Digital Rock Reconstruction with User-Defined Properties Using Conditional Generative Adversarial Networks

Qiang Zheng[1] · Dongxiao Zhang[2]

Abstract

Uncertainty is ubiquitous with multiphase flow in subsurface rocks due to their inherent heterogeneity and lack of *in-situ* measurements. To complete uncertainty analysis in a multi-scale manner, it is a prerequisite to provide sufficient rock samples. Even though the advent of digital rock technology offers opportunities to reproduce rocks, it still cannot be utilized to provide massive samples due to its high cost, thus leading to the development of diversified mathematical methods. Among them, two-point statistics (TPS) and multi-point statistics (MPS) are commonly utilized, which feature incorporating low-order and high-order statistical information, respectively. Recently, generative adversarial networks (GANs) are becoming increasingly popular since they can reproduce training images with excellent visual and consequent geologic realism. However, standard GANs can only incorporate information from data, while leaving no interface for user-defined properties, and thus may limit the representativeness of reconstructed samples. In this study, we propose conditional GANs for digital rock reconstruction, aiming to reproduce samples not only similar to the real training data, but also satisfying user-specified properties. In fact, the proposed framework can realize the targets of MPS and TPS simultaneously by incorporating high-order information directly from rock images with the GANs scheme, while preserving low-order counterparts through conditioning. We conduct three reconstruction experiments, and the results demonstrate that rock type, rock porosity, and correlation length can be successfully conditioned to affect the reconstructed rock images. The randomly reconstructed samples with specified rock type, porosity and correlation length will contribute to the subsequent research on pore-scale multiphase flow and uncertainty quantification.

Keywords Digital rock · Three-dimensional reconstruction · Conditional generative adversarial network (cGAN) · User-defined property

✉ Dongxiao Zhang
zhangdx@sustech.edu.cn

[1] Department of Mathematics and Theories, Peng Cheng Laboratory, Shenzhen 518000, People's Republic of China

[2] School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, People's Republic of China

1 Introduction

Due to the scarcity of observations, uncertainty exists in hydrogeological problems and should be considered when modeling flow and transport in subsurface porous media. During the modeling process, some material properties, such as permeability, possessing huge stochasticity due to the heterogeneity of porous media, usually serve as input parameters of the physical model, and thus lead to uncertain model outputs. To evaluate the uncertainty in a systematic and multi-scale manner, some researchers focused on how randomness in the micro pore structure will affect macro permeability and eventually determine model outputs, such as pressure head (Icardi et al. 2016; Wang et al. 2018a; Xue et al. 2019; Zhao et al. 2020). In terms of such uncertainty analysis workflow, the rapid and accurate reconstruction of porous media for random realizations constitutes the initial step and is critically important.

In the last few decades, advances in three-dimensional imaging techniques, e.g., X-ray computed tomography (e.g., Micro-CT and Nano-CT) (Bostanabad et al. 2018; Chen et al. 2013; Li et al. 2018), focused ion beam and scanning electron microscope (FIB-SEM) (Archie et al. 2018; Tahmasebi et al. 2015) and helium-ion-microscope (HIM) for ultra-high-resolution imaging (King Jr et al. 2015; Peng et al. 2015; Wu et al. 2017), have contributed to the development of digital rock technologies, which provide new opportunities to investigate flow and transport in rocks via numerical simulation based on digital representation of rock samples. Constrained by the high cost of three-dimensional imaging, digital rock technologies remain unsuitable for uncertainty analysis because of the lack of random rock samples, thus leading to the development of diversified mathematical methods for random reconstruction of digital rocks. These methods can be divided into the following three groups: object-based methods; process-based techniques; and pixel-based methodologies (Ji et al. 2019). Object-based methods treat pores and grains as a set of objects which are defined according to prior knowledge of the pore structure (Pyrcz and Deutsch 2014). Such methods can be stably implemented, because they have relatively explicit objective functions, so that optimization methods, such as simulated annealing, can converge easily. However, the disadvantage is that they cannot reproduce long-range connectivity of the pore space because they only utilize low-order information. Process-based techniques, on the other hand, can produce more realistic structures than object-based methods through imitating the physical processes that form the rocks (Biswal et al. 2007; Øren and Bakke 2002). This process is relatively time-consuming, however, and necessitates numerous calibrations.

Pixel-based methodologies work on an array of pixels in a regular grid, with the pixels representing geological properties of the rocks. Within these methods, geostatistics is the core technique that contributes to the reconstruction of rock samples, primarily including two-point statistics (TPS) and multi-point statistics (MPS) (Journel and Huijbregts 1978; Kitanidis 1997). One of the most commonly used TPS methods is the Joshi-Quiblier-Adler (JQA) method, which is named according to its three contributors (Adler et al. 1990; Joshi 1974; Jude et al. 2013; Quiblier 1984). The JQA method was designed to truncate the intermediate Gaussian field that satisfies a specified two-point correlation, and finally obtain a binary structure with given porosity and two-point correlation function (Jude et al. 2013). Even though TPS is a very common concept in geostatistics and its related methods are easy to implement, it cannot reproduce the full characteristics of the pore structures because only low-order information is adopted. Under this circumstance, MPS was proposed to address this issue. Unlike TPS, MPS is able to extract local multi-point features

by scanning a training image using a certain template, leading to the incorporation of high-order information, and thus better reproducing performance (Okabe and Blunt 2005). Motivated by multiple applications, MPS methods are flourishing and have diversified variants, mainly comprising single normal equation simulation (SNESIM) (Strebelle 2002), direct sampling method (Mariethoz et al. 2010), and cross-correlation-based simulation (CCSIM) (Tahmasebi and Sahimi 2013). Although MPS features sufficiently capture global information, it is challenged by the conditioning of local information, or equivalently honoring low-order statistics as TPS does. As a consequence, certain fundamental properties, such as volumetric proportion, may not be satisfied even though they should have been honored in the generated samples (Mariethoz and Caers 2014).

In recent years, deep learning has achieved great success in image synthesis, primarily for facial, scenery, and medical imaging (Abdolahnejad and Liu 2020; Choi et al. 2018; Nie et al. 2018). Due to the similar data format and task object, the success of deep learning in image synthesis inspired its application to digital rock reconstruction. The biggest advantage of deep learning is that it constitutes purely data-driven workflow with no prior information needed, and the totally learnable machine avoids complex hand-crafted feature design. Another significant benefit is that once the deep learning model is trained, prediction (generating new structures) can be accomplished within a few moments, which is precisely the bottleneck of traditional methods, such as MPS. Among deep learning methods, generative adversarial networks (GANs) (Goodfellow et al. 2014) have achieved the most popularity in digital rock reconstruction, since they can learn images in a generative and unsupervised manner, and eventually reproduce highly realistic new images that are similar to the training ones.

Concerning GANs-related applications in reconstructing digital rocks, the early work is that Mosser et al. applied GANs to learn three-dimensional images of micro structures with three rock types, i.e., bead pack, Berea sandstone and Ketton limestone, and successfully reconstructed their stochastic samples with morphological and flow characteristics maintained (Mosser et al. 2017, 2018). On top of GANs, Shams et al. (2020) integrated it with auto-encoder networks to produce sandstone samples with multiscale pores, enabling GANs to predict inter-grain pores while auto-encoder networks provide GANs with intra-grain pores. Some other representative applications also exist, such as adopting GANs to reconstruct shale digital cores (Zha et al. 2020), utilizing GANs to augment resolution and recover the texture of micro-CT images of rocks (Wang et al. 2019, 2020), and reconstructing three-dimension structures from two-dimension slices with GANs (Feng et al. 2020; Kench and Cooper 2021; You et al. 2021).

Although GANs are successfully verified to reconstruct several kinds of rocks, the information source for learning may be excessively single, i.e., only the rock images, and prior information about the rocks cannot be incorporated in the current GANs workflow. As a result, the generated samples could be too random and less representative, which may limit their potential for downstream research about pore-scale flow modeling. For instance, it is easy to produce sandstone samples with realistic porosity in previous work, but hardly possible to synthesize plentiful samples with specified porosity or other user-defined properties. In addition, current GANs are strictly developed for reconstructing rocks of a specific type, which means that it is necessary to restart GANs training when a new rock image is prepared for random reconstruction. Such kind of processing will invisibly aggravate computational burden, and thus needs to be improved.

To enable GANs to study images of different rock types simultaneously and enhance the representativeness of generated samples according to user-defined properties, we leverage the GANs in a conditioning manner, which was originally proposed by Mirza and Osindero

(2014). In this study, we adopt progressively growing GANs as a basic architecture (Karras et al. 2017), which has demonstrated excellent performance in image synthesis (Liang et al. 2020; Wang et al. 2018), and has been successfully applied in geological facies modeling (Song et al. 2021). Inspired by the work of generating MNIST digits based on class labels with conditional GANs (Mirza and Osindero 2014), we make the rock type as one of the conditional information, aiming to generate samples with respect to a specified rock type. Moreover, the two basic statistical moments, i.e., porosity and two-point correlations, usually adopted to describe morphological features, also serve as another conditional information to produce samples with user-defined first two moments. Compared with object-based, process-based, pixel-based methods and standard GANs, the conditional GANs in our work can incorporate information not only from the data but also user-defined statistical moments (i.e., the prior information), which cannot be realized by any of the aforementioned approaches. Specifically, the proposed framework can integrate the advantages of commonly used MPS and TPS, since it is able to obtain high-order information from rock images using the GANs scheme and honor low-order moments through the conditioning manner simultaneously. The randomly reconstructed samples with specified low-order moments can bring benefits for the subsequent research on multiphase flow modeling and uncertainty analysis. For example, the moment-based stochastic models can be developed for modeling pore-scale flow and transport (Zhang 2001), and further help characterize the uncertainties of macro properties, e.g., relative permeability. Moreover, 3D printing techniques can be applied to reconstructing realistic samples with specified porosity or correlation length, which can be used to study mechanism of pore-scale flow and transport through experiments.

The remainder of this paper is organized as follows. Section 2 reviews the fundamental concepts and loss functions of the original GANs, and their conditional variants. Section 3 introduces the experimental data and the network architectures used in this study, and then demonstrates the experimental results with three conditioning settings. Finally, conclusions are given in Sect. 4.

2 Methods

2.1 Generative Adversarial Networks

GANs were initially proposed by Goodfellow et al. (2014) to learn an implicit representation of a probability distribution from a given training dataset. Suppose that we have a dataset **y,** which is sampled from P_{data}. Our goal is to build a model, named generator ($\mathcal{G}$), to generate fake samples that nearly support the real distribution P_{data}. The generator is parameterized by $\boldsymbol{\theta}$, and takes the random noise $z \in P_z$ as inputs to generate the fake samples $\mathcal{G}_{\boldsymbol{\theta}}(z)$. In order to discern the real samples from the fake samples, another model, named discriminator ($\mathcal{D}$), is adopted. The discriminator $\mathcal{D}_{\boldsymbol{\phi}}$ with parameters $\boldsymbol{\phi}$ can be viewed as a standard binary classifier, aiming to label the fake samples as zeros and the real samples as ones. Therefore, the generator and the discriminator are placed in a competition, in which the generator wants to deceive the discriminator, while the discriminator attempts to avoid being deceived. When the discriminator exceeds the generator, it will also provide valuable feedback to help improve the generator and, upon reaching the Nash equilibrium point, they can both learn maximum knowledge from the training dataset.

