can be considered as an opening for the deep learning techniques in the vast world of DRP. In the following sections, we will introduce the CNN, SegNet and HYPPS algorithms in Section 2. In Section 3, the utilized Berea sandstone is described as a benchmark data in the DRP studies. Then, we produce a limited number of ground-truth (or segmented) images semi-manually and increase the numbers using the HYPPS method. Next, the two versions of the SegNet are used for segmentation and their respective segmented results are compared. Finally, the results are discussed in Section 4.

2. Basic concepts

Two main algorithms that are used in this paper are the SegNet for image-based segmentation and the HYPPS method for simulation and data augmentation. In this part, a brief description of these algorithms as well as CNN, as the core of the image-based neural network, is provided.

2.1. Convolutional Neural Networks

Convolutional Neural Networks are a part of a large group of deep learning methods. They attracted attention due to their strong abilities in image classification/recognition (He et al., 2015). CNN are trained through their convolutional layers to recognize various patterns in the input images. Small size kernels are pillars of the convolutional layers. Indeed, they effectively extract high-level characteristics of the input image. Convolutional layers are followed by a fully connected neural network, which is used to translate those features obtained from the previous layers to the given output phases. The basic layers in a common CNN are as follow:

1. **Input layer:** Images are considered as input data, which are introduced to CNN in this layer.
2. **Convolutional layer:** In this layer, input images or feature maps from the last layer are convolved with some small size filters (or kernels) to generate new feature maps. These convolutions are being performed with a shift of 'n' pixels, which are called *stride* (Krizhevsky et al., 2012). In fact, stride controls how the filter convolves around the input image.
3. **ReLU (Rectified Linear Units) Layer:** The purpose of this layer is to introduce nonlinearity to a system that basically has just been computing by linear operations (multiplications and summations) during the convolutional layers. The ReLU layer applies the function $f(x) = \max(0, x)$ to all the values in the input volume. The logic behind ReLU is that this layer changes all the negative values to zero.
4. **Max-pooling layer:** This layer, usually known as down-sampling, is used to summarize data by choosing the local maximum in a sliding window moving across the feature maps with a stride of the same length.
5. **Fully connected layer:** It is similar to the traditional Multi-Layer Perceptron (MLP) neural networks (Haykin, 1999) and is used to translate feature maps or patterns obtained in previous layers to a known classification.
6. **Soft-max layer:** The soft-max or normalized exponential function is another activation function, which produces a categorical

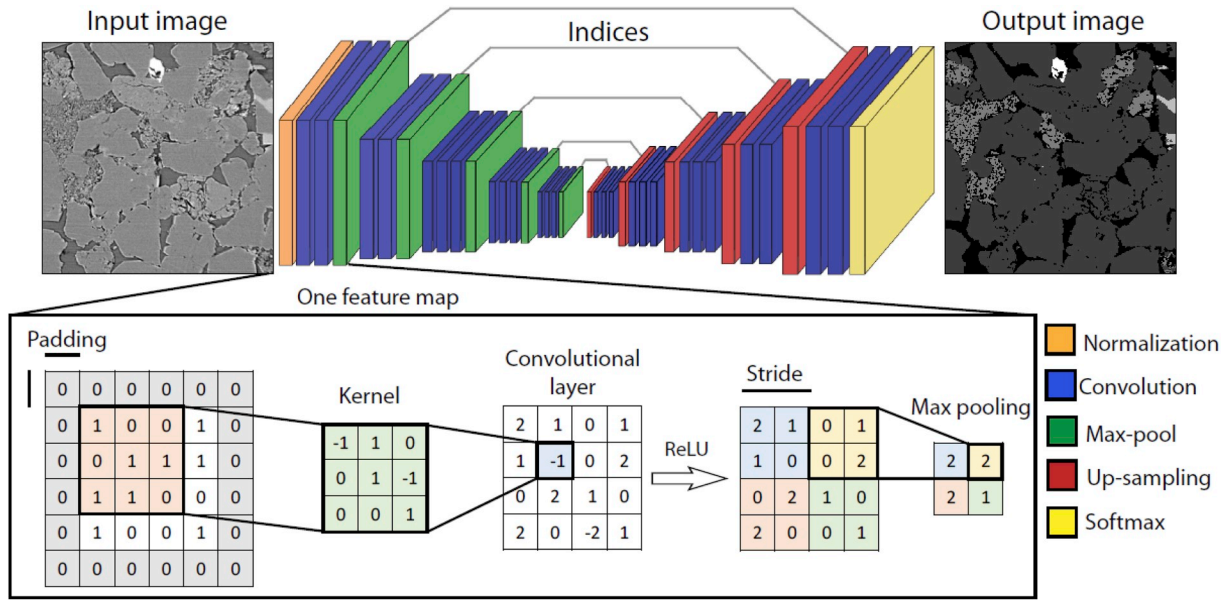


Fig. 1. The general architecture of the SegNet (after Badrinarayanan et al. (2015)).

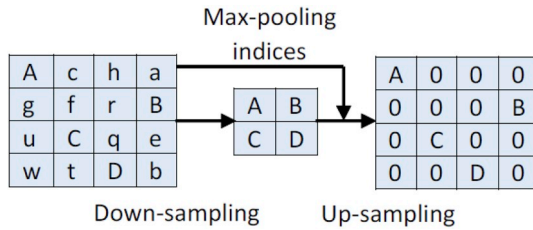


Fig. 2. Max-pooling indices are used to up-sample low-resolution maps in SegNet (after Badrinarayanan et al. (2015)).

probability distribution such that the total sum of the outputs is equal to one. This layer is located in the final layer.

