



OPEN

Generating 3D images of material microstructures from a single 2D image: a denoising diffusion approach

Johan Phan^{1,2,3✉}, Muhammad Sarmad^{1,3}, Leonardo Ruspini², Gabriel Kiss¹ & Frank Lindseth¹

Three-dimensional (3D) images provide a comprehensive view of material microstructures, enabling numerical simulations unachievable with two-dimensional (2D) imaging alone. However, obtaining these 3D images can be costly and constrained by resolution limitations. We introduce a novel method capable of generating large-scale 3D images of material microstructures, such as metal or rock, from a single 2D image. Our approach circumvents the need for 3D image data while offering a cost-effective, high-resolution alternative to existing imaging techniques. Our method combines a denoising diffusion probabilistic model with a generative adversarial network framework. To compensate for the lack of 3D training data, we implement chain sampling, a technique that utilizes the 3D intermediate outputs obtained by reversing the diffusion process. During the training phase, these intermediate outputs are guided by a 2D discriminator. This technique facilitates our method's ability to gradually generate 3D images that accurately capture the geometric properties and statistical characteristics of the original 2D input. This study features a comparative analysis of the 3D images generated by our method, *SliceGAN* (the current state-of-the-art method), and actual 3D micro-CT images, spanning a diverse set of rock and metal types. The results shown an improvement of up to three times in the Fréchet inception distance score, a typical metric for evaluating the performance of image generative models, and enhanced accuracy in derived properties compared to *SliceGAN*. The potential of our method to produce high-resolution and statistically representative 3D images paves the way for new applications in material characterization and analysis domains.

Keywords 3D image generation, Material microstructures, Denoising diffusion probabilistic models, Single-image-based generation

Three-dimensional (3D) volumetric images are a critical resource in various disciplines, including geophysics, petroleum, and materials science, due to their role in the numerical analysis and computational modeling of materials' internal structures. These data, represented as a 3D grid of voxels (volumetric pixels), provide an intricate view of the internal structure of diverse materials, which is vital for deriving physical properties in different industries.

The acquisition of 3D images often poses significant challenges. Traditional image acquisition methods, such as computed tomography (CT) scanners, require substantial financial investment and skilled operators. Conventional techniques like micro-CT are often limited by the resolution capabilities of the imaging equipment. Practical issues related to sample preparation can further complicate acquiring high-resolution 3D images for specific materials or structures. Micro-CT scanners, which utilize X-rays, may also encounter difficulties penetrating radiodense materials, particularly metals, creating additional challenges in capturing comprehensive 3D images of such materials.

In contrast, sub-micrometer 3D scanning solutions such as nano-CT¹ and FIB-SEM (Focused Ion Beam Scanning Electron Microscope)² may offer higher resolution capabilities. However, these technologies come with a significantly higher price tag compared to micro-CT, have limited scanning sample sizes, and face limitations related to image quality, inhibiting their widespread application across various sectors within academia and industry^{3,4}.

¹Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway. ²Petricore Norway, Trondheim, Norway. ³These authors contributed equally: Johan Phan and Muhammad Sarmad. ✉email: johan.phan@ntnu.no

Given these challenges, there has been growing interest in developing techniques that can generate 3D volumetric data using 2D images. This approach offers a promising alternative, as 2D imaging methods such as optical microscopes or Scanning Electron Microscopes (SEM) are in many cases more cost-effective and flexible in terms of resolution capabilities, including at the nanoscale⁵. Currently, 2D images are primarily used alongside 3D images for quality control or as supplementary material when the resolution of 3D imaging fails to fully capture the studied sample's structure. Consequently, the capability to create 3D images from 2D images could reduce costs associated with the imaging process, enhance accessibility, and improve the efficiency of 3D image generation across diverse scientific fields.

Most previous approaches for generating 3D voxelized data from 2D input require learning from 3D image data. Our work belongs to the category of methods that generate 3D images from a single 2D input, as shown in Fig. 1. In addition, they can do so by learning from 2D images. The uniqueness of our work is that we propose a framework that utilizes a denoising diffusion-based probabilistic model (DDPM)⁶. Since DDPM requires ground truth (GT) data for training, we propose modifications to enable it to operate without needing 3D GT. This is achieved through utilizing the reverse diffusion process (chain sampling). In addition, we utilize the generative adversarial network (GAN) loss⁷ since it helps to generate realistic samples. GANs can be unstable when used in a standalone setting to learn from 2D images generating 3D images. However, we propose to stabilize training by combining GANs with diffusion models.

Additionally, addressing the practical requirements of 3D image generation for industrial applications, we have adapted the diffusion process to converge into the final image with just a few denoising steps, specifically 11 in this study, as opposed to the thousands of steps typically used in 2D image generation. This modification is crucial for reducing computation time when working with 3D data, where a large image with over a billion voxels ($1000 \times 1000 \times 1000$) is often necessary for comprehensive material characterization.

Our results prove to be more accurate than previous works in both visual quality and physical/statistical properties. We further showed that our model can successfully learn to generate 3D images from a single 2D input across a wide variety of cases, ranging from rocks to metal alloys.

The contributions of our work can be summarized as follows:

- We propose a method based on the Denoising Diffusion Probabilistic Model (DDPM) for generating 3D microstructures from a single 2D image.
- We demonstrated the feasibility of applying DDPM without the need for training on GT data.
- Our method significantly outperforms existing approaches in terms of visual quality and statistical properties. Moreover, it demonstrates robust performance even on complex images with high heterogeneity, where current state-of-the-art methods fail.

The ability to generate 3D images from a single 2D image would allow us to perform characterizations and analyses that require the availability of 3D data. This technique would be suitable for application in cases where micro-CT imaging is not feasible, such as when capturing features at sub-micrometer resolution or for materials without any density contrast, or in the case of high-density materials like metals⁵.