More formally, the two-player game between $\mathcal{G}$ and $\mathcal{D}$ can be recast as a min–max optimization problem, which is defined as follows (Goodfellow et al. 2014):

$$\min_{\theta} \max_{\phi} \mathbb{E}_{y \sim P_{\text{data}}} \left[\log \mathcal{D}_{\phi}(y) \right] + \mathbb{E}_{z \sim P_z} \left[\log \left(1 - \mathcal{D}_{\phi} \left(\mathcal{G}_{\theta}(z) \right) \right) \right]. \tag{1}$$

In practice, $\mathcal{G}$ and $\mathcal{D}$ are trained iteratively with the gradient descent-based method, with the purpose of approximating P_{data} with $P_{\mathcal{G}(\mathbf{z})}$, or equivalently, narrowing the distance between the two distributions. In the original GANs, the distance is measured by Jensen-Shannon (JS) divergence. However, JS divergence is not continuous everywhere with respect to the generator parameters, and thus cannot always supply useful gradients for the generator. Consequently, it is very difficult to train the original GANs stably. To stabilize the training performance, Wasserstein GANs (WGANs) were proposed by switching the JS divergence with Wasserstein-1 distance, which is a continuous function of the generator parameters under a mild constraint (Arjovsky et al. 2017). Specifically, the constraint is Lipschitz continuity imposed on the discriminator, and it is realized by clamping the parameters of the discriminator to a preset range (e.g., $[-0.01, 0.01]$) during the training process. In order to enforce the Lipschitz constraint without imprecisely clipping the parameters of the discriminator, Gulrajani et al. (2017) proposed to constrain the norm of the discriminator's output with respect to its input, leading to a WGAN with gradient penalty (WGAN-GP), whose objective function is defined as:

$$\min_{\phi} \max_{\theta} \mathbb{E}_{z \sim P_z} \left[\mathcal{D}_{\phi} \left(\mathcal{G}_{\theta}(z) \right) \right] - \mathbb{E}_{y \sim P_{\text{data}}} \left[\mathcal{D}_{\phi}(y) \right] + \lambda \mathbb{E}_{\hat{x} \sim P_{\hat{x}}} \left[\left(||\nabla_{\hat{x}} \mathcal{D}(\hat{x})||_2 - 1 \right)^2 \right], \tag{2}$$

where $P_{\hat{x}}$ is defined as sampling uniformly along the straight lines between pairs of points sampled from P_{data} and $P_{\mathcal{G}(\mathbf{z})}$; and λ is the penalty coefficient ($\lambda = 10$ is common in practice).

Apart from stabilizing training in the form of modifying loss function as WGAN-GP did, there is another kind of method that aims to speed up and stabilize the training process via revolutionizing its traditional scheme. Progressively growing GAN (ProGAN), proposed by Karras et al. (2017), is the most representative framework of such kind. The key idea in ProGAN lies in training both the generator and the discriminator progressively, as shown in Fig. 1, i.e., starting from low-resolution data, and progressively adding new layers to learn finer details (high-resolution data) when the training proceeds. This incremental property allows the models to firstly discover large-scale features within the data distribution, and then shift attention to finer-scale details by adding new training layers, instead of having to learn the information from all scales simultaneously. Compared to traditional learning architectures, the progressive training has two main advantages, i.e., more stable training and less training time. In this work, we adopt ProGAN as a basic architecture, and also utilize the loss of WGAN-GP to further ensure stable training.

2.2 Conditional ProGAN

Although GANs are confirmed to work well in reproducing data distributions, or equivalently, generating fake samples that closely approximate the real ones, it is still worth noting that such kind of workflow is totally unconditional, which means that the training data provide the whole information that GANs can learn. However, in many cases, some conditional information needs to be considered for some specified generations, e.g., generating MNIST digits conditioned on class labels or generating face images based on gender.
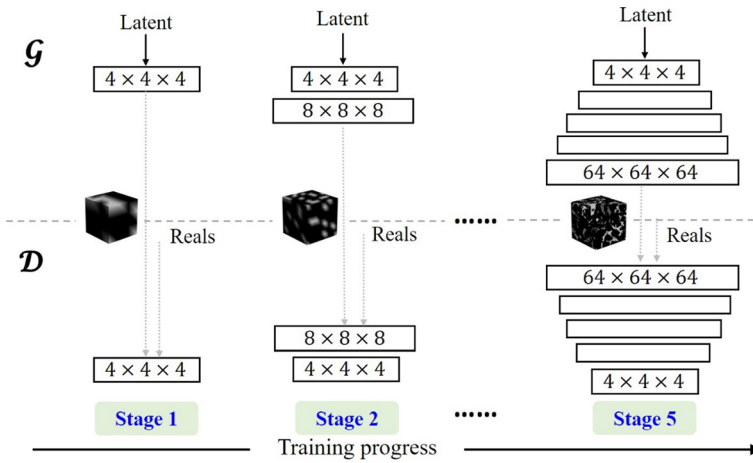
Fig. 1 Schematic of the progressive training process. The original training data in our study are shaped as $64 \times 64 \times 64$ voxels, and they are down-sampled to several low-resolution versions, i.e., from $32 \times 32 \times 32$ voxels to $4 \times 4 \times 4$ voxels with the rate as half of the previous edge length. The training procedure starts from the data with the lowest resolution, i.e., $4 \times 4 \times 4$ voxels, and then goes to the subsequent stage to learn data with higher resolutions through adding new network layers, until the data reach the largest resolution, i.e., $64 \times 64 \times 64$ voxels

Likewise, in digital rock reconstruction, we aim to not only reproduce the training samples, but also allow them to satisfy some user-defined properties, e.g., porosity and two-point correlations, which can be viewed as conditional information in the generating process. Moreover, in previous work, the GANs are always trained on datasets from a single rock type and, as a consequence, the trained GANs can only generate rock samples of this kind. If one wants to obtain samples from n rock types, it is necessary to separate GANs training n times according to previous practice, which means a large amount of computational cost. Motivated by how to generate samples with different rock types using one trained GAN and enable these samples to incorporate user-defined properties, we propose a new Pro-GAN in a conditioning manner for digital rock reconstruction.

Conditional GANs (cGANs) were initially introduced by Mirza and Osindero (2014), aiming to generate MNIST digits conditioned on class labels. The framework of cGANs is highly similar to that of regular GANs. In the simplest form of cGANs, some extra information **c** is concatenated with the input noise, so that the generation process can condition on this information. In a similar manner, the information **c** should also be seen by the discriminator, because it will affect the distance between the real and fake distributions, which is measured by the discriminator. The original cGANs adopted the loss function in the form of JS divergence, which are deficient in training stability, as discussed in the above section. Therefore, in this work, we only borrow the conditioning manner in cGANs, and build a new loss function on top of WGAN-GP loss.

As shown in Fig. 2, the generator of conditional ProGAN takes as input the three-dimensional image-like augmented noises, which are composed of random noise with known distribution and conditional labels, and they are concatenated along the channel axis. Regarding why the input noise should be formulated as image-like data, this is because the generator adopts fully convolutional architectures in order to achieve scalable generations, i.e., generate images with changeable sizes by taking noise with different shapes as input.
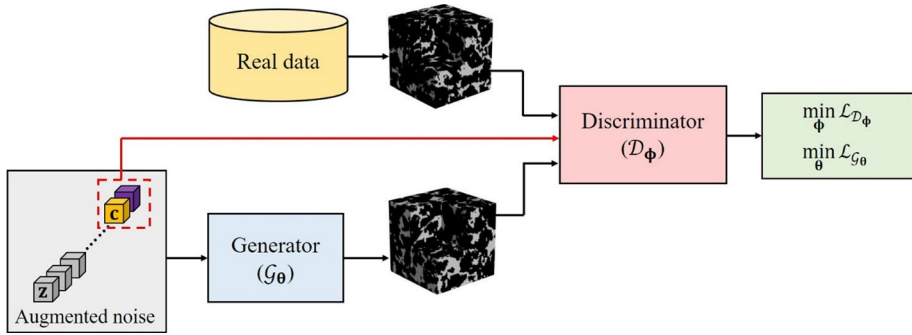
Fig. 2 Schematic of the proposed conditional ProGAN. The input noise to the generator is augmented by conditional labels, i.e., one-hot encoded rock types, porosity, and parameters of two-point correlation functions. The labels should also serve as input of the discriminator by concatenating with the images

Followed by the objective function defined in Eq. (2), the loss function of the generator and the discriminator in conditional ProGAN can be written as follows:

$$
\begin{aligned}
\mathcal{L}_{\mathcal{G}} &= -\mathbb{E}_{z \sim P_z}\big[\mathcal{D}_{\boldsymbol{\phi}}\big(\mathcal{G}_{\boldsymbol{\theta}}(z, \boldsymbol{c}), \boldsymbol{c}\big)\big], \mathcal{L}_{\mathcal{D}} = \mathbb{E}_{z \sim P_z}\big[\mathcal{D}_{\boldsymbol{\phi}}\big(\mathcal{G}_{\boldsymbol{\theta}}(z, \boldsymbol{c}), \boldsymbol{c}\big)\big] \\
&\quad - \mathbb{E}_{\boldsymbol{y} \sim P_{\text{data}}}\big[\mathcal{D}_{\boldsymbol{\phi}}(\boldsymbol{y}, c)\big] + \lambda \mathbb{E}_{\hat{\boldsymbol{x}} \sim P_{\hat{x}}}\Big[\big(||\nabla_{\hat{x}}\mathcal{D}_{\boldsymbol{\phi}}(\hat{\boldsymbol{x}}, \boldsymbol{c})||_2 - 1\big)^2\Big],
\end{aligned}
\tag{3}
$$

where **c** represents the conditional labels containing rock types, porosity, and parameters of two-point correlation functions.

3 Experiments

3.1 Experimental Data

To evaluate the applicability of the proposed conditional ProGAN framework for reconstructing digital rocks, we collect five kinds of segmented rock images, i.e., Berea sandstone, Doddington sandstone, Estaillade carbonate, Ketton carbonate and Sandy multiscale medium, from public datasets, i.e., the Digital Rocks Portal (https://www.digitalrockspor

Table 1 Basic information of five kinds of rock samples

Rock type	Original size	Original resolution	Sample resolution	References
Berea sandstone	$1000 \times 1000 \times 1000$	2.25 μm	9.00 μm	(Neumann et al. 2020)
Doddington sandstone	$700 \times 700 \times 700$	5.40 μm	15.12 μm	(Moon and Andrew 2019)
Estaillade carbonate	$650 \times 650 \times 650$	3.31 μm	8.60 μm	(Muljadi 2015)
Ketton carbonate	$1000 \times 1000 \times 1000$	3.00 μm	12.00 μm	(Raeini et al. 2017)
Sandy multiscale medium	$512 \times 512 \times 512$	3.00 μm	6.14 μm	(Mohammadmoradi 2017)

tal.org/), and their basic information is listed in Table 1. The basic strategy to prepare training samples is to extract three-dimensional subvolumes from the original big sample. In the condition of limited computational resources, the size of training data cannot be very large. Meanwhile, the size of the training sample should meet the requirements of representative elementary volume (REV) to capture relatively global features (Zhang et al. 2000). Consequently, we down-sample the original rock images to 250^3 voxels via spatial interpolations, and then we extract samples of size 64^3 voxels with a spacing of 12 voxels, whose consequent resolutions are also listed in Table 1. Additionally, we conduct an experiment about porosity of extracted samples with respect to their edge lengths, with an aim to determine whether 64 voxels can meet the requirements of REV. As shown in Fig. 3, when the edge lengths are larger than 64 voxels, the curves become much smoother than the previous parts. Even though the Sandy multiscale medium cannot get as smooth as the other four kinds of rocks when the edge length exceeds 64 voxels, it becomes much smoother than itself with smaller edge lengths. Considering that the edge length cannot be very large in order to guarantee a certain amount of extracted samples and save computational resources, and hence the edge length is determined as 64 voxels in this work. Finally, for each kind of rock, we can extract 4096 training samples of size 64^3 voxels with a spacing of 12 voxels between them in the original image. To further amplify the sample size, we rotate the samples by 90° along one axis for two times, and then the total size of each training dataset can be three times of 4096, i.e., 12,288.