2.2. The SegNet architecture

SegNet is considered as a fully convolutional encoder-decoder network (Badrinarayanan et al., 2015). For the encoder part, it uses a general architecture of CNN and removes the fully connected layers to produce a low-resolution feature map of the input image. Conversely, in the decoder part, a similar architecture is used to reproduce a high-resolution feature map. This map is fed to a multiphase soft-max layer to classify it into a pixel-wise multiphase segmented output; see Fig. 1 (Garcia-Garcia et al., 2017).

Each encoder consists of various convolutional layers responsible for generating feature maps. To introduce non-linearity, a ReLU layer is used as an efficient activation function. Then, a max-pooling window with a non-overlapping stride produces sub-sampled feature maps. After using several steps of convolutional layers, a low-resolution feature map is achieved, which is up-sampled in the decoder part. The problem lies with the decoder learning to deconvolve or decode the low-resolution map. Thus, boundary information of the encoder feature maps is stored before sub-sampling and used in the decoder step. Badrinarayanan et al. (2015) showed that it is sufficient to store only the max-pooling indices. Since the maximum feature values exist in each encoder feature maps. These memorized max-pooling indices are used in the decoder part for up-sampling and producing dense feature maps (Fig. 2).

SegNet architecture is similar to DeconvNet (Noh et al., 2015) and U-Net (Ronneberger et al., 2015), but there are some differences. Comparing to DeconvNet, it uses fully connected layers, thus its

parameters are very large leading to more expensive computation and harder training. On the other hand, U-net transfers the entire encoder feature map to decoder, which requires an extensive memory. There are also some minor architectural differences between the SegNet and U-net, which are found in Badrinarayanan et al. (2015).

In this paper, we use two versions of the SegNet network: 1. Basic SegNet and 2. Extended SegNet. In the basic SegNet, four encoders and four decoders are used. In each encoder or decoder, a $7 \times 7 \times 64$ convolutional/deconvolutional layer with a 2×2 max-pooling/un-pooling window and stride 2 were applied. This means the 64 channels feature map of each internal step is sub-sampled or up-sampled by a factor of two. The ReLU activation function is used only in decoder part. The extended SegNet is similar to the basic one, but with a more massive architecture. In this study, five encoders and five decoders are used, while two convolutional/deconvolutional layers applied in four end-encoders and decoders. Besides, three convolutional/deconvolutional layers are applied in six central part (similar to Fig. 1). Therefore, each basic and extended SegNet contains 18 and 38 layers, respectively.

2.3. Data augmentation

As mentioned earlier, conducting a successful training for the artificial intelligence techniques require a massive number of data (Zhao, 2017). Such a requirement becomes more crucial when the input data are complex, which in this study comprises a set of gray-scale images of the rock sample. Furthermore, depending on the complexity of the problem, the number of input data for training varies significantly and, thus, it requires much cost and time to provide such big data. In this study, however, the necessary data is provided wherein the stochastic algorithm uses the input-image and can produce images with different structures. Therefore, we first describe the implemented stochastic algorithm and, then, other issues related to imposing variability among the training to ensure using a comprehensive dataset will be discussed.

In this study, the cross-correlation simulation algorithm, which belongs to a more inclusive method of the HYPPS technique, is used. The reason for using this algorithm and not the original HYPPS method is discussed later. The implemented method works with an/(a set of) image(s) and produces various equiprobable realizations. In this study, we used the same size as the input image. However, there is no relationship between the input and output images and they can be of any sizes. First, an empty simulation grid is generated for the output image.

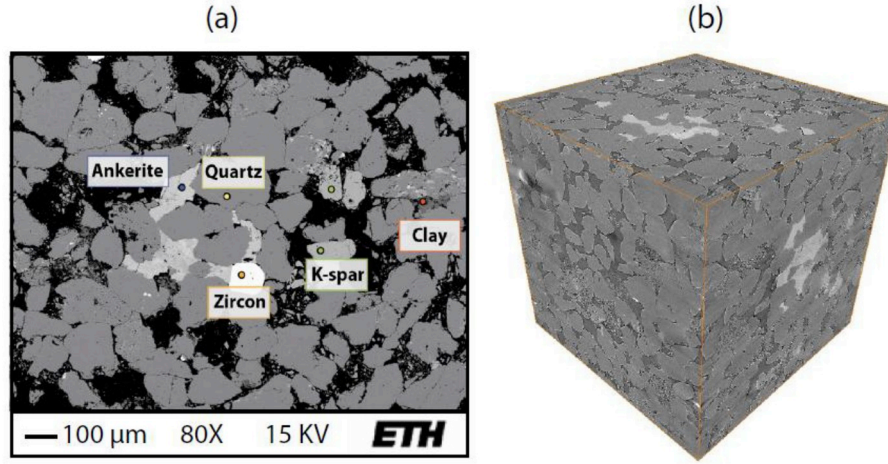


Fig. 3. (a) Different minerals of Berea sandstone detected in a scanning electron microscopy (SEM) (after (Madonna et al., 2013)) and (b) 3D μ CT image of Berea sandstone with a size of $1024 \times 1024 \times 1024$ voxels and a resolution of $0.74 \mu\text{m}$.