Related works

Creating 3D models or images of specific porous structures or materials has been a long-standing research challenge since the advent of image-based numerical analysis. Existing methodologies for tackling this problem

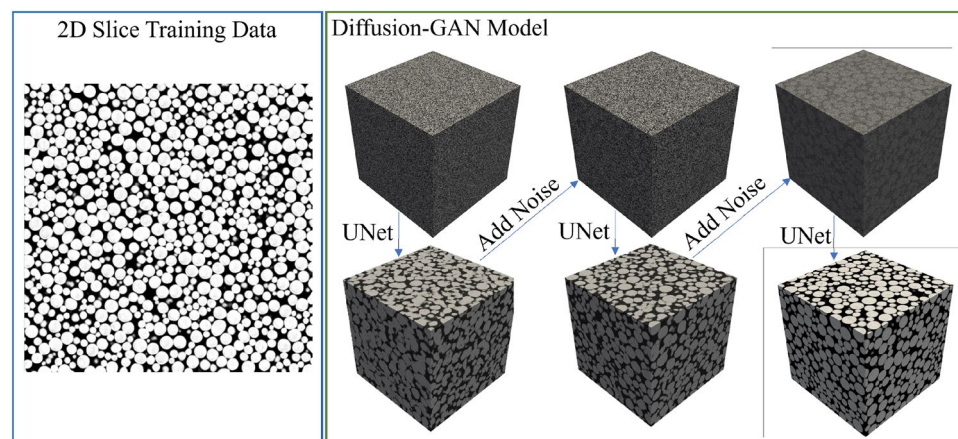


Figure 1. Diffusion-GAN model: the proposed method is based on a denoising diffusion process combined with a generative adversarial framework. In this setting, at test time, starting from a cube of noise, the noise is iteratively estimated using a Unet, removed and then added back to the sample. This process is repeated to obtain a noise-free representative sample. Our method can learn from only a few 2D slices of the training image, as shown on the left.

can be broadly classified into three main categories: process-based modeling, properties-based generation, and machine learning-based generation.

Process-based modeling

Process-based modeling approaches aim to emulate the mechanisms underlying the natural formation of materials. In these models, the physical and chemical processes that occur during material formation, such as deposition, compaction, cementation, dissolution, and fracturing, are translated into mathematical and computational algorithms. By closely imitating these natural processes, process-based models allow for extensive control over the properties and characteristics of the generated samples, making them useful for hypothesis testing or simulating a wide array of possible scenarios^{8–13}.

Despite their benefits, process-based models also have certain limitations. Simulating natural processes with satisfactory accuracy is computationally intensive, time-consuming, and challenging due to the complex interplay of numerous factors and the stochastic nature of many processes. Furthermore, operating these models requires a comprehensive understanding of the processes being replicated and the simplification of actual phenomena. As a result, there can be substantial discrepancies between the structures produced by these models and the real materials.

Properties-based generation

Generating 3D images can also be achieved through an iterative generation process that aims to converge toward a structure with desired statistical properties. This approach encompasses both stochastic-based modeling and optimization-based modeling^{14–16} where statistical descriptors such as the Minkowski functional and the n-point correlation function are commonly used. One of the main advantages of this method is its capability to generate models with specific desired properties. However, this approach is restricted to binary segmented images since most statistical descriptors for images are specifically designed for binary data. In more complex scenarios, especially for heterogeneous material, accurately capturing non-statistical representative features becomes challenging, potentially leading to the generation of unrealistic images even when the material's statistical properties are matched.

Machine learning-based generation

The recent advancements in 3D image generation have predominantly focused on utilizing machine learning techniques with existing 3D data to generate new images. One prominent approach in this field is the use of Generative Adversarial Networks (GANs)⁷. GANs have been applied for unconditional generation^{17,18} and conditional generation^{19–23}, of 3D images for micro-CT data. However, training a GAN-based model requires careful attention to ensure stability^{24,25}. Challenges such as mode collapse and catastrophic forgetting can arise when using GANs for conditional generation tasks, necessitating the incorporation of additional consistency loss^{26,27}.

To overcome the limitations of GAN-based methods, hybrid models that combine transformers and VQ-VAEs have emerged as an alternative solution. These models offer stable training and the ability to generate high-fidelity 3D rock samples from 2D conditional images²⁸.

However, all of the mentioned works rely on the availability of 3D GT data for training, which can pose limitations in terms of accessibility, particularly when dealing with samples that contain a significant number of sub-micrometer features. In a recent study, Kench et al.²⁹ showcased the capability of generating 3D microstructures with only 2D images as training data. Nevertheless, their approach relied on GANs, which are prone to common issues like unstable training and mode collapse.

In contrast to existing approaches, we propose a novel and stable diffusion-based method that achieves 3D image generation of material microstructures using only a single 2D image.

Background

Denoising diffusion probabilistic models

This section provides a basic understanding of Denoising Diffusion Probabilistic Models (DDPM), also known as diffusion models, which serve as the foundation for our proposed method. DDPM consists of two main processes: the forward and reverse processes⁶.

In the forward process, noise, typically Gaussian noise, is gradually added to the data distribution $q(\mathbf{x}_0)$, where \mathbf{x}_0 is the noise-free target. This process proceeds step by step, with the variance of the added noise changing according to a predefined schedule β_t (β_1, \dots, β_T). The forward process can be expressed as follows:

$$\begin{aligned} q(\mathbf{x}_{1:T}|\mathbf{x}_0) &= \prod_{t \geq 1} q(\mathbf{x}_t|\mathbf{x}_{t-1}) \\ &= \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}), \end{aligned} \quad (1)$$

In the reverse process, the aim is to recover the data from noise in steps. A diffusion model is required, which is parameterized by θ with mean $\mu_\theta(\mathbf{x}_t, t)$ and variance σ_t^2 . The reverse denoising process is given as:

$$\begin{aligned} p_\theta(\mathbf{x}_{0:T}) &= p(\mathbf{x}_T) \prod_{t \geq 1} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) \\ &= \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \sigma_t^2\mathbf{I}), \end{aligned} \quad (2)$$

To train this model, the variational bound on the negative log-likelihood objective $p_\theta(\mathbf{x}_0)$ is optimized, defined as $\int p_\theta(\mathbf{x}_{0:T}) d\mathbf{x}_{1:T}$. The variational lower bound is equivalent to matching the true denoising distribution $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$ with the parameterized denoising model $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ using the loss function:

$$\mathcal{L} = - \sum_{t \geq 1} \mathbb{E}_{q(\mathbf{x}_t)} [D_{\text{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t) \| p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))] + C \tag{3}$$

where D_{KL} represents the Kullback-Leibler (KL) divergence between the two distributions, i.e., the true denoising distribution $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$ and the parameterized denoising model $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$. C is a constant.