3.2 Network Architecture

We built the network architecture based on that of (Karras et al. 2017), and made a few necessary modifications. As shown in Fig. 4a, the generator network consists of five blocks, and they will be added into the network progressively once the training data enter into a higher resolution. For example, only block 1 needs to be used in stage 1 to generate fake samples with size $4 \times 4 \times 4$ voxels. When the training in stage 1 finishes and enters into stage 2, block 2 will be added to the network to produce fake samples with size $8 \times 8 \times 8$ voxels. Likewise, the subsequent stages will be conducted by introducing new blocks. To enable scalable generations, i.e., generate samples with arbitrary sizes, we replace
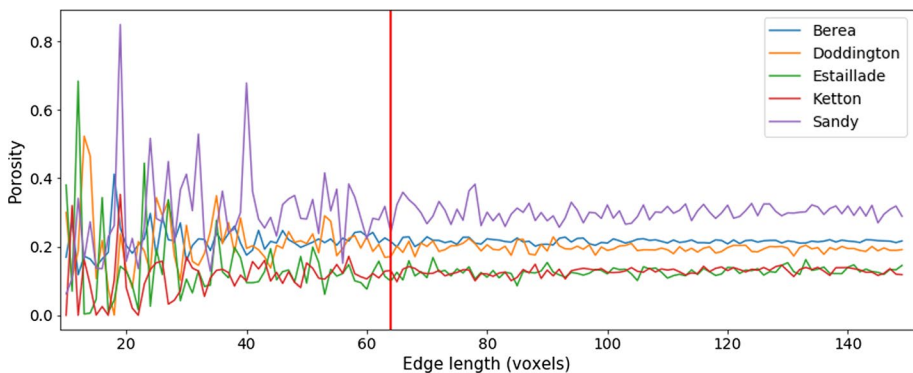

Fig. 3 The porosity of cubic subvolumes with different edge lengths extracted from the original sample with size 250^3 voxels. The vertical red line means that the edge length equals to 64 voxels. For simplicity, we take the first word of names of each rock type to represent them, respectively, in this and subsequent figures
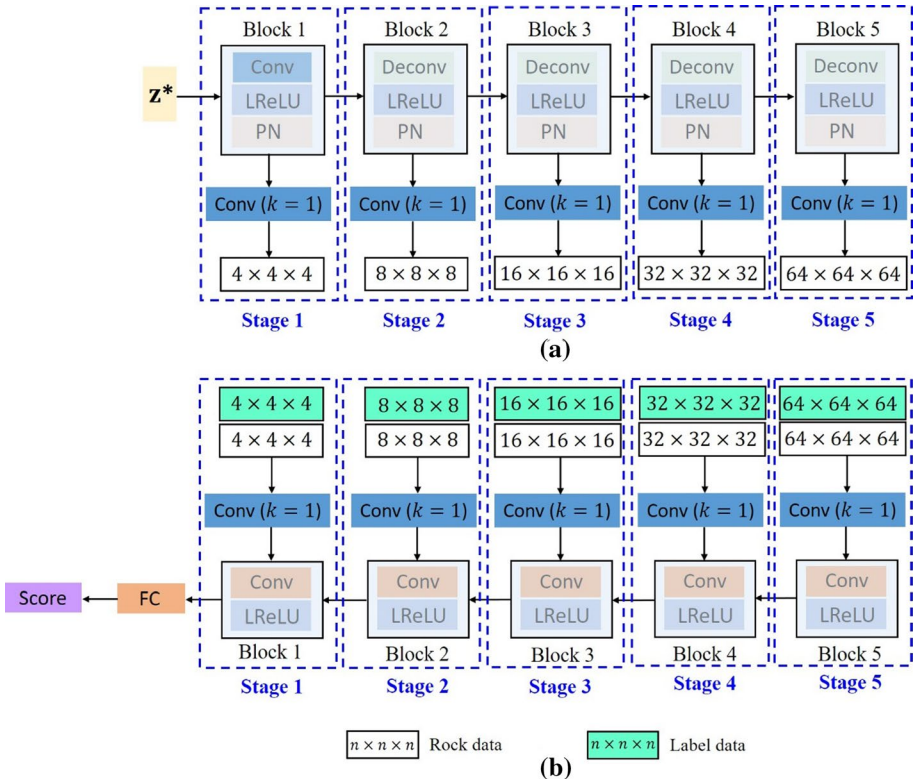
Fig. 4 The network architecture of **a** generator and **b** discriminator. Due to the progressive training scheme, the training process is split into five stages, and they are paired for the generator and the discriminator, which means that the output of the generator should share the same size as the input of the discriminator in each training stage. When entering new stages, new blocks will be added progressively to handle higher-resolution data. The input tensor z* for the generator, augmented by random noise and labels, firstly goes through a convolutional layer (Conv), a Leaky ReLU activation layer (LReLU), and a pixel-wise normalization layer (PN) in block 1. The subsequent blocks are the same as block 1 except for the deconvolutional layer (Deconv), which is used for enlarging the size of feature maps. In each stage, the output of the final block will go through a Conv layer with kernel and stride size as 1 to reduce data channels and obtain fake data with the same shape as training data. The discriminator takes multi-resolution data and labels as input, which share the same size and concatenate along the channel axis. The input data firstly go through a Conv layer with kernel and stride size as 1 to increase channels, and then pass some blocks, which are all composed of a Conv layer (with cubic kernel size as 3 and strides as 2) and a LReLU activation layer. Finally, a fully-connected layer (FC) is used to transform the outputs of block 1 to one-dimensional scores

fully-connected layers in original block 1 with convolutional layers, whose kernel size and stride are both $1 \times 1 \times 1$ voxel. Consequently, the inputs should be 5D tensors to meet the demands of convolutional operations, and here we set it as shape $N \times C \times 4 \times 4 \times 4$, in which N represents batch size and C means channels. In block 1, leaky ReLU activation function and pixel-wise normalization are added after the convolutional layer. To feed the discriminator single-channel data with resolution $4 \times 4 \times 4$ voxels, a convolutional layer, which adopts cubic kernel and stride with edge length as 1 voxel, is added to reduce the channels of outputs of block 1 while preserving their size unchanged. After stage 1 training, the subsequent blocks are the same in network layers, and the only difference from

block 1 is that all of them adopt deconvolutional layers (with kernel size $3 \times 3 \times 3$ and strides $2 \times 2 \times 2$) to enlarge the size of feature maps by two times of the previous ones. The same as in stage 1, there is also a convolutional layer being arranged after each block to decrease the channels of their outputs. The input z^* to the generator is an augmented vector, which is obtained by the random noises z concatenated by the conditional labels c along the channel axis, i.e., $z^* = z \oplus c$, where $\oplus$ represents the concatenation operation.

The discriminator network (see Fig. 4b) is almost symmetric to the generator network with several exceptions. Likewise, the discriminator has the same number of blocks as the generator since they should undergo the paired training stages, which means that the outputs of the generator and the inputs of the discriminator must have the same resolution in each training stage. The input Y^* to the discriminator is also an augmented vector like that to the generator, and the conditional labels c should be reformulated as 5D tensors with the same width as training images Y so that they can be concatenated along the channel axis, i.e., $Y^* = Y \oplus c$. In each stage, the training data will firstly go through a convolutional layer, whose cubic kernel and stride have edge length as one voxel, to enlarge channels while maintaining the size unchanged. Subsequently, the outputs will undergo several blocks, which all contain a convolutional layer (with kernel size $3 \times 3 \times 3$ and strides $2 \times 2 \times 2$) and leaky ReLU activation. Finally, the outputs of block 1 will be transformed to one-dimensional scores by a fully-connected layer. It is worth noting that, when entering new training stages, the newly added blocks to the generator and the discriminator will fade in smoothly, with an aim to avoid sudden shocks to those already well-trained and smaller-resolution blocks. For additional details about the network architecture and training execution of ProGAN, one can refer to (Karras et al. 2017). In this work, we almost follow its network hyper-parameters settings, e.g., activation functions and number of convolutional kernels, since they have been verified to work well in diversified image synthesis tasks. Moreover, it is also promising to tune the hyper-parameters with the aim to balance the efficiency and accuracy and enable key physical constraints to be satisfied (Zhang et al. 2020a; b).

3.3 Experimental results

3.3.1 Reconstruction Conditioned on Rock Type

Based on the training samples of five rock types prepared above, we train the conditional ProGAN networks using the Adam optimizer (Kingma and Ba 2014). The learning rate (lr) is set based on the current training stage or data resolution, and here we set $lr = 5e - 3$ when the resolution is less than or equal to 16^3 voxels, $lr = 3.5e - 3$ for resolution as 32^3 voxels, and $lr = 2.5e - 3$ for resolution as 64^3 voxels. In this section, we consider the rock type as conditional labels, aiming to study several kinds of rock data simultaneously with one model rather than wasting time to build several models. The conditional rock type are discretized labels, and thus should be one-hot encoded so as to concatenate with random noise. As mentioned in Sect. 3.2, we set the inputs of generator (z^*) as shape $N \times C \times 4 \times 4 \times 4$, and thus the random noise z sampled from standard Gaussian distribution P_z has shape $N \times 1 \times 4 \times 4 \times 4$. To concatenate with image-like noise along the channel axis, the original label with size $N \times 5$ should be reshaped and repeated as a tensor with size $N \times 5 \times 4 \times 4 \times 4$. Likewise, when feeding the discriminator, the labels should also be reshaped and repeated so as to concatenate with multi-resolution rock data. Therefore, in this case, the channel C of inputs to the generator and

the discriminator is 6. The batch size N is chosen as 32 for each training stage in this work. In each stage, we set 320,000 iterations of alternative training when data resolution is less than or equal to 16^3 voxels, and 640,000 iterations for all larger resolutions. We train the networks for 2,880,000 iterations in total, which requires approximately 23 h running time on four GPUs (Tesla V100).

With the trained model, we can produce new realizations by feeding the generator with new random noise sampled from P_z and the specified rock type. Here, in its simplest form, we assign the one-hot code [1, 0, 0, 0, 0] as Berea sandstone, [0, 1, 0, 0, 0] as Doddington sandstone, and so on in a similar fashion for other rocks. Since we utilized a fully convolutional architecture to build a scalable generator, we can produce new samples with different sizes by feeding the generator with noise of different shapes. During the model training, the input z^* is shaped as cubic image-like data with edge length as 4 voxels, and the output is also cubic data with side length as 64 voxels. To make the generated cubic samples become larger, we increase edge length of z^* as 6, 8, and 10 voxels to obtain the corresponding outputs with edge length as 96, 128, and 160 voxels. We randomly select one sample for each kind of rock to visualize the reconstruction performance. As shown in Fig. 5, the left column represents training samples, while the others are generated samples with different sizes. Firstly, the reconstructed rocks of different kinds are distinguishable, but visually similar to their own kinds, which means that the conditioning of rock type does work. Secondly, the scalable generator can effectively produce samples with larger sizes than those of training ones without losing morphological realism. It is worth noting that, owing to the great power of the progressive training scheme, the multiscale features can be effectively incorporated into the model, which endows the model with enormous potential to produce structures with dual porosities, such as Sandy multiscale medium.