Then, starting from a corner of the simulation grid, a pattern with a specific size from the input image is selected and pasted on the simulation grid. Note, the first pattern is inserted on an empty simulation grid and it is selected in a completely random fashion. The size of the selected pattern depends on the heterogeneity of the input image. A larger template size can be used if the input image represents very complex patterns. Similarly, smaller pattern size can be chosen when the input image is homogeneous. However, it should be kept in mind that a larger template often reduces the variability between the produced realizations yet generating high-quality images. On the other hand, smaller template size increases the variability, whereas the final realizations may not be very similar to the input image. Therefore, after defining the appropriate size, a small overlap between the previously inserted patterns OL ($OL_x \times OL_y$) is selected and the cross-correlation of the selected region(s) with the input image is calculated using:

$$\psi(i, j; x, y) = \sum_{x=0}^{OL_x-1} \sum_{y=0}^{OL_y-1} \mathbf{DI}(x+i, y+j) \mathbf{D}_T(x, y), \quad (1)$$

where \mathbf{DI} is the input 2D image, \mathbf{D}_T is the visiting data-event at point (x, y) and T represents its size (i.e. template size). The resulted similarity map is used, and the patterns are sorted based on their similarities and a certain number of the most similar patterns are selected. Finally, one of such patterns is selected and inserted in the visiting point (Tahmasebi, 2017). This process continues until the simulation grid is filled. Some of the produced realization using an input sandstone image are shown in Fig. 5.

The HYPPS algorithm was originally developed to deal with complex secondary continuous and point data (Tahmasebi, 2017). As such, reproducing the conditioning data can be difficult using a sole pattern strategy. Therefore, the HYPPS method offers some flexibilities such as simulating the point-data through a pixel-based method. In this study, however, since none of such data are available, the HYPPS method is used in its pattern-based mode.

One of the main differences between artificial intelligence and physics-based modeling is the amount of the necessary variability. In other words, the deep learning methods can represent their best outcome when the input data show large standard deviation so that the new models can be built from a rich database. To leverage the spatial relationship between the input images and making new patterns, the HYPPS method was modified. Thus, new realizations/images are generated using an ensemble of inputs. In other words, all the available images are searched during the pattern selection phase. Doing so will result in more variability and also producing new transition patterns that might not be available in each of the input images individually.

3. Benchmark data

Andrä et al. (2013) introduced several standard digital rock samples such as Berea and Fontainebleau sandstone and Grosmont carbonate¹. These benchmark samples have been frequently used for DRP studies. Among them, we used the Berea sandstone to evaluate our segmentation results. This sample is mainly composed of quartz and some small minerals such as clay, K-feldspar, ankerite, and zircon; see Fig. 3(a). The acquired image consists of $1024 \times 1024 \times 1024$ voxels with a resolution of $0.74 \mu\text{m}$ (Fig. 3(b)). Andrä et al. (2013) implemented three segmentation methods, namely VSG, SU, and KJ, to obtain a mono-mineral sample. The results indicate that the porosity ranges from 18.4 to 20.9%. This is mostly because even with a bimodal image histogram (i.e. two distinct modes of porosity and mineral), choosing a threshold value strongly affects the estimated porosity.

4. Results

4.1. Semi-automatic segmentation

Fig. 4(a) represents an example of the original grayscale image used in this study. The original image size is 1024×1024 pixels, but to avoid the existing streak artifacts around the central Z-axis, especially in the boundary of the image, an image with 512×512 pixels was selected from the center of the original image. Then, we resampled the image by a factor of 0.5 to produce an image with 256×256 pixels and a resolution of $1.48 \mu\text{m}$, mostly for the sake of smaller computational time.

To obtain a multi-mineral segmented image using several threshold values require addressing two crucial issues. As illustrated in Fig. 4(b): 1. Grain boundaries are brighter than grain surfaces so that they are misclassified, and 2. Different minerals are similarly classified because of their close color values or intensities. Although these minerals have different textures, the above segmentations are insensitive to such features. Based on several trials, the watershed algorithm enhances the segmentation process. This algorithm, however, also failed to differentiate complex structures. Therefore, we decided to manually label each misclassified pixel as an expert supervisor. Fig. 4(c) shows the result of our semi-manually segmentation. Although there are many minor minerals with a similar color in this sample, we decided to categorize all of them as one phase. Therefore, five phases were

¹ These images and their corresponding segmentations are available on: <https://github.com/fkrzikalla/drp-benchmarks>.

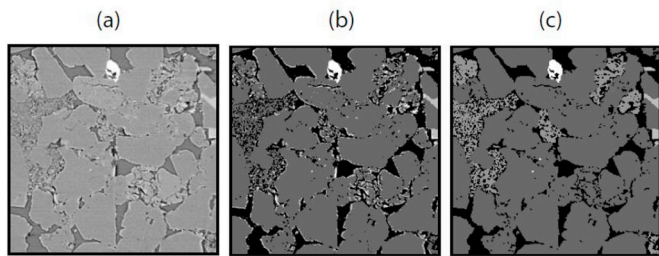


Fig. 4. An example of (a) original image, (b) automatic multi-thresholding segmentation and (c) semi-manually segmentation of Berea sandstone.

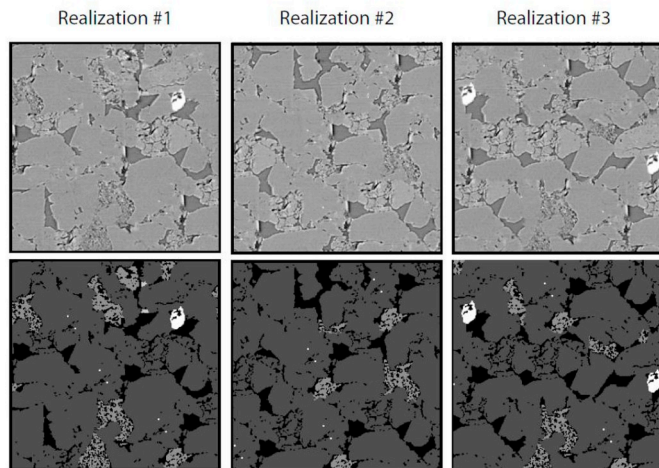


Fig. 5. Three realizations and their corresponding segmented images produced using HYPPS algorithm.

distinguished as pore space (Φ), quartz (Qtz), K-feldspar (K-Fld), zircon (Zrc) and other minerals (i.e. mainly clays). Though semi-manually segmentation is an efficient approach, it requires a long time and can be very limited in practice. Therefore, only 20 images were segmented in this study.