Two fundamental assumptions are commonly made in diffusion models: First, the denoising distribution $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ is modeled as a Gaussian distribution. Second, the number of denoising steps T is assumed to be large.

Denoising diffusion GANs

To address the challenge of requiring a large number of denoising steps in diffusion models, Xiao et al.³⁰ proposed a combination of Denoising Diffusion Probabilistic Models (DDPM) and Generative Adversarial Networks (GANs). Their work introduced two major modifications to the original diffusion process:

- **Adversarial Loss:** Instead of using typical loss functions like Mean Squared Error (MSE) or Mean Absolute Error (MAE), this method used an adversarial loss from a conditional discriminator.
- **Direct Output of Noise-Free Images:** Rather than training the DDPM to output the noise for a given image, which is then subtracted to obtain the noise-free image, this method directly generates the noise-free image.

By combining DDPM with GANs and implementing the mentioned modifications, the method proposed by Xiao et al. achieved a drastic reduction in the number of denoising steps required by a factor of 10^3 , while also improving the quality of the generated images and maintaining stable training.

Methods

Denoising diffusion models alone are inherently unsuitable for solving the problem of generating 3D images from a single 2D image, as they require a noise-free GT image \mathbf{x}_0 for training. The absence of \mathbf{x}_0 renders the forward process $q(\mathbf{x}_{1:T}|\mathbf{x}_0)$ mentioned in Eq. (1) invalid. Consequently, to adapt DDPM to our specific problem, we changed the original DDPM pipeline to cater to learning with only 2D GT data.

Inspired by the *SliceGAN* architecture introduced by Kench et al.²⁹, which utilized a 2D discriminator as the adversarial loss for a 3D generator, we have adapted the method proposed in the Denoising Diffusion GANs work by Xiao et al.³⁰ to a similar setting as *SliceGAN*. An overview of our method is shown in Fig. 2. The main difference between *SliceGAN* and our method is that *SliceGAN* uses only generative adversarial networks. However, we use the diffusion model with GANs to get a more stable and robust training pipeline. However, using the diffusion model for this task is not trivial. Therefore, we provide a novel method to deploy diffusion models for 3D image generation from 2D slices.

Chain sampling as a means of data generation

The lack of a noise-free GT \mathbf{x}_0 presents a significant challenge when adapting the Denoising Diffusion Probabilistic Models (DDPM) framework to the task of generating 3D images from a single 2D input. To overcome this challenge, we propose a novel approach called chain sampling, which involves leveraging the reverse diffusion

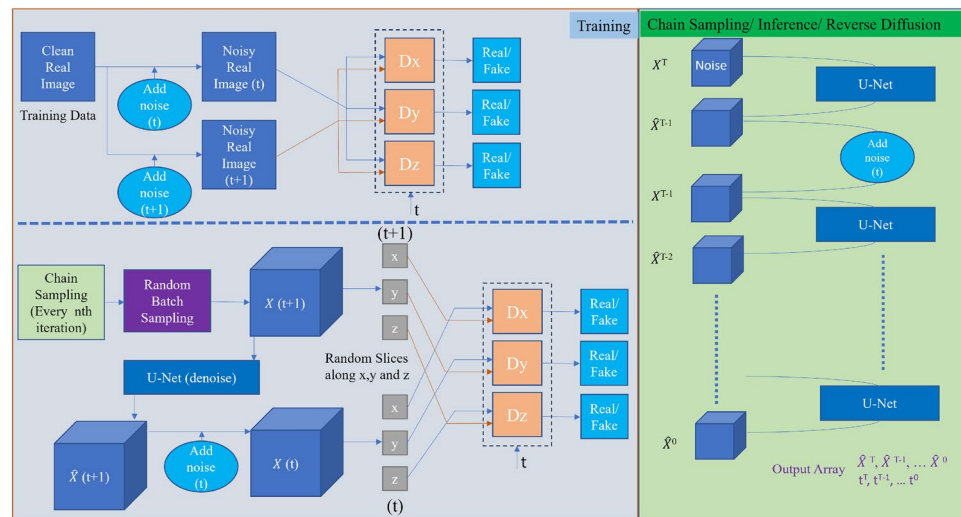


Figure 2. Our method: this figure shows the overview of our diffusion GAN based model for 3D generation using only 2D data for training.

process during training alongside the forward diffusion process. This departure from the conventional usage of the reverse process solely for inference or testing purposes is a key distinction in our method. By employing chain sampling, we can utilize intermediate results during training as a substitute for the missing noise-free 3D GT \mathbf{x}_0 , under the assumption that the denoising model is still progressing in the correct direction. The chain sampling process, illustrated in Fig. 2, involves adding the corresponding level of noise using a simplified noise addition process as shown in Eq. (4). The denoising model G then performs the denoising operation on the image, as described in Eq. (5).

$$\mathbf{x}_t = q(\hat{\mathbf{x}}_{t+1}) = \beta_t * \hat{\mathbf{x}}_{t+1} + (1 - \beta_t) * \mathcal{N}(\boldsymbol{\mu}_\theta(\hat{\mathbf{x}}_{t+1}, t + 1), \sigma_t^2) \quad (4)$$

$$\hat{\mathbf{x}}_t = G(\mathbf{x}_t, t) \quad (5)$$

Discriminator setting

To ensure the training of our denoising model in the absence of noise-free GT \mathbf{x}_0 , we employ a 2D discriminator trained on both the 2D image and the intermediate results from the chain sampling process. Similar to any GANs-based architecture, our discriminator requires both real data and fake (generated) data to train:

Fake Data: In the lower part of Fig. 2, we illustrate the process of generating fake data. During each training iteration, we sample an image from the output array of the chain sampling process, which corresponds to a denoised generated image $\hat{\mathbf{x}}_t$. From this image, we randomly select a slice from each axis (X, Y, Z) and feed these slices to their respective discriminators D_x, D_y, D_z (Eq. 6). While it is possible to use a single discriminator, we have found that utilizing three separate discriminators leads to more stable training and enables us to handle asymmetrical images effectively.