Apart from visual realism, we need to quantitatively evaluate the reconstruction performance by computing statistical similarities between training samples and generated ones. In the original ProGAN work, Karras et al. (2017) asserted that a good generator should produce samples with local image structures similar to the training set over all scales, and they proposed multi-scale sliced Wasserstein distance (SWD) to evaluate the multi-scale similarities. The multi-scale SWD metric is calculated based on the local image patches drawn from Laplacian pyramid representations of generated and training images, starting at a low resolution of 16^3 voxels and doubling it until reaching the full resolution. Here, we sample 4000 training and generated images, and randomly extract 32 7×7-pixel slice patches from the Laplacian pyramid representation of each image at each resolution level, to prepare 128,000 patches from the training and generated dataset at each level. Since each saved model during the training process can be utilized to calculate multi-scale SWD, we can obtain the changes of SWD with respect to the iteration steps at each level. In this work, we average the multi-scale SWD over different levels to acquire a mean value to evaluate the distance between two distributions. Meanwhile, we also calculate the average SWD at the highest level of real samples as a benchmark, i.e., similarities between real patches, and their values are 7.23×10^3, 7.41×10^3, 7.22×10^3, 7.31×10^3, and 7.50×10^3 for Berea sandstone, Doddington sandstone, Estaillade carbonate, Ketton carbonate, and Sandy multiscale medium, respectively. As shown in Fig. 6, the average SWD curves of generated samples from the five rocks can converge to a relatively small value, and are extremely close to the above benchmark values, which means that when we stop training, the generator can produce very realistic samples.
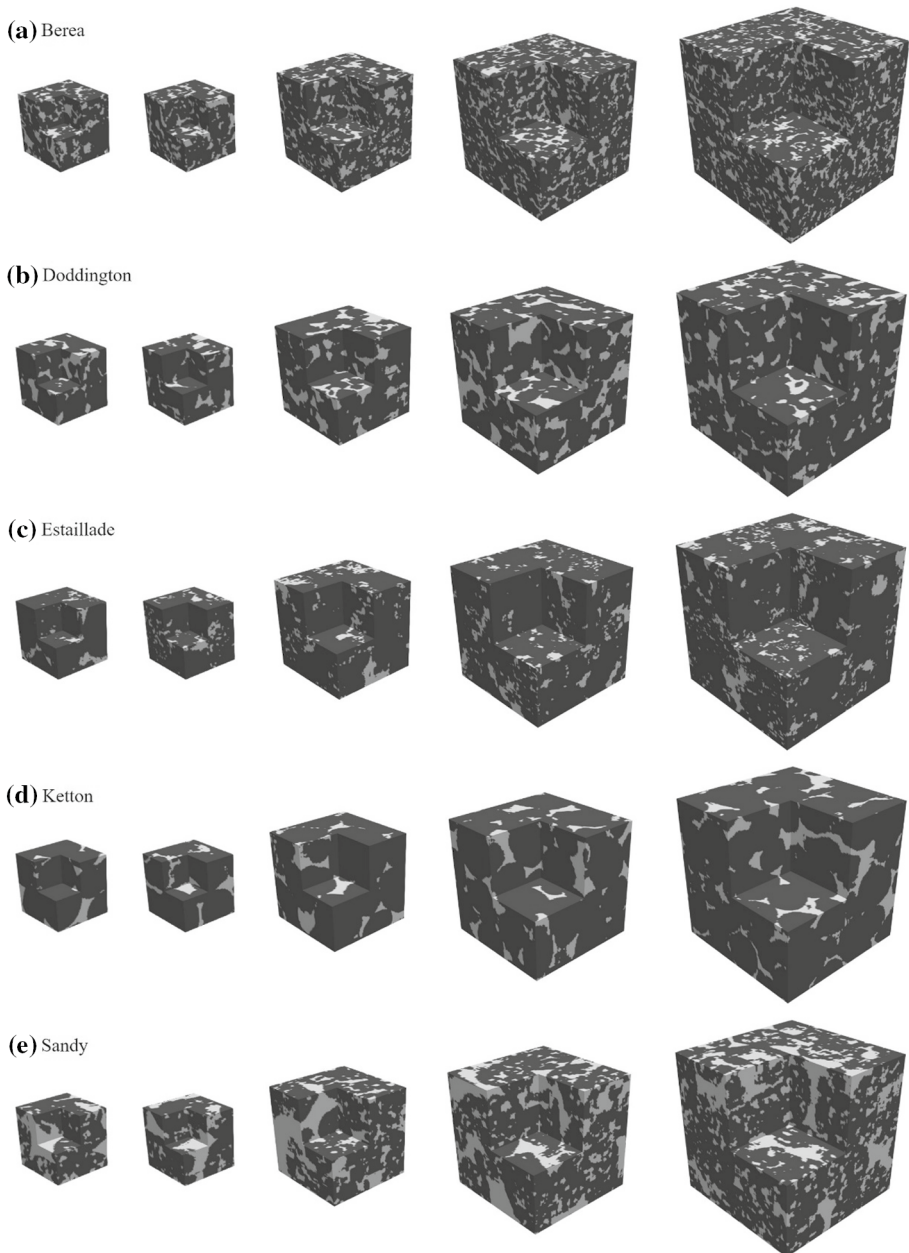
Fig. 5 Training samples (the first column in each subfigure) and synthetic samples (the subsequent four columns in each subfigure) with different sizes of five rock types: **a** Berea sandstone; **b** Doddington sandstone; **c** Estaillade carbonate; **d** Ketton carbonate; and **e** Sandy multiscale medium
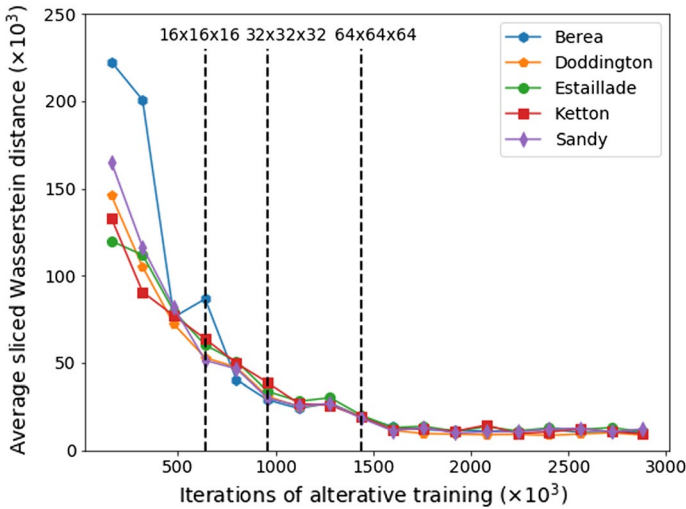
Fig. 6 Average sliced Wasserstein distance (SWD) of five rock types during the training process

In addition to using SWD to assess the generative performance, we also conduct geo-statistical analysis to further verify the morphological realism of reconstructed samples. The rock media in this experiment can be defined as a binary field $F(x)$ as follows:

$$F(x) = \begin{cases} 1 & x \in \Omega_{\text{pore}} \\ 0 & x \in \Omega_{\text{solid}} \end{cases}, \tag{4}$$

where x represents any point in the image of porous media; and Ω_{pore} and Ω_{solid} are the space occupied by the pores and solid grains, respectively. In order to characterize the structures of rocks, the first- and second-order moments of $F(x)$, defined as follows, can be utilized:

$$\phi_F = \overline{F(x)}, \tag{5}$$

$$R_F(x, x + r) = \frac{\overline{\left[\phi_F - F(x)\right] \cdot \left[\phi_F - F(x + r)\right]}}{\phi_F - \phi_F^2}, \tag{6}$$

where ϕ_F represents the porosity; and $R_F(x, x + r)$, termed the normalized two-point cor-relation function, represents the probability that two points x and $x + r$, separated by lag vector $\mathbf{r}$, are located in the pore phase Ω_{pore}. In addition, we also calculate specific surface area to measure morphological similarities, which is expressed as:

$$S_a = \frac{1}{V} \int dS, \tag{7}$$

where integration occurs at the solid-pore interface S; and V is bulk volume.

In order to calculate the above three metrics, we randomly select 200 samples from the training dataset with respect to each rock type, and simultaneously generate new realiza-tions with the same amount. As shown in Fig. 7, the porosity ranges of generated samples of size 64^3 voxels agree well with those of training ones for all rock types. When enlarging
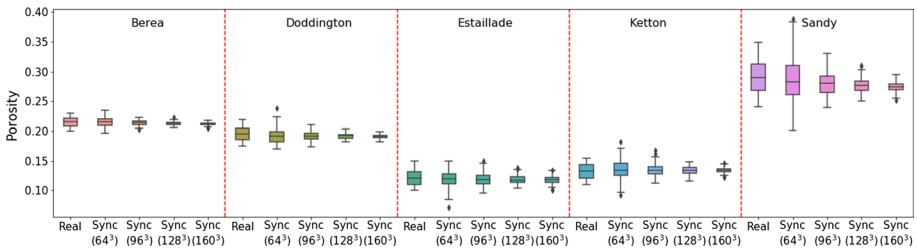
Fig. 7 Comparisons of the porosity of real training samples and synthetic (Sync) samples with different sizes of five rock types

the size of generated samples, their porosity ranges become narrower while keeping median values close to the real ones. This can be easily understood from the perspective of stochastic process, in which the random field with larger scale is more closely approaching ergodicity, and thus its statistical properties, such as porosity, should be more stable than that with a smaller size (Papoulis and Pillai 2002; Zhang 2001). The normalized two-point correlation curves in three dimensions for each kind of rock are presented in Fig. 8. Obviously, both the means and the ranges of the two-point correlation functions of generated samples can match well with those of the original training ones. When samples become larger, the two-point correlation functions have narrower ranges and longer tails, and keep the mean curves tightly close to the real ones. This means that the second-order structure of rock media can be maintained when we produce larger samples than the training ones. The comparisons of specific surface area are shown in Fig. 9. Likewise, this metric also matches well between the training samples and the synthetic ones, and approaches median values when the sample size becomes larger.

To evaluate the physical accuracy of reconstructed samples, we calculate their absolute permeability using the single-phase Lattice Boltzmann method (Eshghinejadfard et al. 2016). In this case, we ignore the anisotropy of each kind of rock and plot the results in Fig. 10. It is obvious that the permeability of generated samples matches well with that of training ones, and also gets closer to median values when enlarging sample size. Such a trend is quite similar to that of above three geometric indicators, since they are totally dependent on the structure, which can be viewed from the perspective of stochastic process. Furthermore, with sample size increasing, the abnormal values of permeability become less and less. This means that the larger are the samples, the more probable it is for them to achieve realistic and stable morphological and physical properties, which is crucial for downstream research of pore-scale flow based on the reconstructed samples.