Although there is no straightforward instruction for defining the number of training data, a few thousand images per class are required to properly train a CNN for classification purposes (Ciresan et al., 2012). Thus, the HYPPS algorithm based on data augmentation approach is used to train the SegNet network. This is discussed in detail below.

4.2. Data augmentation

As mentioned, the main obstacle in using the CNN for digital rock images is the limited available data for training. Although image transforms are used as data augmentation method (Garcia-Garcia et al., 2017), our proposed method in Section 2.3 is a more efficient method. Therefore, HYPPS method is applied in this study to enrich the dataset. It has been proven that this algorithm can be considered as an efficient augmentation method to generate as many images as required in the deep learning studies.

To avoid a secondary segmentation, we made changes in the source code to produce the segmented images simultaneously. This means no extra round of segmentation is required and the segmented image is generated all at once when a new realization is produced. The input image size is 256×256 pixels and optimal template and overlap sizes used for the reconstructions are 90×90 and 10×10 respectively. According to the heterogeneity of this sample, five images are considered. An example of such a simultaneous reconstruction is shown in Fig. 5. In this example, we used the gray-scale image shown in Fig. 4(a) and its corresponded segmented image in Fig. 4(c) as inputs, and three realizations, as well as their segmented images, are produced. These results demonstrate how efficient and accurate images are generated in this study.

To avoid overfitting, the original 20 images (Section 4.1) are divided into training, validation and testing groups with a fraction of 80, 10 and 10%, respectively. Then, 16000 stochastic images are generated using the images in the training group. In a similar way, 4000 other images are produced for the validation and testing phases, each with 2000 images. In the next step, the produced images in the training group are used to train the SegNet and the network is adjusted using the existing images in the evaluation dataset. Finally, the designed network is tested using the unseen images in the testing subset.

4.3. Segmentation using SegNet

The basic and extended SegNet are used for digital rock image segmentation. Detailed specifications of these networks are mentioned in Section 2.2. Augmented dataset generated in the previous section is used with these networks. Fig. 6 shows the categorical accuracy and loss values of training and validation procedures along epochs.

According to Fig. 6, the extended SegNet reaches the required accuracy faster. As can be seen, it takes about 100 epochs of training, while the basic SegNet requires 500 epochs (i.e. five times longer) to regularize the convolutional coefficients and reaching the appropriate accuracy (we mean here more than 90%). Both categorical accuracy

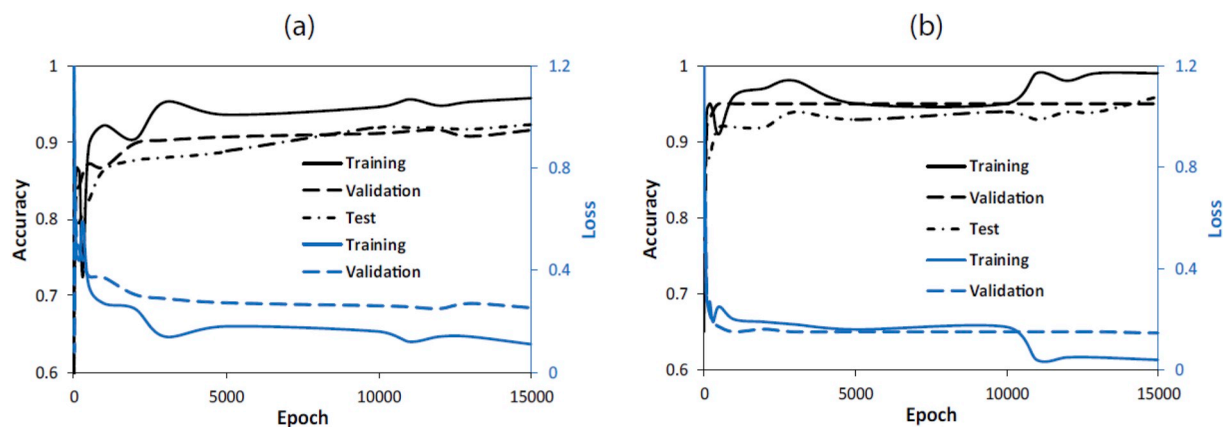


Fig. 6. The categorical accuracy and loss of (a) basic and (b) extended SegNet used in this study.