$$D_\phi(\mathbf{x}_{t-1}^{3D}, \mathbf{x}_t^{3D}, t) = D_x(\mathbf{x}_{t-1}^{3D_x}, \mathbf{x}_t^{3D_x}, t) + D_y(\mathbf{x}_{t-1}^{3D_y}, \mathbf{x}_t^{3D_y}, t) + D_z(\mathbf{x}_{t-1}^{3D_z}, \mathbf{x}_t^{3D_z}, t) \quad (6)$$

Real data: In the upper part of Fig. 2, we depict the process of generating real data for training the discriminator. Since our discriminator consists of 2D convolutional layers, we can easily add different levels of noise into the 2D image to retrieve the corresponding \mathbf{x}_{t-1} and \mathbf{x}_t . These noisy images, with noise added according to predefined β_t values, serve as the real data inputs for training the discriminator.

$$D_\phi(\mathbf{x}_{t-1}^{2D}, \mathbf{x}_t^{2D}, t) = D_x(\mathbf{x}_{t-1}^{2D_x}, \mathbf{x}_t^{2D_x}, t) + D_y(\mathbf{x}_{t-1}^{2D_y}, \mathbf{x}_t^{2D_y}, t) + D_z(\mathbf{x}_{t-1}^{2D_z}, \mathbf{x}_t^{2D_z}, t) \quad (7)$$

$$\min_\theta \sum_{t \geq 1} \mathbb{E}_{q(\mathbf{x}_t)} [D_x(q(\mathbf{x}_{t-1}^{2D} | \mathbf{x}_t^{2D}) \| q(\mathbf{x}_{t-1}^{3D_x} | \mathbf{x}_t^{3D_x})) + D_y(q(\mathbf{x}_{t-1}^{2D} | \mathbf{x}_t^{2D}) \| q(\mathbf{x}_{t-1}^{3D_y} | \mathbf{x}_t^{3D_y})) + D_z(q(\mathbf{x}_{t-1}^{2D} | \mathbf{x}_t^{2D}) \| q(\mathbf{x}_{t-1}^{3D_z} | \mathbf{x}_t^{3D_z}))], \quad (8)$$

To manage the computational intensity of generating large 3D images, we had to restrict the number of denoising timesteps to a smaller value, specifically $T = 11$. Consequently, this resulted in larger β_t values for each diffusion step. Since our approach involved a significantly reduced number of denoising timesteps compared to the original Denoising Diffusion GANs, we paid close attention to selecting suitable β_t values. The aim was to maintain a similar level of denoising complexity for each step, despite the reduced overall number of steps.

For the adversarial training, we define the sum of the three time-dependent discriminators as $D_\phi(\mathbf{x}_{t-1}, \mathbf{x}_t, t) : \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R} \rightarrow [0, 1]$, with parameters ϕ_x, ϕ_y , and ϕ_z , as shown in Eq. (8). This discriminator takes the N -dimensional 2D slices \mathbf{x}_{t-1}^{2D} and \mathbf{x}_t^{2D} of \mathbf{x}_{t-1} and \mathbf{x}_t as inputs and determines whether the input is a plausible denoised version of \mathbf{x}_t^{2D} or not.

Experiments

Data

In this study, we used a combination of internal data and publicly available data to evaluate the performance of our model. In the first case, we validated the quality of the generated images compared to 3D GT (Table 1) using four micro-CT 3D images: a Glass Bead image, two sandstone images with different resolutions, and a Savoniere carbonate image from the digital rock portal³¹. For each 3D image, we randomly selected five unrelated 2D slices to train our model.

In the second case (Table 2), we considered scenarios where only a single 2D image was available for both training and evaluation. The images used in this case include a Cast Iron with magnesium-induced spheroidized graphite and a Brass (Cu 70%, Zn 30%) with recrystallized annealing twins from Microlib³². Both images were captured using reflected light microscopy³³. In addition, we also used an SEM image of kaolinite clay minerals.

Evaluation metric

We utilize Fréchet Inception Distance (FID) as our evaluation metric³⁴. FID is a popular choice for assessing the quality of generated images in tasks such as GAN evaluations. It can serve as a measure of similarity between two

Rock type	Dimension	Our (x, y, z)			SliceGAN (x, y, z)		
Glassbead	200 × 200 × 200	54.78	60.95	59.67	87.37	99.02	72.33
Bentheimer sandstone	256 × 256 × 256	35.62	49.13	40.99	46.91	61.59	54.65
Sandstone	500 × 500 × 500	23.58	23.81	20.46	25.25	29.82	21.76
Savoniere	250 × 250 × 250	171.57	186.60	172.20	476.33	436.95	435.58

Table 1. Measured FID score across three dimensions (x, y, z) between the generated image and the GT, the closer to 0 the better.

Material type	Dimension	Our (x, y, z)			SliceGAN (x, y, z)		
Magnesium treated cast iron	800 × 528	60.03	62.90	63.37	63.81	67.89	66.04
Brass (Cu 70%, Zn 30%)	542 × 800	107.18	97.37	109.18	266.78	252.17	216.54
Kaolinite clay mineral	256 × 256	177.78	177.21	183.78.65	360.52	335.13	317.98

Table 2. Measured FID score across three dimensions (x, y, z) between the generated image and the GT.

datasets of images. The FID metric calculates the Fréchet distance between two multivariate Gaussian distributions that are fitted to feature representations of the Inception network. One distribution represents real images, while the other represents generated images. The lower the FID score, the more similar the two datasets of images are in terms of their distribution in the high-dimensional space defined by the Inception network. Hence, a lower FID signifies a higher quality of generated images. It captures how well the generated images mimic the real ones.

Resources and hyper-parameters

Each experiment in this study is conducted using PyTorch on a single Nvidia RTX 3090 GPU. The training time for both *SliceGAN* and our method was set to 24 hours. In our experiments, we used 11 denoising time steps ($T = 11$) with corresponding β_t values ranging from 0.9100 to 0.0000, as follows: [0.9100, 0.8109, 0.7058, 0.5985, 0.4929, 0.3931, 0.3025, 0.2238, 0.1586, 0.1070, 0.0685, 0.0000].