3.3.2 Reconstruction Conditioned on Porosity

After validating that the conditioning manner worked successfully on rock type, we continue to consider the first-order moment, i.e., the porosity, as another kind of conditional label with an aim to generate random realizations with a specified porosity. During the model training, the scalar porosity labels should be reshaped and repeated as the same dimension with inputs for the generator and the discriminator, respectively, just as it was done in the last section for rock type. When training the discriminator, i.e., minimizing $\mathcal{L}_{\mathcal{D}}$ in Eq. (3), the conditional porosity should be the corresponding values of the training samples. Meanwhile, when training the generator, i.e., minimizing $\mathcal{L}_{\mathcal{G}}$ in Eq. (3), we hope that
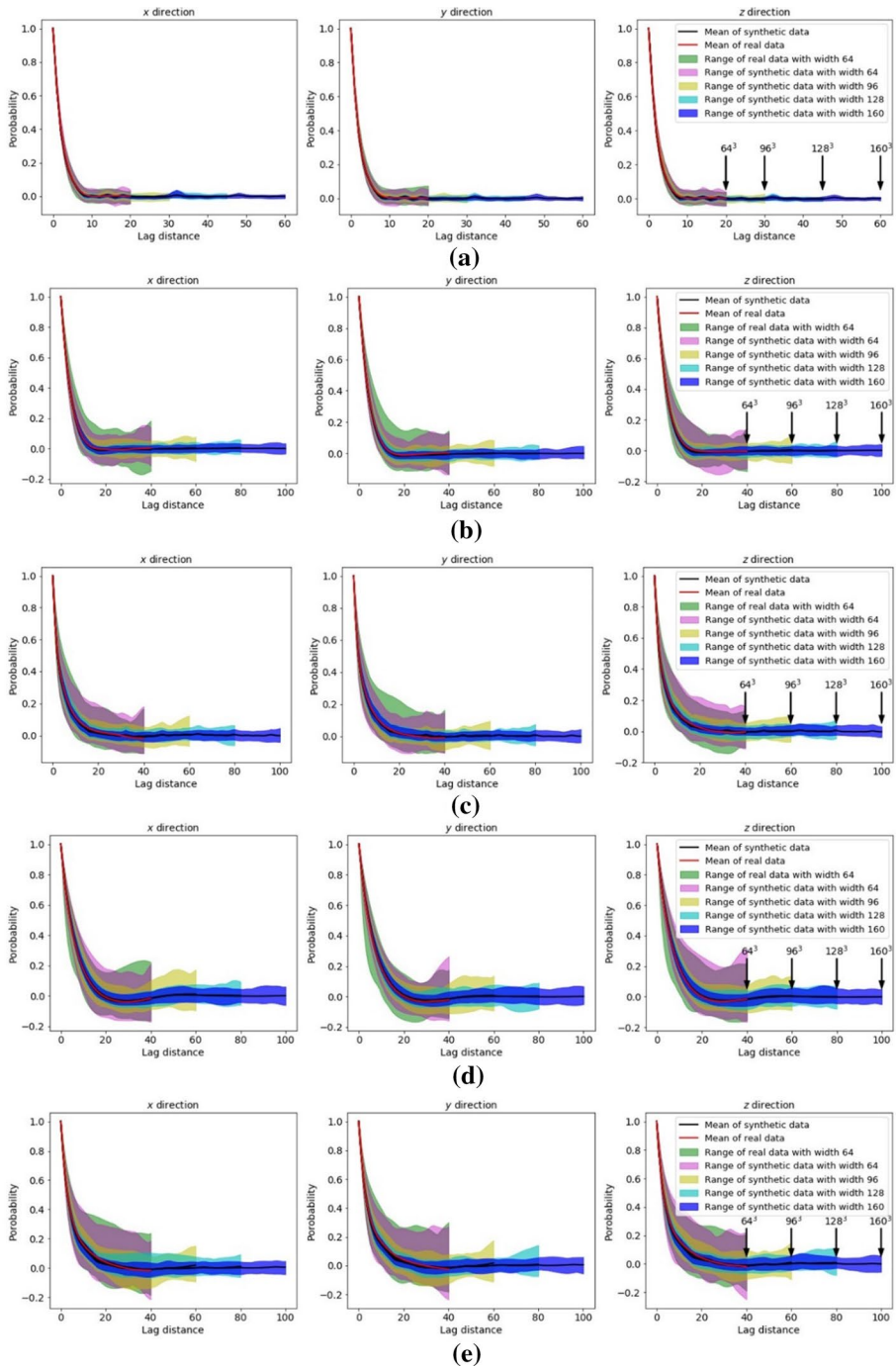
Fig. 8 Comparisons of the normalized two-point correlation function of training samples and synthetic samples with different sizes of five rock types: **a** Berea sandstone; **b** Doddington sandstone; **c** Estaillade carbonate; **d** Ketton carbonate; and **e** Sandy multiscale medium
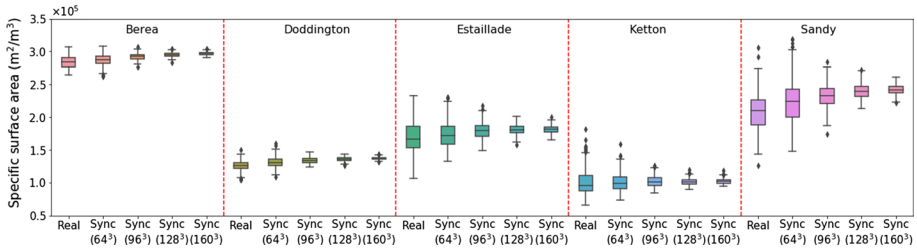
Fig. 9 Comparisons of the specific surface area of training samples and synthetic samples with different sizes of five rock types
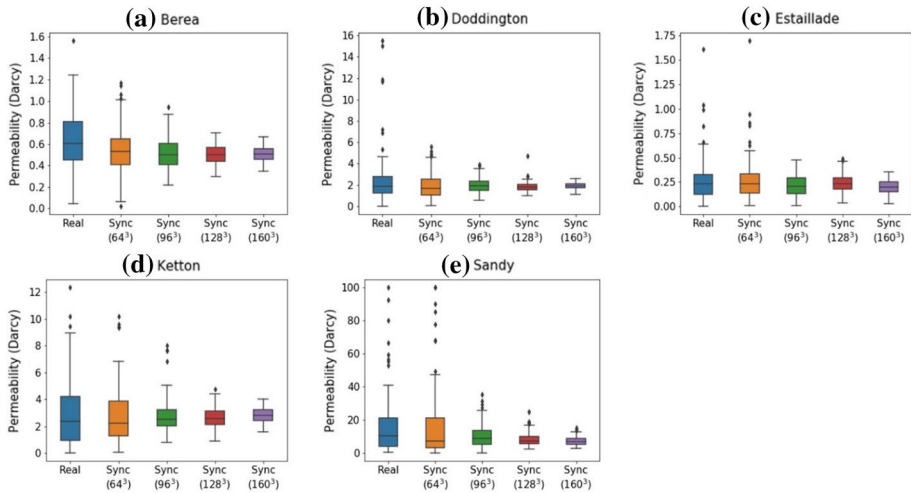

Fig. 10 Comparisons of the permeability of real training samples and synthetic samples with different sizes of five rock types

the generator can see reasonable porosities with equal possibilities, and thus we extract the maximum and minimum values of porosity from the training dataset of each kind of rock, and then uniformly sample from that range to feed the generator. In this case, we select three rock types, i.e., Doddington sandstone, Estaillade carbonate and Sandy multiscale medium, to test the performance of porosity conditioning.

With the trained model, we can produce random rock realizations with specified porosity by feeding the generator new random noise $z \sim P_z$ after concatenating it with a given porosity label. In this case, we set different porosity targets with respect to different rock types, i.e., $\phi_{target} = 0.21$ for Doddington sandstone, $\phi_{target} = 0.10$ for Estaillade carbonate, and $\phi_{target} = 0.22$ for Sandy multiscale medium. The target values are determined according to the porosity distribution of the real dataset. Specifically, they are randomly selected from two interval ranges, i.e., one range is between the upper quartile and the maximum and the other one is between the lower quartile and the minimum. These ranges are chosen because they are relatively more difficult to access than those around median values, with the aim to show conditioning performance as fully as possible. To validate the performance of conditioning on porosity targets, we produce
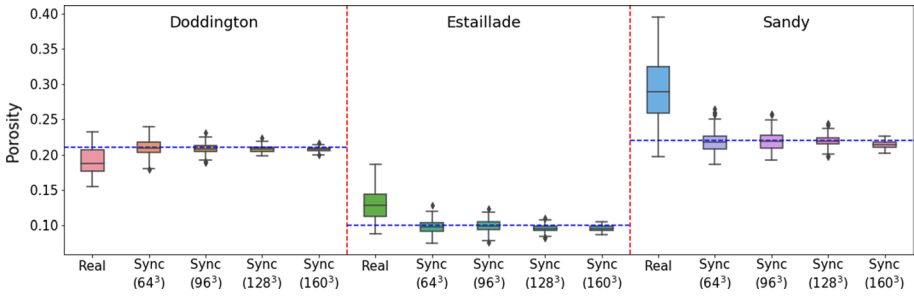
Fig. 11 Porosity of the training samples and generated samples with different sizes for three rock types. The blue dashed line represents the preset porosity targets

200 samples with different sizes for each rock type, and also randomly select training samples with the same amount, to calculate their porosities. It can be seen from Fig. 11 that the porosities of the generated samples with different sizes all exhibit a global offset from those of the training ones to the preset targets, and their median values are rather close to the targets. Furthermore, when the samples get larger, their porosities
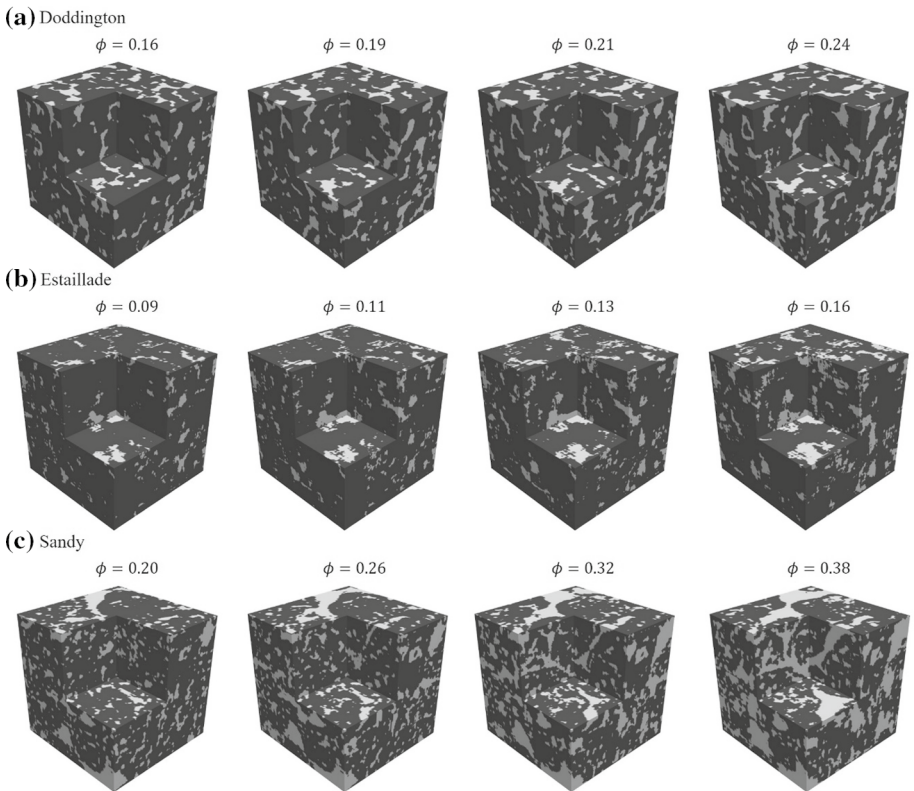
Fig. 12 Synthetic samples of varying porosity with size 160^3 voxels. The three rows represent Doddington sandstone, Estaillade carbonate, and Sandy multiscale medium, respectively

will approach targets in a more determinate manner. In other words, the large samples are more reliable when we want to reconstruct rock samples with a specified porosity.