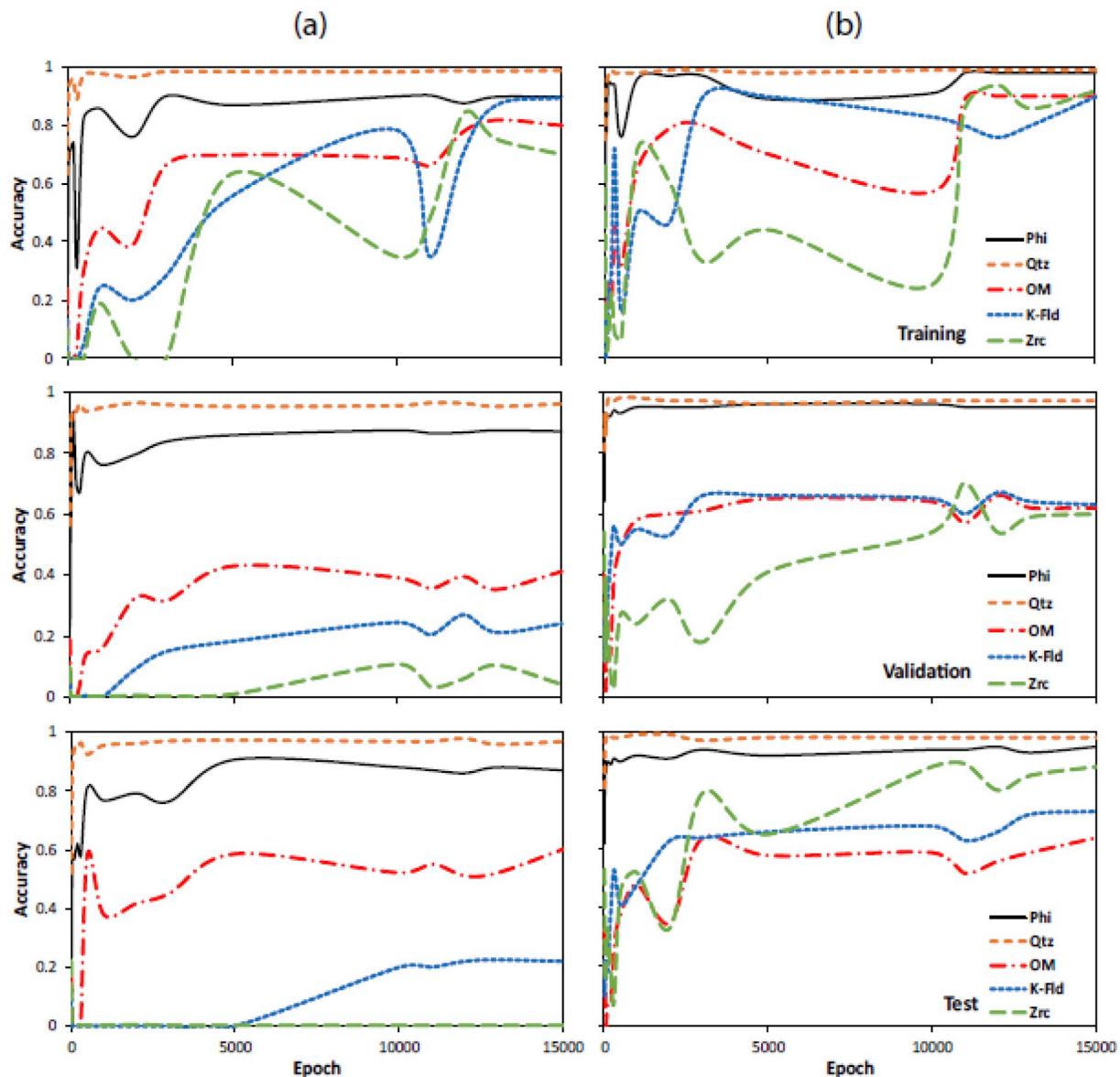


Fig. 7. Categorical accuracy values of (a) the basic and (b) extended SegNet for each phase.

Table 1

A summary of the basic and extended SegNet performances.

Categorical accuracy	Basic			Extended		
	Training	Validation	Test	Training	Validation	Test
Phase #1 (Phi)	0.89	0.87	0.87	0.98	0.95	0.95
Phase #2 (Qtz)	0.98	0.96	0.97	0.99	0.90	0.98
Phase #3 (OM)	0.80	0.41	0.60	0.90	0.62	0.64
Phase #4 (K-Fld)	0.89	0.24	0.22	0.91	0.63	0.73
Phase #5 (Zrc)	0.69	0.04	0.00	0.92	0.60	0.88
Overall	0.96	0.92	0.92	0.99	0.97	0.96

and loss values of two networks for each training, validation and testing phases represent a steady convergence. Moreover, the validation and testing results are more accurate in the extended SegNet, which indicates that this network can produce more valid segmentation results for the new and unseen images. The strength of the extended SegNet is more revealed using the categorical accuracy convergence of each phase. Fig. 7 shows the categorical accuracy values for each phase individually along the epoch for both networks. Table 1 also summarizes

such values for the final epoch.

The categorical accuracy curves and the values presented in Fig. 7 and Table 1 demonstrate that in all cases the two first phases (Phi and Qtz) can reach a reasonable accuracy very fast. Note that these phases comprise the majority part of the utilized images (Fig. 4), which their categorical accuracy values highly affect the overall performance (Fig. 6). In other words, even if other phases are not properly determined, an overall categorical accuracy still more than 90% can be achieved (Table 1). This stresses that one must not only rely on the overall categorical accuracy or loss values as it is mostly controlled by most dominated minerals. Fig. 7(a) shows that categorical accuracy of other phases (OM, K-Fld, and Zrc) in the basic SegNet never reaches a value more than 90%, although its overall value is 96% for the training step in the final epoch (Table 1). In addition, the categorical accuracy of these phases using the validation and test data shows that the basic SegNet can barely detect the minerals. For instance, the Zircon phase is never distinguished from the other existing minerals. The network performances for labeling the unseen images in the testing data are 87, 97, 60, 22 and 0% for Phi, Qtz, OM, K-Fld, and Zrc, respectively, with an overall categorical accuracy of 92%. As can be seen, the results are more promising in the extended SegNet. This can be verified through