Results and discussions

Glass beads generation

Our study introduces a novel 3D image generation method, the efficacy of which we assessed through a comparative analysis with glass bead pack images. These images, which naturally depict spherical formations in a densely packed array, are reduced to varying sizes of 2D circles in their planar representations. Our validation approach involved contrasting our method's output with results from similar studies, focusing specifically on the fidelity of reconstructing 3D spherical shapes from these 2D circular projections. For this purpose, we selected benchmark studies by^{22,29,35,36} and¹⁸ for comparison.

Figure 3 showcases the cross-sectional views of structures synthesized using our method, alongside those generated by an authentic glass bead image, the *SliceGAN* algorithm, and the methods employed in the aforementioned studies. This highlights our method's unique capability in accurately rendering spherical shapes in 3D from 2D inputs, a feature distinctively absent in the comparative methods, especially in terms of artifact-free shape generation.

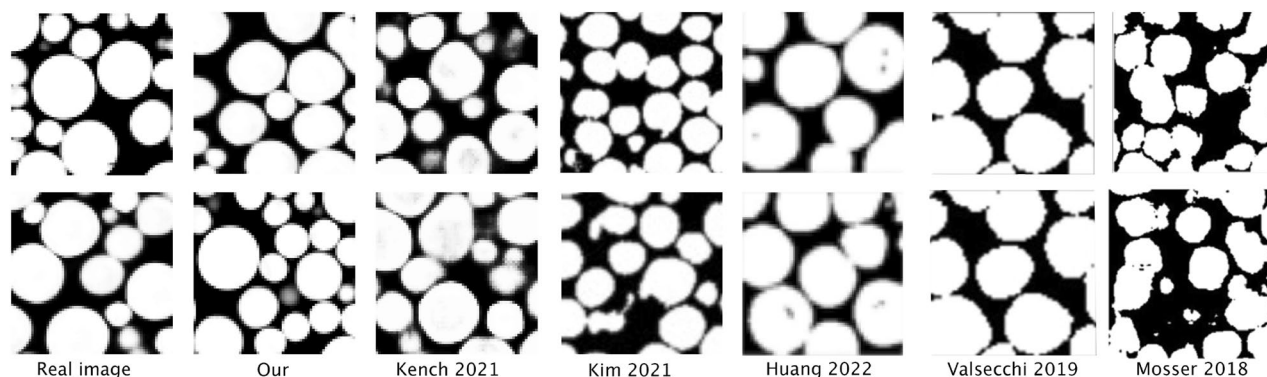


Figure 3. Visual comparison of spherical shape generation from 2D circular inputs in glass bead packs. This study contrasts our method's results with previous deep learning-based 3D image generation techniques, highlighting our approach's enhanced accuracy in generating spherical shapes.

Comparison with 3D GT

Visual comparison and FID score

Our first case study used four different 3D micro-CT images to evaluate both the visual quality and the accuracy of the characterized properties of our generated 3D images against the GT. For each image, five 2D slices of the xy plane, taken from different locations along the z-axis, were used to train our model and *SliceGAN*. We chose to use five 2D images since a single 2D slice might not fully capture the range of structural variation present in the 3D image. A step length of 11 was selected for our model to ensure fast generation times for large images. During each training iteration, we used a batch of random 64x64 pixel crops as input, which subsequently produced outputs of 64x64x64. The cross-sections of the images used in the training, as well as the images generated by both methods, are shown in Fig. 4.

In assessing the performance of our method compared to *SliceGAN*, we used the Fréchet inception distance (FID) scores as a measure of visual quality³⁷. To compute the FID score, the original requirement was for 2D images as input. To adapt this calculation to 3D images, we treated them as stacks of 2D images and computed the FID score across three dimensions (x, y, z). In comparison to other studies that used the FID score for image generation evaluation, the FID scores presented in Table 1 are notably higher. These higher FID scores are due to the few slices from the 3D image used for training not being able to cover the real data distribution of the 3D GT, especially for heterogeneous materials like the Savoniere Carbonate.