To further vividly elucidate the performance of porosity conditioning, we fix the input noise and only change the porosity for each rock type when feeding the generator to produce samples with size 160^3 voxels. As shown in Fig. 12, almost all pores become larger when gradually increasing the conditional porosity. In contrast, the patterns of the pore structure remain almost unchanged from left to right in the figure due to the fixed noise. This phenomenon reveals that the conditional information could be disentangled from the random noise, and they respectively control different aspects in the reconstructed samples, with pore structure generally encoded in the input noise and pore size controlled by the conditional label. It is interesting to find that the changes in Sandy multiscale medium mainly lie in the inter-grain pores, whose sizes grow dramatically when increasing the global porosity, while keeping the inner-grain pores less changed.

Apart from evaluating conditioning performance in terms of porosity, we also investigate the effects of changing porosity on physical property. Based on the same samples used in Fig. 11, we calculate the corresponding absolute permeability via the Lattice Boltzmann method. As shown in Fig. 13, the permeability of the generated samples of Doddington sandstone exhibits a slight increase compared to that of training samples, since its porosity target is located above the upper quartile and larger than that of major samples. Estaillade carbonate presents a similar, but opposite, trend. Actually, the permeability of these two rocks did not respond sensitively to the porosity changes. In contrast, the permeability of Sandy multiscale medium presents remarkable decreases when setting the porosity target below the lower quartile. To more quantatively clarify the markedly different responses in permeability, we calculate the relative deviations of porosity targets with respect to the median values of the real dataset for each rock type, and also compute that between two permeabilities, i.e., the median values of permeabilities of generated samples with size 160^3 voxels and that of the real dataset. The results demonstrate that the relative deviations for porosity targets are 11.8%, -21.6%, and -23.8% for Doddington sandstone, Estaillade carbonate, and Sandy multiscale medium, respectively. The minus deviations mean that the targets are smaller than the median values of real porosities, and vice versa. The corresponding relative deviations for permeabilities are 21.2%, -18.1%, and -97.4%. It is obvious that, for Sandy multiscale medium, even though the porosity deviation is slightly larger than that of the other two types, the permeability deviation is quite larger than that of the other two. Regarding the reason for this, we guess that the inter-grain pores, which affect discharge capacity to a large extent, shrink dramatically when decreasing the porosity, as
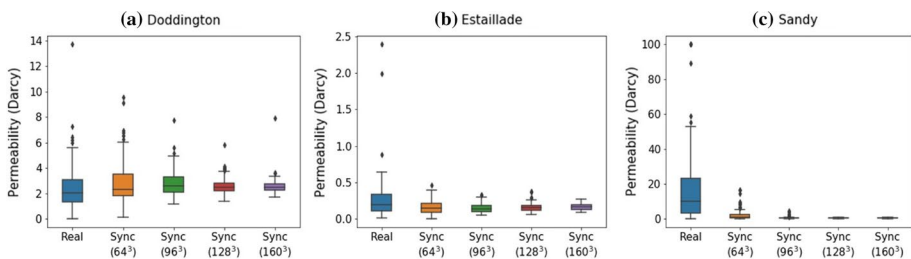

Fig. 13 Permeability of the training samples and the generated samples with different sizes for three rock types. The generated samples are the same as those in Fig. 11, which adopt different preset porosity as targets

denoted in Fig. 12c, and mainly contribute to the decrease of permeability. On the other hand, in Doddington sandstone or Estaillade carbonate, even though the pore sizes are changing, the pore throats may change a little, which have direct effects on rock permeability. Therefore, we can find that porosity conditioning can actually work through the proposed method. However, whether the changing porosity will affect permeability effectively is different among rock types, which requires further and systematic investigations but goes beyond the scope of this article.

3.3.3 Reconstruction conditioned on correlation length

In addition to porosity, we also test the conditioning performance of the second-order moment, i.e., the two-point correlations. Considering that the two-point correlation function is relatively hard to be conditioned on, we instead leverage a more underlying variable, i.e., the correlation length with respect to each correlation function. The normalized two-point correlation curves shown in Fig. 8 are close to the exponential model, defined as follows:

$$R(r) = \exp\left(-r/\lambda\right) \tag{8}$$

where r is the lag distance; and λ is the correlation length. They both use voxels as units in this case. The non-linear least square method can be utilized to fit $R(r)$ to the two-point correlation curves, and eventually obtain λ corresponding to each sample. In practice, we can make use of the Python library, named *scipy.optimize*, to realize it very conveniently.

In this case, we select Berea sandstone, Ketton carbonate, and Sandy multiscale medium to test the conditioning performance. We ignore anisotropy of correlations, since we found that it is not obvious in the datasets of selected rock types. Therefore, the conditional label λ is still a scalar value, and should concatenate with inputs for the generator and the discriminator as done for the porosity in the last section. In the same manner as the last section, when feeding the discriminator, the conditional λ should correspond to the training data; while feeding the generator, λ is uniformly sampled from the range that is determined by the extremum of real labels.

After training, we can use the learned model to generate realizations by assigning a specific correlation length. In this case, we set the target correlation lengths as $\lambda_{\text{target}} = 2.4$ voxels for Berea sandstone, $\lambda_{\text{target}} = 7.0$ voxels for Ketton carbonate, and $\lambda_{\text{target}} = 3.5$ voxels for Sandy multiscale medium. We select these target values following the same rule for porosity conditioning in the previous section. We generate 200 samples for each rock type
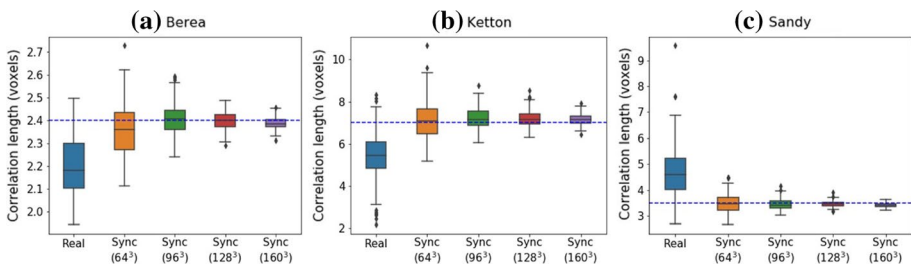

Fig. 14 Isotropic correlation lengths (voxels) of training samples and generated samples with different sizes of three rock types. The blue dashed lines represent the respective preset targets for each rock type

with different sizes, and randomly select the training samples with the same amount to demonstrate the performance of correlation length conditioning. As shown in Fig. 14, the correlation length of generated samples with different sizes are approaching the preset targets, and their ranges become narrower with increasing size, as we found for the porosity in the previous section. Meanwhile, we plot the correlation functions of training samples and synthetic samples with size 160^3 voxels in Fig. 15. We can find that the mean curves of the generated samples get very close to the targets, and this is especially obvious for Ketton carbonate and Sandy multiscale medium samples.

To further clearly show the conditioning effects of correlation length, we fix the input noise and increase the correlation length gradually to produce samples with size 96^3 voxels. Here, we choose 96^3 since the relatively smaller size is beneficial to visualize the local changes. Theoretically, the increase of correlation length will contribute to the connection of pores. As shown in Fig. 16, the Ketton carbonate sample presents the most visible changes, with pores becoming more connected from left to right in the figure. Similar to the porosity conditioning, the changes of the Sandy multiscale medium sample mostly lie in inter-grain pores, which become larger and thus have better connectivity when increasing correlation length. In contrast, even though the statistical metrics reveal good performance (as shown in Figs. 14 and 15), little obvious variations can be found in the Berea sandstone sample, which may be partly attributed to its rather small ranges of correlation length, i.e., approximately from 2.0 to 2.5.

Furthermore, as we did in the last section, we also evaluate the effects of changing correlation length on permeability. Firstly, we calculate the absolute permeability of the plotted samples in Fig. 16 to prove that correlation length does change. As shown in Table 2, the permeability of three rocks generally increases from sample 1 to sample 4 when increasing correlation length, which means that their correlation lengths did change even though the plotted samples may not reveal it very obviously. In addition, we also calculate the permeability of samples with different sizes, i.e., the samples used in Fig. 14, to investigate its statistical trend. As shown in Fig. 17, with preset targets larger than those of most samples, the permeability of Berea sandstone and Ketton carbonate samples exhibit a growing trend compared to training samples. In contrast, the permeability of Sandy multiscale medium samples decreases remarkably due to a relatively lower preset target. Therefore, it can be validated that the correlation length has a statistically positive correlation with permeability.

After validating the conditioning performance of isotropic correlation length, we move forward to test whether the model can condition on anisotropic correlation length. We
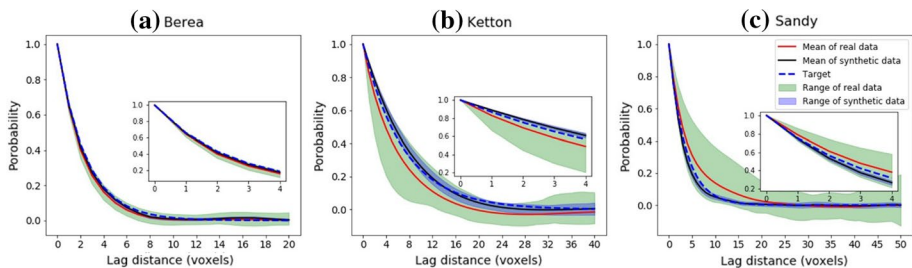

Fig. 15 Comparisons of the normalized two-point correlation function of real and synthetic samples of three rock types. The size of synthetic samples is 160^3 voxels
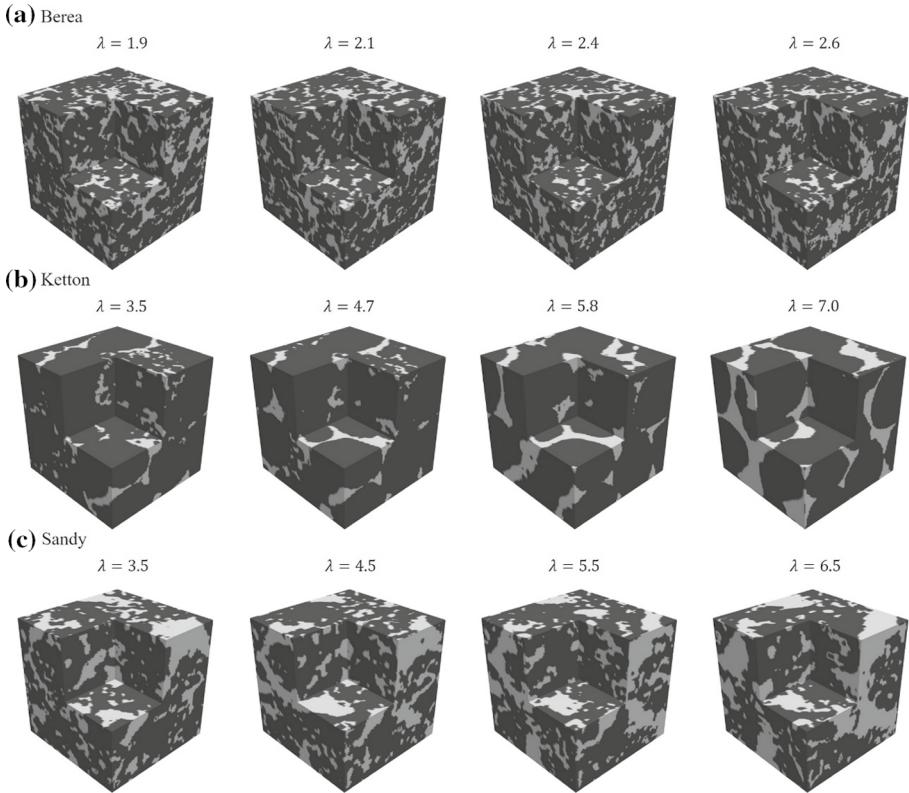
(a) Berea