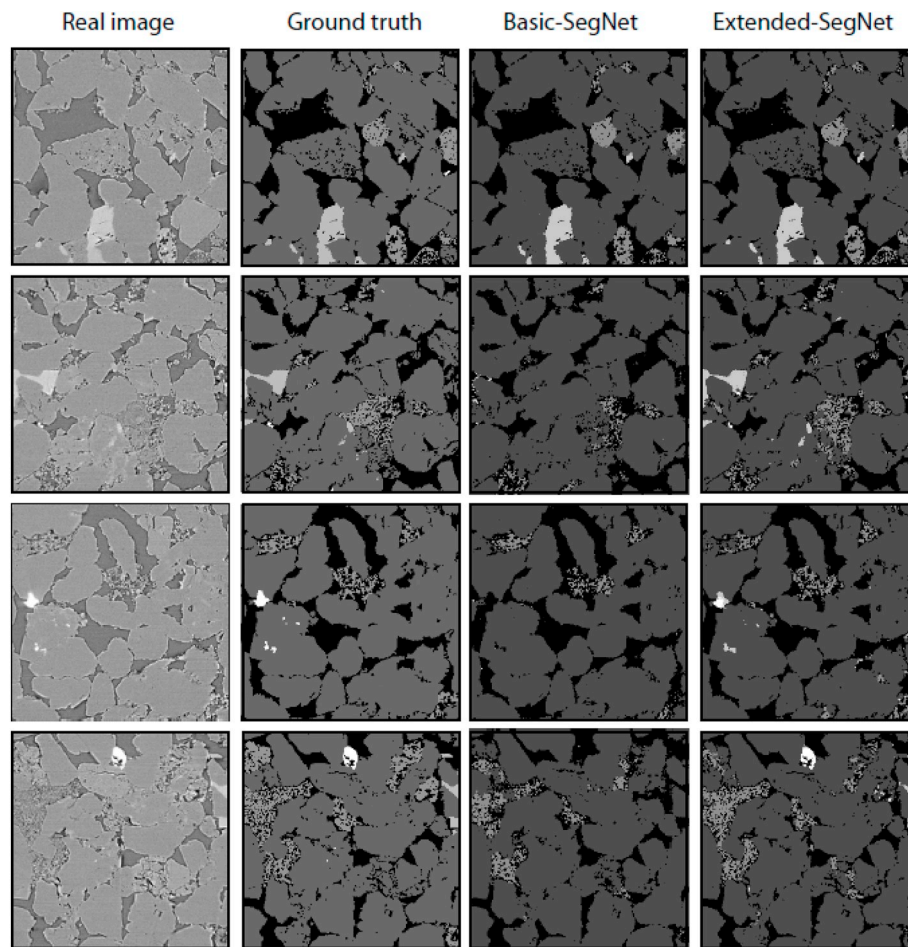


Fig. 8. Four examples of SegNet segmentation.

the categorical accuracy values of all phases after 10000 epochs, which are more than 90% (Fig. 7(b) and Table 1). Although the overall categorical accuracy for the training step in the last epoch is 99%, the network does not perform well for the validation and test images. The extended SegNet, however, segmented the testing dataset with 95, 98, 64, 73 and 88% of categorical accuracy for the Phi, Qtz, OM, K-Fld, and Zrc, respectively, with an overall categorical accuracy of 96%.

According to these results, the extended SegNet is considered being a more reliable network. Fig. 8 illustrates the gray-scale image, the ground truth and the basic and extended SegNet results for four unseen images. It is clear that the extended SegNet has been successful in the training and image segmentation. Whereas, the basic SegNet algorithm introduces a large number of noises in the segmented images.

A key point that suppresses the CNN-based segmentation compared to conventional methods is that these methods are based on small convolutional kernels combined in a deep architecture, which enable them to consider both color and texture simultaneously. To

demonstrate this capability, the results of a multi-segmented image using the conventional and the CNN methods are compared. The multi-thresholding method is done by taking the following steps:

1. Due to the artifact associated with image acquisition and reconstruction, the image was smoothed using a median filter with a size of two pixels.
2. The gradient image was computed to identify the boundary between the minerals. The transition areas between separate phases are those with high gradient magnitude. Such areas are then masked as the transition regions.
3. Except for the masked transition area, multiphase segmentation is implemented by choosing four thresholds manually to achieve five different phases based on color contrasts of minerals.
4. Finally, each phase is extended to the transition area using a watershed algorithm.

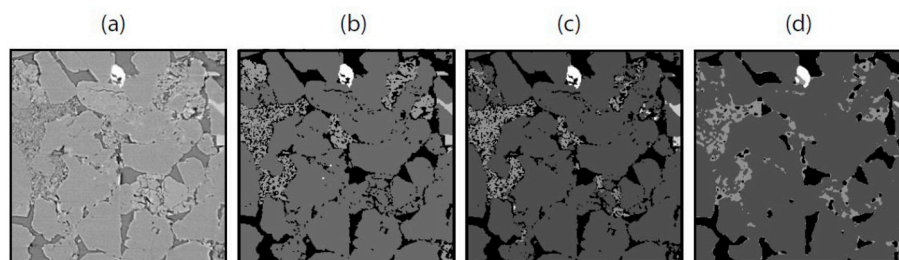


Fig. 9. (a) Real, (b) ground truth, (c) SegNet and (d) multiphase thresholding segmentation.