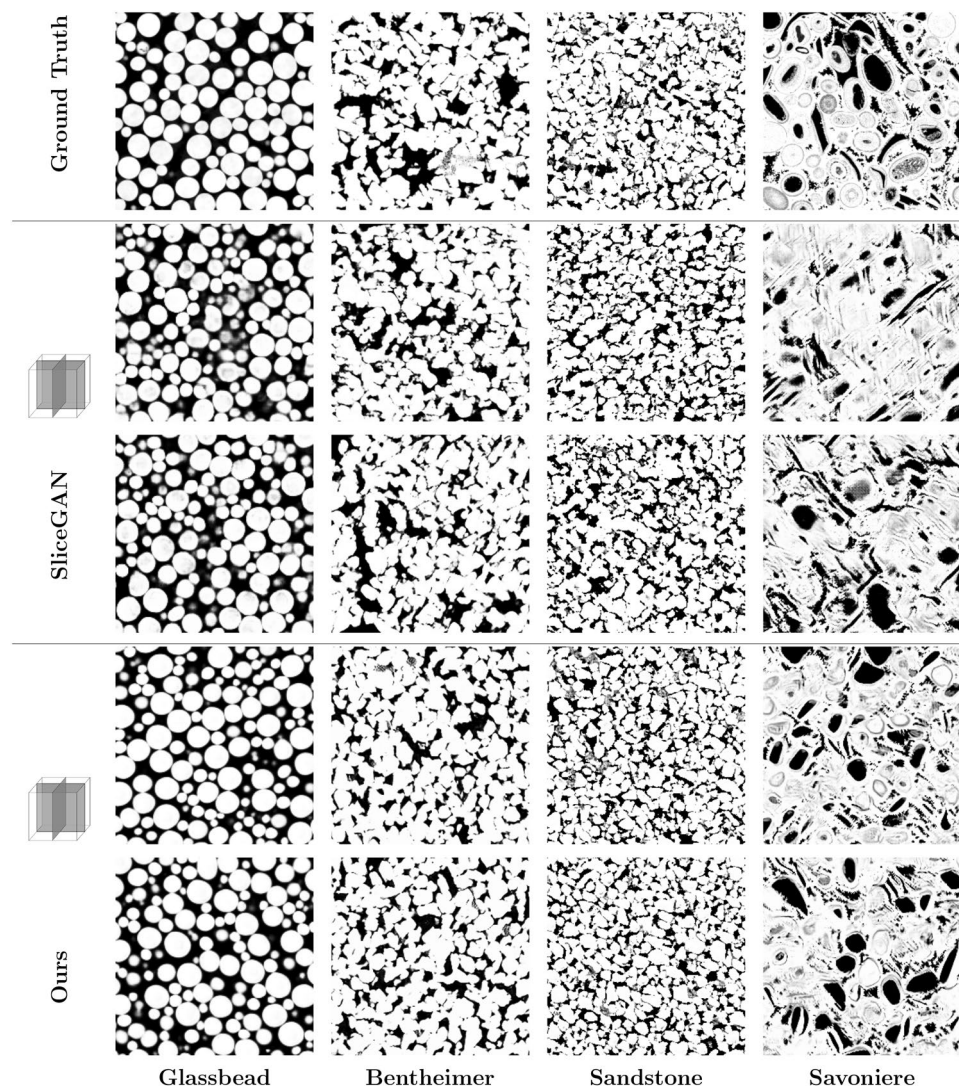


Figure 4. Visual comparison with micro-CT images: cross-sections of 3D images generated by our method and *SliceGAN*, alongside their respective ground truth or training data. The GT images are 3D X-ray microCT scans obtained at varying resolutions. The Glassbeads case showcases our method's superior performance over *SliceGAN*. Our model can capture the spherical shape of the object, even though it only sees circles at the 2D input. In more challenging cases like the Savoniere—a carbonate of fossilized microorganism—our method proves its robustness by generating images that bear a higher resemblance to reality, despite the heterogeneous nature of the original image.

Comparison of porous media properties

In the context of porous media, it is crucial to evaluate our model's performance in terms of physical properties. To calculate these properties, we used Porespy³⁸, an open-source tool specifically designed to analyze 3D images of porous materials. With Porespy, we calculated local porosity, the two-point correlation function, and pore size distribution of the images depicted in Fig. 4.

Porosity (ϕ)—Porosity, representing the volume fraction of void spaces, is a fundamental characteristic of porous media. To calculate porosity, we first convert the images into binary format through thresholding. Subsequently, we divide the generated images into overlapping cubes with a side length of 128 voxels. The process of calculating porosity is then applied to these cubes, and the results are visualized using box plots shown in Fig. 5.

Two-point correlation function (ξ)—the two-point correlation function is a significant metric in image analysis, utilized to describe the spatial arrangement and connectivity of the porous structure. In this study, we calculated the probability that a pair of points, separated by a certain distance, both reside within the pore space. This statistical measure is sensitive to the image's degree of homogeneity and isotropy, thus allowing us to capture subtle geometric features of the pore network. The two-point correlation function plots are shown in Fig. 6.

Pore size distribution—pore size distribution is a metric that characterizes the range of pore sizes within a porous material, and it plays a vital role in determining how fluids flow and permeate through the material. Ensuring the accuracy of pore size distribution in our 3D generation from 2D images is important because the sizes and arrangement of pores define the transport properties of the porous medium. Accurate pore size distribution in the generated 3D images is essential for maintaining physical accuracy and predictive usefulness in representing the actual material. We calculated the pore size distribution through a process known as porosimetry, which interprets each voxel in the image as the radius of the largest sphere that would overlap it. For this, we used the Porespy library and the result is shown in Fig. 7.

3D generation experiments from 2D image of a sample

In the final part of our evaluation, we went beyond testing our method's performance with 3D GT data and tested it in scenarios where only a single 2D image was available, as shown in Fig. 8. We used two metal images (Cast Iron and Brass) from Microlibs³², an online database for images generated by *SliceGAN*, as our test data. We trained our model on the downloaded 2D image and subsequently compared our output with both the original training image and the 3D image created by *SliceGAN*, also sourced from the same Microlibs platform. Our study also included an SEM-acquired image of kaolinite clay, recognized for its complex nanostructure that necessitates capturing in 2D. This is due to the high resolution required to capture the sub-microscopic structure of kaolinite, which is beyond the capabilities of current 3D CT scanners.

These cases represent the scenarios where our algorithm may find its most practical use—situations where acquiring 3D images is infeasible, hence the necessity to generate a 3D model from a 2D image.

The comparison in terms of FID score is shown in Table 2.

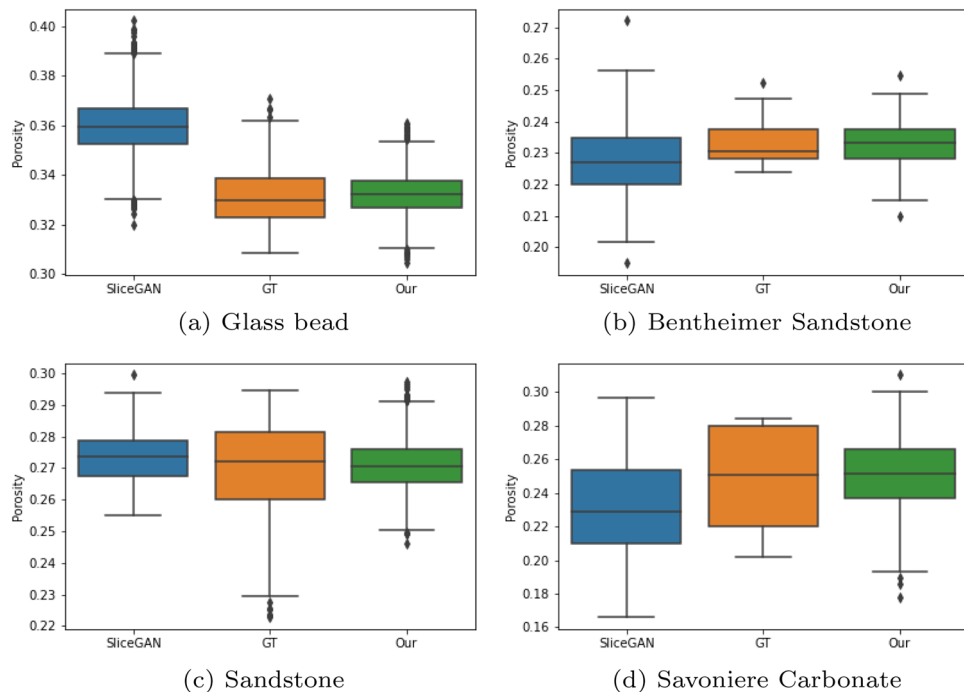


Figure 5. Porosity—these box plots show the comparison of porosity between the ground truth, our model and *SliceGAN*.