(b) Ketton

(c) Sandy

Fig. 16 Synthetic samples with size 96^3 voxels when only varying correlation lengths λ. The three rows represent Berea sandstone, Ketton carbonate, and Sandy multiscale medium, respectively

Table 2 The permeability (Darcy) of samples in Fig. 16. The correlation length increases from sample 1 to sample 4

Rock type	Sample 1	Sample 2	Sample 3	Sample 4
Berea sandstone	0.56	0.62	1.08	1.16
Ketton carbonate	0.04	0.35	1.58	8.44
Sandy multiscale medium	0.63	6.04	9.32	22.03

adopt Doddington sandstone for this experiment. Based on the initially extracted samples, i.e., 4096 samples as stated in Sect. 3.1, we compute the correlation lengths of three directions according to Eq. (8), and then calculate the standard deviations of correlation lengths among three directions of each sample, and rank them. We select the former 2500 samples with relatively large standard deviations, aiming to eliminate the samples with no obvious anisotropy. Finally, we rotate the selected samples for two times as was done in Sect. 3.1 to obtain the eventual training dataset with size 7500.

Upon finishing the training, we use the trained model to generate 200 samples with different sizes by setting the correlation length target as $\lambda_x = 5.0$ voxels, $\lambda_y = 3.5$ voxels, and
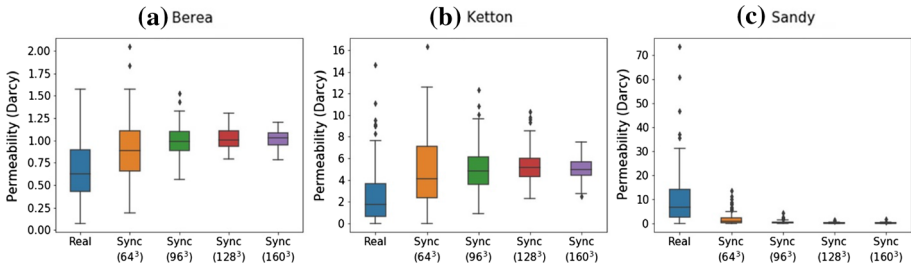
Fig. 17 Permeability of the training samples and the generated samples with different sizes for three rock types. The synthetic samples are generated by using the correlation length target set in Fig. 14
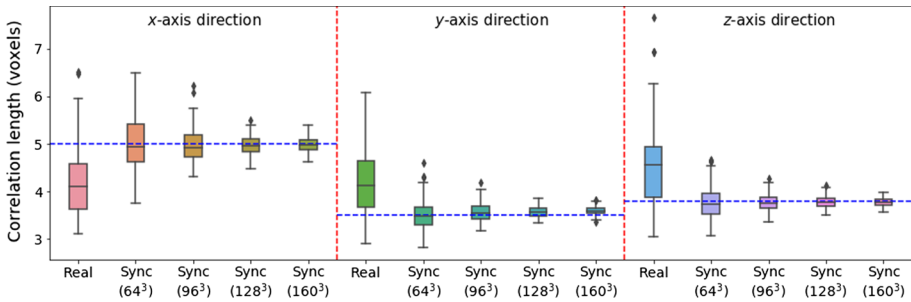

Fig. 18 Anisotropic correlation lengths (voxels) of the Doddington sandstone training samples and generated samples with different sizes. The blue dashed lines represent the respective preset targets for each direction
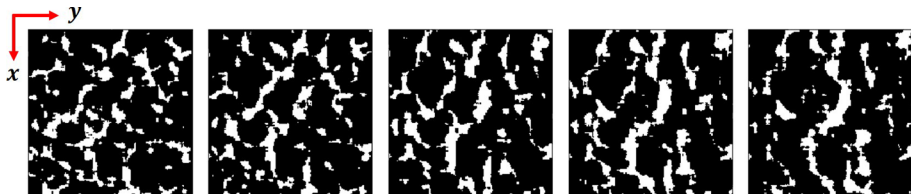

Fig. 19 x–y sections (160×160) of generated Doddington sandstone samples extracted at a randomly selected but same location, which are produced by one fixed input noise and changing correlation lengths, i.e., increasing λ_x from 3.5 voxels to 6.0 voxels with approximately equal intervals (i.e., 3.50, 4.13, 4.75, 5.38, 6.00) and maintaining $\lambda_y = 3.5$ voxels and $\lambda_z = 3.8$ voxels

$\lambda_z = 3.8$ voxels. As shown in Fig. 18, the conditioning performance is excellent for each direction, and the samples with larger size also show better matching with corresponding targets. To demonstrate the anisotropic conditioning more clearly, we produce samples by using a fixed input noise and a gradually changing correlation length in one direction. We fix $\lambda_y = 3.5$ voxels and $\lambda_z = 3.8$ voxels, and increase λ_x from 3.5 voxels to 6.0 voxels with roughly equal intervals (i.e., 3.50, 4.13, 4.75, 5.38, 6.00), and randomly select a x–y section from the same location of each sample. It can be seen from Fig. 19 that, through only changing λ_x, the pores become increasingly connected along the x-axis while keeping the connection almost unchanged along another direction. This means that the anisotropic
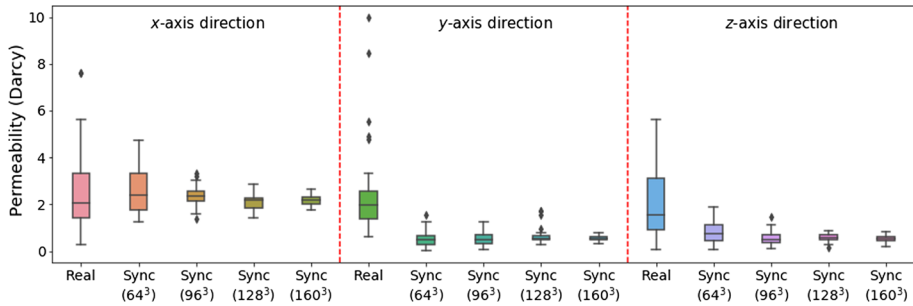
Fig. 20 Anisotropic permeability of Doddington sandstone training samples and generated samples with different sizes. The samples are the same as those used in Fig. 18, which adopt different preset correlation lengths as targets in three directions

conditioning does work, and the pores' connection can be modified along one specific direction. In addition, we also test the effects of conditioning of anisotropic correlation length on permeability. We adopt the same samples used in Fig. 18 to calculate anisotropic permeability, and plot the results in Fig. 20. From the figure, it can be seen that the permeability of generated samples in the y- and z-axis directions witness an obvious decrease compared to that of training ones, while the permeability of generated samples in the x-axis direction exhibits a slight increase. This phenomenon can further validate that the correlation length is positively correlated with permeability statistically. Through the above tests and discussions, we can conclude that the proposed conditional ProGAN can work in both isotropic and anisotropic correlation length conditioning, to provide synthetic samples with modified pore connections.

4 Conclusion

In this work, we extended the framework of digital rock random reconstruction with GANs by introducing a conditioning manner, aiming to produce random samples with user-specified rock types and first two moments. In order to stabilize the training process and naturally learn multiscale features, the conditional ProGAN using WGAN-GP loss was adopted in this work. Through three experiments, it was successfully verified that rock type, porosity, and correlation length could be effectively conditioned on to guide the morphology and geostatistical property of the generated samples without losing visual and geologic realism. Due to the conditioning manner, this study offers a new way to simultaneously incorporate experimental data and user-defined properties, which can be viewed as prior information, when reconstructing digital rocks. Those synthetic samples with user-defined rock type, porosity and correlation length, can be provided for researching the pore-scale multiphase flow behavior, especially for its uncertainty analysis, since the low-order statistical moments can be honored by the random samples.

Regarding comparisons with previous representative methods, the commonly used two-point statistics methods can only consider user-defined low-order moments, and subsequent multi-point statistics methods extract high-order information directly from the training images while leaving no interface for users to specify low-order properties. Therefore, the proposed framework in this work actually integrated their advantages, i.e., learning

high-order information directly from images using the GANs scheme, while honoring low-order properties through additional conditioning manner. Moreover, the conditioning of rock type may provide an alternative solution to previous isolated GANs training developed for a specific rock type, and consequently save computational cost. Additionally, a potential deficiency exists about rock type conditioning that needs to be mentioned. Since different kinds of rocks have different REV sizes, we refer to their maximum value to design the shape of training samples, aiming to make it satisfy the REV demands of all rock types simultaneously. However, even though the shape may be slightly large for those with small REV size, it will not negatively affect the learning performance of the proposed method.

An alternative research about how to manipulate the generated images is disentangled representation learning, which has been increasingly promoted to improve the interpretability of GANs (Chen et al. 2016; He et al. 2019; Karras et al. 2019) or other generative models, such as variational auto-encoders (Chen et al. 2018; Higgins et al. 2016). The disentanglement of the latent space, including conditional labels, i.e., making them uncorrelated and connected to individual aspects of the generated data, constitutes the key part in such kind of research. Inspired by this concept and in order to clearly demonstrate the effects of each conditional label, we did not carry out the experiment conditioned on the porosity and correlation length together because they are correlated in the binary microstructure (Lu and Zhang 2002). If the data are continuous, such as the heterogeneous field of geological parameters, it is certain that the first two moments can be conditioned on simultaneously since they are uncorrelated. Furthermore, we only consider the reconstruction of binary structures in this work, which are certain to be less realistic than gray-scale reconstructions, such as those in You et al. (2021). However, the digital rock reconstruction in this study is mainly developed for the subsequent multiphase flow modeling and uncertainty analysis, which only requires binary structures. Meanwhile, the proposed framework can be definitely adapted to gray-scale images if a sufficient investigation of morphological characteristics is needed.