Fig. 9 shows the results of multiphase segmentation. A visual comparison between the SegNet and conventional methods reveals that the SegNet segmentation is more successful, in particular for identifying the OM phase (clay and other minerals). The color of this phase is very similar to Qtz phase, which makes it difficult to accurately detect them by only using the color. However, SegNet has demonstrated an excellent performance. The implemented kernels in CNN extract the necessary texture and boundary of different components of a rock image. In other words, the CNN methods are not only sensitive to color fluctuation, but they successfully differentiate components with various textures.

According to the results of this study, automatic CNN-based segmentation produces reliable outputs with higher categorical accuracy. This study shows that even with the very small number of images one can train a CNN if an efficient data augmentation method is used. The implemented reconstruction method presented a successful performance in this study by generating a vast number of images quickly (e.g. here about 28,000). It should be noted that the CPU time for producing each realization is less than 15 ms with a Core-i7 CPU and 8 GB RAM. This type of augmentation may seem unnecessary for the common applications of CNN in other fields, but it is inevitable in DRP and maybe other geosciences related applications. This is mainly due to limited access to a large dataset of images in the engineering processes such as DRP and subsurface problems.

5. Conclusion

In this study, we aimed to improve the segmentation process of digital rock images using CNN. Although CNN has been used in a variety of fields from semantic segmentation, autonomous driving, and medical image processing, due to limited available data their applications in DRP have not yet been investigated significantly. Most of the convolutional networks need a large number of images (in order of 10^4). We approached this issue by using a stochastic reconstruction method by which one can generate a large dataset even with a few images.

Manual segmentation is one of the best methods for segmentation of small datasets, but it requires too much of time. Due to using a small number of images, however, we preferred to perform the segmentation manually. In the next step, the segmented images of every realization generated by the utilized stochastic method can be produced simultaneously in the simulation process with the accuracy similar to those prepared by the expert.

Our results showed that the SegNet with a deeper architecture can be trained more effectively by a large dataset and produce more reliable results. The point is that the overall categorical accuracy (or loss) of the network cannot be used to verify the performance of the network in a multiphase segmentation application. Nevertheless, phase-by-phase categorical accuracy should be considered to accurately evaluate the network's performance. Although the overall accuracies of the basic and extended SegNet in our study were close to each other, the phase-by-phase study showed that the extended SegNet is trained more effectively (with a categorical accuracy of 99%). This network also produced valid results for unseen images with a categorical accuracy of about 96%. A comparison with the results produced by multiphase thresholding also revealed that a substantial improvement is achieved when our reconstruction method is used along with the extended SegNet.

The results of this study are not limited to sandstone samples, but the proposed framework can be for any sample. We anticipate the other related fields with a similar obstacle, namely a limited number of observations or input images can benefit from the findings of this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cageo.2019.02.003>.

References

- Andrá, H., Combaret, N., Dvorkin, J., Glatt, E., Han, J., Kabel, M., Keehm, Y., Krzikalla, F., Lee, M., Madonna, C., Marsh, M., Mukerji, T., Saenger, E.H., Sain, R., Saxena, N., Ricker, S., Wiegmann, A., Zhan, X., 2013. Digital rock physics benchmarks-Part I: imaging and segmentation. *Comput. Geosci.* 50, 25–32. <https://doi.org/10.1016/j.cageo.2012.09.005>.
- Badrinarayanan, V., Kendall, A., Cipolla, R., 2015. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation.
- Beucher, S., Meyer, F., 1992. *The Morphological Approach to Segmentation: the Watershed Transformation*, vol. 34. Opt. Eng. YORK-MARCEL DEKKER Inc., pp. 433.
- Cheng, G., Guo, W., 2017. Rock images classification by using deep convolution neural network. *J. Phys. Conf. Ser.* 887, 012089. <https://doi.org/10.1088/1742-6596/887/1/012089>.
- Ciresan, D.C., Meier, U., Schmidhuber, J., 2012. Transfer learning for Latin and Chinese characters with deep neural networks. In: *The 2012 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–6. <https://doi.org/10.1109/IJCNN.2012.6252544>.
- Fattahi, H., Karimpouli, S., 2016. Prediction of porosity and water saturation using pre-stack seismic attributes: a comparison of Bayesian inversion and computational intelligence methods. *Comput. Geosci.* 20, 1075–1094. <https://doi.org/10.1007/s10596-016-9577-0>.
- Ferreira, A., Giraldo, G., 2017. Convolutional Neural Network approaches to granite tiles classification. *Expert Syst. Appl.* 84, 1–11. <https://doi.org/10.1016/j.eswa.2017.04.053>.
- García-García, A., Orts-Escolano, S., Oprea, S., Villena-Martínez, V., García-Rodríguez, J., 2017. A Review on Deep Learning Techniques Applied to Semantic Segmentation. <https://doi.org/10.1007/978-1-4471-4640-7>.
- Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., Larochelle, H., 2017. Brain tumor segmentation with deep neural networks. *Med. Image Anal.* 35, 18–31. <https://doi.org/10.1016/j.media.2016.05.004>.
- Haykin, S.S., 1999. *Neural Networks: a Comprehensive Foundation*. Prentice Hall.
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1026–1034. <https://doi.org/10.1109/ICCV.2015.123>.
- Iassonov, P., Gebrenegus, T., Tuller, M., 2009. Segmentation of X-ray computed tomography images of porous materials: a crucial step for characterization and quantitative analysis of pore structures. *Water Resour. Res.* 45. <https://doi.org/10.1029/2009WR008087>.
- Karimpouli, S., Fattahi, H., 2016. Estimation of P- and S-wave impedances using Bayesian inversion and adaptive neuro-fuzzy inference system from a carbonate reservoir in Iran. *Neural Comput. Appl.* <https://doi.org/10.1007/s00521-016-2636-6>.
- Karimpouli, S., Khoshlesan, S., Saenger, E.H., Koochi, H.H., 2018. Application of alternative digital rock physics methods in a real case study: a challenge between clean and cemented samples. *Geophys. Prospect.* <https://doi.org/10.1111/1365-2478.12611>. Accepted, Published Online.
- Karimpouli, S., Tahmasebi, P., 2016. Conditional reconstruction: an alternative strategy in digital rock physics. *Geophysics* 81, D465–D477. <https://doi.org/10.1190/geo2015-0260.1>.
- Karimpouli, S., Tahmasebi, P., Lamei, R.H., Mostaghimi, P., Saadatfar, M., 2017. Stochastic modeling of coal fracture network by direct use of micro-computed tomography images. *Int. J. Coal Geol.* 179, 153–163. <https://doi.org/10.1016/j.coal.2017.06.002>.
- Kendall, A., Badrinarayanan, V., Cipolla, R., 2015. Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding. <https://doi.org/10.13140/RG.2.1.2985.2407>.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. *ImageNet Classification with Deep Convolutional Neural Networks*.
- Laloy, E., Hérault, R., Lee, J., Jacques, D., Linde, N., 2017. Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network. *Adv. Water Resour.* 110, 387–405. <https://doi.org/10.1016/j.advwatres.2017.09.029>.
- Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. <https://doi.org/10.1109/5.726791>.
- Li, H., Lin, Z., Shen, X., Brandt, J., Hua, G., 2015. A convolutional neural network cascade for face detection. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 5325–5334. <https://doi.org/10.1109/CVPR.2015.7299170>.
- Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. *Med. Image Anal.* 42, 60–88. <https://doi.org/10.1016/j.media.2017.07.005>.
- Madonna, C., Quintal, B., Frehner, M., Almqvist, B.S.G., Tisato, N., Pistone, M., Marone, F., Saenger, E.H., 2013. Synchrotron-based X-ray tomographic microscopy for rock physics investigations. *Geophysics* 78, D53–D64. <https://doi.org/10.1190/geo2012-0113.1>.
- Mosser, L., Dubrue, O., Blunt, M.J., 2017. Reconstruction of Three-Dimensional Porous Media Using Generative Adversarial Neural Networks. <https://doi.org/10.1103/PhysRevE.96.043309>.
- Nanfack, G., Elhassouny, A., Thami, R.O.H., 2017. Squeeze-SegNet: A New Fast Deep Convolutional Neural Network for Semantic Segmentation.
- Noh, H., Hong, S., Han, B., 2015. Learning deconvolution network for semantic