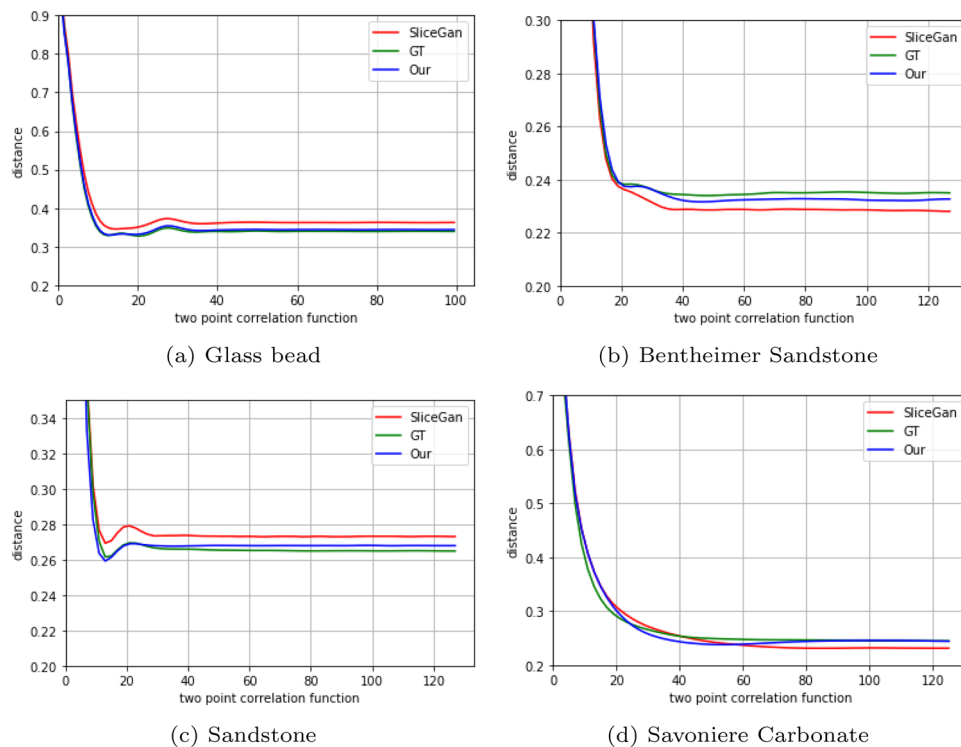


Figure 6. Two-point correlation—these plots depict the two-point correlation function in 3D for the Ground Truth (GT) and the images generated by both SliceGAN and our method. They display the relationship between distance and the probability of a given pixel appearing in a binarized segmented image.

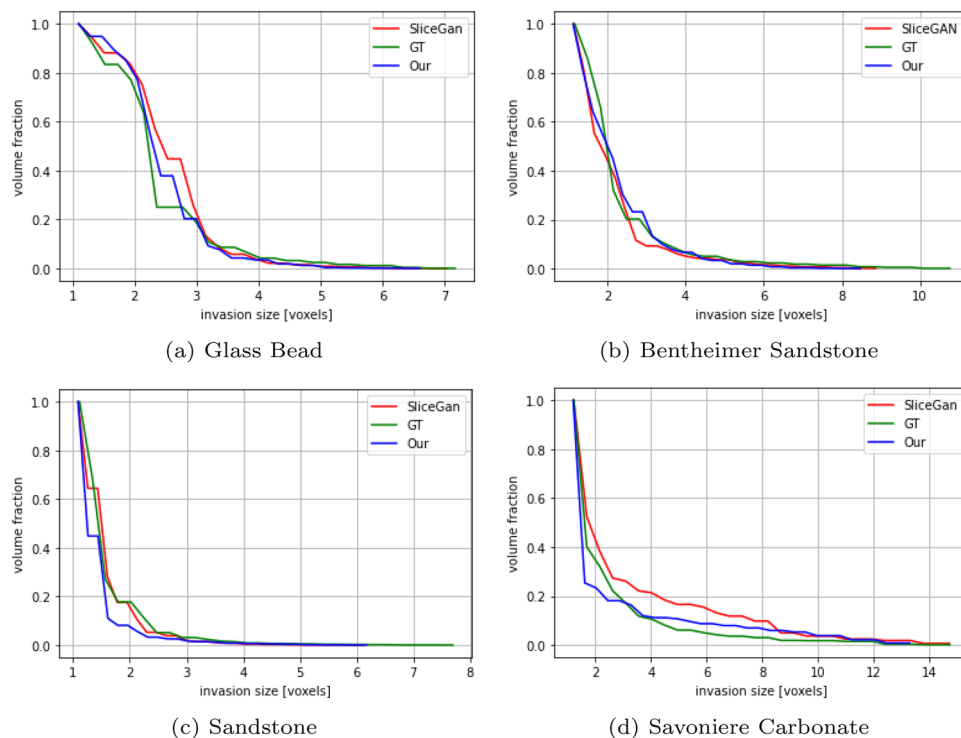


Figure 7. Pore size distribution—these plots shown the comparison of pore size distribution between the generated images and the ground truth. They aid in understanding our model’s effectiveness and SliceGAN’s matching the pore structures to the original 3D micro-CT images.