References

Abdolahnejad, M., Liu, P.X.: Deep learning for face image synthesis and semantic manipulations: a review and future perspectives. Artif. Intell. Rev. **53**(8), 1–34 (2020)

Adler, P.M., Jacquin, C.G., Quiblier, J.A.: Flow in simulated porous media. Int. J. Multiphas. Flow **16**(4), 691–712 (1990)

Archie, F., Mughal, M.Z., Sebastiani, M., Bemporad, E., Zaefferer, S.: Anisotropic distribution of the micro residual stresses in lath martensite revealed by FIB ring-core milling technique. Acta Mater. **150**, 327–338 (2018)

Arjovsky, M., Chintala, S., and Bottou, L.: Wasserstein gan. arXiv preprint https://arxiv.org/abs/1701.07875(2017)

Biswal, B., Øren, P., Held, R.J., Bakke, S., Hilfer, R.: Stochastic multiscale model for carbonate rocks. Phys. Rev. E **75**(6), 61303 (2007)

Bostanabad, R., Zhang, Y., Li, X., Kearney, T., Brinson, L.C., Apley, D.W., Liu, W.K., Chen, W.: Computational microstructure characterization and reconstruction: review of the state-of-the-art techniques. Prog. Mater. Sci. **95**, 1–41 (2018)

Chen, C., Hu, D., Westacott, D., Loveless, D.: Nanometer-scale characterization of microscopic pores in shale kerogen by image analysis and pore-scale modeling. Geochem. Geophys. Geosyst. **14**(10), 4066–4075 (2013)

Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., and Abbeel, P.: Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In: Advances in neural information processing systems, pp. 2172–2180 (2016).

Chen, R.T., Li, X., Grosse, R.B., and Duvenaud, D.K.: Isolating sources of disentanglement in variational autoencoders. In: Advances in Neural Information Processing Systems, pp. 2610–2620 (2018).

Choi, Y., Choi, M., Kim, M., Ha, J., Kim, S., and Choo, J.: Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8789–8797 (2018).

Eshghinejadfard, A., Daróczy, L., Janiga, G., Thévenin, D.: Calculation of the permeability in porous media using the lattice Boltzmann method. Int. J. Heat Fluid Fl. **62**, 93–103 (2016)

Feng, J., Teng, Q., Li, B., He, X., Chen, H., Li, Y.: An end-to-end three-dimensional reconstruction framework of porous media from a single two-dimensional image based on deep learning. Comput. Method. Appl. M. **368**, 113043 (2020)

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems, pp. 2672–2680 (2014).

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A.C. Improved training of wasserstein gans. In: Advances in neural information processing systems, pp. 5767–5777 (2017).

He, Z., Zuo, W., Kan, M., Shan, S., Chen, X.: Attgan: Facial attribute editing by only changing what you want. IEEE t. Image Process. **28**(11), 5464–5478 (2019)

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., and Lerchner, A.: beta-vae: Learning basic visual concepts with a constrained variational framework. In: International Conference on Learning Representations (2016).

Icardi, M., Boccardo, G., Tempone, R.: On the predictivity of pore-scale simulations: Estimating uncertainties with multilevel Monte Carlo. Adv. Water Resour. **95**, 46–60 (2016)

Ji, L., Lin, M., Jiang, W., Cao, G.: A hybrid method for reconstruction of three-dimensional heterogeneous porous media from two-dimensional images. J. Asian Earth Sci. **178**, 193–203 (2019)

Joshi, M.: A class of stochastic models for porous materials. University of Kansas, Lawrence (1974)

Journel, A.G., Huijbregts, C.J.: Mining geostatistics. Academic press London (1978)

Jude, J.S., Sarkar, S., and Sameen, A.: Reconstruction of Porous Media Using Karhunen-Loève Expansion. In: proceedings of the International Symposium on Engineering under Uncertainty: Safety Assessment and Management. India, Springer, pp. 729–742 (2013)

Karras, T., Aila, T., Laine, S., and Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation. arXiv preprint https://arxiv.org/abs/1710.10196(2017)

Karras, T., Laine, S., and Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4401–4410 (2019).

Kench, S., Cooper, S.J.: Generating three-dimensional structures from a two-dimensional slice with generative adversarial network-based dimensionality expansion. Nat. Mach. Intell. **3**(4), 299–305 (2021)

King, H.E., Jr., Eberle, A.P., Walters, C.C., Kliewer, C.E., Ertas, D., Huynh, C.: Pore architecture and connectivity in gas shale. Energy Fuels **29**(3), 1375–1390 (2015)

Kingma, D.P., and Ba, J.: Adam: A method for stochastic optimization. arXiv preprint https://arxiv.org/abs/1412.6980 (2014)

Kitanidis, P.K.: Introduction to geostatistics: applications in hydrogeology. Cambridge University Press (1997)

Li, H., Singh, S., Chawla, N., Jiao, Y.: Direct extraction of spatial correlation functions from limited x-ray tomography data for microstructural quantification. Mater. Charact. **140**, 265–274 (2018)

Liang, J., Yang, X., Li, H., Wang, Y., Van, M.T., Dou, H., Chen, C., Fang, J., Liang, X., and Mai, Z.: Synthesis and Edition of Ultrasound Images via Sketch Guided Progressive Growing GANS. In: IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, pp. 1793–1797 (2020).

Lu, Z., Zhang, D.: On stochastic modeling of flow in multimodal heterogeneous formations. Water Resour. Res. **38**(10), 1–8 (2002)

Mariethoz, G., Caers, J.: Multiple-point geostatistics: stochastic modeling with training images. Wiley (2014)

Mariethoz, G., Renard, P., Straubhaar, J.: The direct sampling method to perform multiple-point geostatistical simulations. Water Resour. Res. (2010). https://doi.org/10.1029/2008WR007621

Mirza, M., and Osindero, S.: Conditional generative adversarial nets. arXiv preprint https://arxiv.org/abs/1411.1784 (2014)

Mohammadmoradi, P.: A Multiscale Sandy Microstructure. Digital Rocks Portal (2017). http://www.digitalrocksportal.org/projects/92.

Moon, C., and Andrew, M.: Intergranular Pore Structures in Sandstones. Digital Rocks Portal (2019) https://www.digitalrocksportal.org/projects/222.

Mosser, L., Dubrule, O., Blunt, M.J.: Reconstruction of three-dimensional porous media using generative adversarial neural networks. Phys. Rev. E **96**(4), 43309 (2017)

Mosser, L., Dubrule, O., Blunt, M.J.: Stochastic reconstruction of an oolitic limestone by generative adversarial networks. Transp. Porous Med. **125**(1), 81–103 (2018)

Muljadi, B.P.: Estaillade Carbonate. Digital Rocks Portal (2015). http://www.digitalrocksportal.org/projects/10.

Neumann, R., Andreeta, M., and Lucas-Oliveira, E.: 11 Sandstones: raw, filtered and segmented data. Digital Rocks Portal (2020). http://www.digitalrocksportal.org/projects/317.

Nie, D., Trullo, R., Lian, J., Wang, L., Petitjean, C., Ruan, S., Wang, Q., Shen, D.: Medical image synthesis with deep convolutional adversarial networks. IEEE Trans. Biomed. Eng. **65**(12), 2720–2730 (2018)

Okabe, H., Blunt, M.J.: Pore space reconstruction using multiple-point statistics. J. Petrol. Sci. Eng. **46**(1–2), 121–137 (2005)

Øren, P., Bakke, S.: Process based reconstruction of sandstones and prediction of transport properties. Transp. Porous Med. **46**(2–3), 311–343 (2002)

Papoulis, A., and Pillai, S.U.: Probability, random variables, and stochastic processes. Tata McGraw-Hill Education (2002).

Peng, S., Yang, J., Xiao, X., Loucks, B., Ruppel, S.C., Zhang, T.: An integrated method for upscaling pore-network characterization and permeability estimation: example from the mississippian barnett shale. Transport Porous Med. **109**(2), 359–376 (2015)

Pyrcz, M.J., Deutsch, C.V.: Geostatistical reservoir modeling. Oxford University Press (2014)

Quiblier, J.A.: A new three-dimensional modeling technique for studying porous media. J. Colloid Interf. Sci. **98**(1), 84–102 (1984)

Raeini, A.Q., Bijeljic, B., Blunt, M.J.: Generalized network modeling: Network extraction as a coarse-scale discretization of the void space of porous media. Phys. Rev. E **96**(1), 13312 (2017)

Shams, R., Masihi, M., Boozarjomehry, R.B., Blunt, M.J.: Coupled generative adversarial and auto-encoder neural networks to reconstruct three-dimensional multi-scale porous media. J. Petrol. Sci. Eng. **186**, 106794 (2020)

Song, S., Mukerji, T., Hou, J.: Geological Facies modeling based on progressive growing of generative adversarial networks (GANs). Computat. Geosci. **25**(3), 1251–1273 (2021)

Strebelle, S.: Conditional simulation of complex geological structures using multiple-point statistics. Math. Geol. **34**(1), 1–21 (2002)

Tahmasebi, P., Sahimi, M.: Cross-correlation function for accurate reconstruction of heterogeneous media. Phys. Rev. Lett. **110**(7), 78002 (2013)

Tahmasebi, P., Javadpour, F., Sahimi, M.: Three-dimensional stochastic characterization of shale SEM images. Transp. Porous Med. **110**(3), 521–531 (2015)

Wang, P., Chen, H., Meng, X., Jiang, X., Xiu, D., Yang, X.: Uncertainty quantification on the macroscopic properties of heterogeneous porous media. Phys. Rev. E **98**(3), 33306 (2018a)

Wang, Y.D., Armstrong, R.T., Mostaghimi, P.: Enhancing resolution of digital rock images with super resolution convolutional neural networks. J. Petrol. Sci. Eng. **182**, 106261 (2019)

Wang, Y.D., Armstrong, R.T., Mostaghimi, P.: Boosting resolution and recovering texture of 2D and 3D Micro-CT images with deep learning. Water Resour. Res. (2020). https://doi.org/10.1029/2019WR026052

Wang, T., Liu, M., Zhu, J., Tao, A., Kautz, J., and Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional gans. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8798–8807 (2018).

Wu, T., Li, X., Zhao, J., Zhang, D.: Multiscale pore structure and its effect on gas transport in organic-rich shale. Water Resour. Res. **53**(7), 5438–5450 (2017)

Xue, L., Li, D., Nan, T., Wu, J.: Predictive assessment of groundwater flow uncertainty in multiscale porous media by using truncated power variogram model. Transport Porous Med. **126**(1), 97–114 (2019)

You, N., Li, Y.E., Cheng, A.: 3D carbonate digital rock reconstruction using progressive growing GAN. J. Geophys. Res.: Solid Earth (2021). https://doi.org/10.1029/2021JB021687

Zha, W., Li, X., Xing, Y., He, L., Li, D.: Reconstruction of shale image based on wasserstein generative adversarial networks with gradient penalty. Adv. Geo-Energy Res. **4**(1), 107–114 (2020)

Zhang, D.: Stochastic methods for flow in porous media: coping with uncertainties. Elsevier (2001)

Zhang, D., Zhang, R., Chen, S., Soll, W.E.: Pore scale study of flow in porous media: Scale dependency, REV, and statistical REV. Geophys. Res. Lett. **27**(8), 1195–1198 (2000)

Zhang, T., Li, Y., Li, Y., Sun, S., Gao, X.: A self-adaptive deep learning algorithm for accelerating multi-component flash calculation. Comput. Method. Appl. M. **369**, 113207 (2020a)

Zhang, T., Li, Y., Sun, S., Bai, H.: Accelerating flash calculations in unconventional reservoirs considering capillary pressure using an optimized deep learning algorithm. J. Petrol. Sci. Eng. **195**, 107886 (2020b)

Zhao, L., Li, H., Meng, J., Zhang, D.: Efficient uncertainty quantification for permeability of three-dimensional porous media through image analysis and pore-scale simulations. Phys. Rev. E **102**(2), 23308 (2020)