- segmentation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1520–1528. <https://doi.org/10.1109/ICCV.2015.178>.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, Cham, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
- Sezgin, M., 2004. Survey over image thresholding techniques and quantitative performance evaluation. *J. Electron. Imag.* 13, 146–168.
- Shelhamer, E., Long, J., Darrell, T., 2017. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 640–651. <https://doi.org/10.1109/TPAMI.2016.2572683>.
- Srisutthiyakorn, N., 2016. Deep-learning methods for predicting permeability from 2D/3D binary-segmented images. In: SEG Technical Program Expanded Abstracts 2016. Society of Exploration Geophysicists, pp. 3042–3046. <https://doi.org/10.1190/segam2016-13972613.1>.
- Tahmasebi, P., 2017. HYPPS: a hybrid geostatistical modeling algorithm for subsurface modeling. *Water Resour. Res.* 53, 5980–5997. <https://doi.org/10.1002/2017WR021078>.
- Tahmasebi, P., 2018a. Accurate modeling and evaluation of microstructures in complex materials. *Phys. Rev. E* 97, 023307. <https://doi.org/10.1103/PhysRevE.97.023307>.
- Tahmasebi, P., 2018b. Nanoscale and multiresolution models for shale samples. *Fuel* 217, 218–225. <https://doi.org/10.1016/j.fuel.2017.12.107>.
- Tahmasebi, P., 2018c. Packing of discrete and irregular particles. *Comput. Geotech* 100, 52–61. <https://doi.org/10.1016/J.COMPGEO.2018.03.011>.
- Tahmasebi, P., Javadpour, F., Sahimi, M., 2017. Data mining and machine learning for identifying sweet spots in shale reservoirs. *Expert Syst. Appl.* 88, 435–447.
- Tahmasebi, P., Kamrava, S., 2018. Rapid multiscale modeling of flow in porous media. *Phys. Rev. E* 98, 052901. <https://doi.org/10.1103/PhysRevE.98.052901>.
- Tahmasebi, P., Sahimi, M., Shirangi, M.G., 2018. Rapid Learning-Based and Geologically Consistent History Matching. *Transp. Porous Media.* <https://doi.org/10.1007/s11242-018-1005-6>.
- Wallach, I., Dzamba, M., Heifets, A., 2015. AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-Based Drug Discovery.
- Zhang, P.Y., Sun, J.M., Jiang, Y.J., Gao, J.S., 2017a. Deep Learning Method for Lithology Identification from Borehole Images. <https://doi.org/10.3997/2214-4609.201700945>.
- Zhang, Y., Chan, W., Jaitly, N., 2017b. Very deep convolutional networks for end-to-end speech recognition. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 4845–4849. <https://doi.org/10.1109/ICASSP.2017.7953077>.
- Zhao, W., 2017. Research on the deep learning of the small sample data based on transfer learning. In: AIP Conference Proceedings. AIP Publishing LLC, 020018. <https://doi.org/10.1063/1.4992835>.