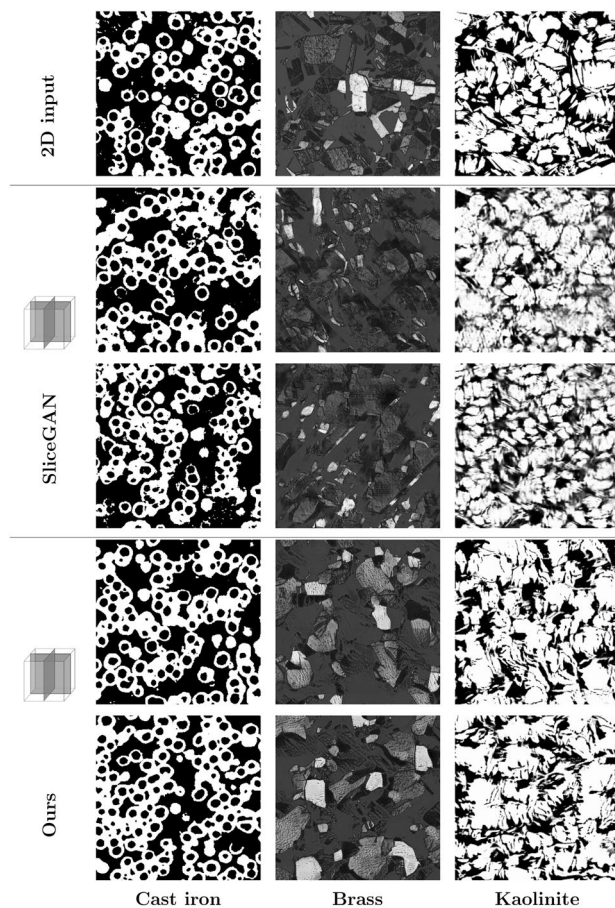


Figure 8. Visual comparison with non-CT data sources: cross-sections of 3D images produced by our method and SliceGAN, presented alongside their respective 2D training images. This figure includes images of cast iron, brass, and kaolinite clay mineral.

Discussion

Our method demonstrates a significant improvement over *SliceGAN*, evident at a visual level, as shown in Figs. 4 and 8. For example, in the glass bead case, our method successfully manages to generate 3D spherical structures from training with 2D circular input, while *SliceGAN* and other machine learning based method failed. In the Sandstone cases, our method demonstrated its capability to handle various resolutions and grain sizes. The Savoniere case, owing to the image's heterogeneity, presents a challenge in generating a representative 3D image solely from 2D information. Despite this, our method manages to produce a more visually accurate output compared to *SliceGAN* with a significantly lower FID score (Table 1).

Our evaluation of porosity (Fig. 5), the two-point correlation function (Fig. 6), and pore size distribution (Fig. 7), further affirms our method's efficacy in representing the statistical properties of the 3D GT, even when trained only on five 2D slices. In the final study case where only a single 2D image is available for training and evaluation, we downloaded the 2D image for training and the 3D image created by *SliceGAN* from Microlib. As seen in Fig. 8, our method exhibits versatility by generating high-quality images across different material types. Due to the absence of 3D GT images for these cases, we relied on visual inspection and FID scores for evaluation. Despite the insignificant difference in the visual quality and FID score for the cast iron case, the FID score measured in the Brass case clearly favors our method over *SliceGAN* (see Table 2).

In all study cases included in this work, our method has shown comparable or better performance compared to *SliceGAN*. Additionally, the generated 3D images for materials with simple and homogeneous structures closely match the real images, exhibiting both comparable visual quality and measured properties. However, for more complex and heterogeneous materials, especially those with asymmetrical 3D features, there are areas that indicate potential for improvement. Nevertheless, the results of this study lay a promising foundation for future exploration in the domain of 3D porous media image generation from 2D inputs.

Conclusion

In this study, we introduced a novel approach to 3D image generation using denoising diffusion probabilistic models (DDPMs) with only a single 2D slice as training data. While DDPMs are not inherently designed for learning from 2D data to represent 3D structures, we introduced a modified reverse diffusion step that effectively

denoises a 3D noise vector using a 2D GAN-based discriminator. Our method outperforms state-of-the-art techniques in terms of key physical validation metrics for various types of materials.

Our work marks a significant advancement in the domain of 3D material microstructure generation from 2D inputs. By reducing the dependency on extensive 3D image data and offering a cost-effective, high-resolution alternative to prevailing imaging techniques, our approach paves the way for novel research and practical applications in material characterization and analysis.

Data availability

The 2D training datasets presented in this study is publicly available at³². The 3D datasets presented in this study available from the corresponding author on reasonable request.

Received: 28 September 2023; Accepted: 12 March 2024

Published online: 18 March 2024

References

- Kampschulte, M. *et al.* *Nano-computed Tomography: Technique and Applications* vol. 188(no. 2) 146–154 (Georg Thieme Verlag KG, 2016).
- Grober, M. A., Haley, B., Uchic, M. D., Dimiduk, D. M. & Ghosh, S. 3d reconstruction and characterization of polycrystalline microstructures using a fib-sem system. *Mater. Charact.* **57**(4–5), 259–273 (2006).
- Xu, C. S. *et al.* Enhanced fib-sem systems for large-volume 3d imaging. *Elife* **6**, 25916 (2017).
- Ahmed, H. M. A. Nano-computed tomography: Current and future perspectives. *Restor. Dent. Endod.* **41**(3), 236–238 (2016).
- CT vs. SEM: Which Is Better?—*imaging.rigaku.com*. <https://imaging.rigaku.com/blog/ct-vs-sem-which-is-better>. Accessed December 27 2023.
- Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **33**, 6840–6851 (2020).
- Goodfellow, I. *et al.* Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **27**, 66 (2014).
- Adler, P. M., Jacquin, C. G. & Qublier, J. A. Flow in simulated porous media. *Int. J. Multiph. Flow* **16**(4), 691–712. [https://doi.org/10.1016/0301-9322\(90\)90025-E](https://doi.org/10.1016/0301-9322(90)90025-E) (1990).
- Strebelle, S. Conditional simulation of complex geological structures using multiple-point statistics. *Math. Geol.* **34**, 1–21 (2002).
- Blair, S. C., Berge, P. A. & Berryman, J. G. Using two-point correlation functions to characterize microgeometry and estimate permeabilities of sandstones and porous glass. *J. Geophys. Res. Solid Earth* **101**(B9), 20359–20375. <https://doi.org/10.1029/96JB00879> (1996).
- Tahmasebi, P., Hezarkhani, A. & Sahimi, M. Multiple-point geostatistical modeling based on the cross-correlation functions. *Computat. Geosci.* **16**, 779–797 (2012).
- Tahmasebi, P., Sahimi, M. & Caers, J. Ms-ccsim: Accelerating pattern-based geostatistical simulation of categorical variables using a multi-scale search in Fourier space. *Comput. Geosci.* **67**, 75–88 (2014).
- Feng, J., Teng, Q., He, X., Qing, L. & Li, Y. Reconstruction of three-dimensional heterogeneous media from a single two-dimensional section via co-occurrence correlation function. *Comput. Mater. Sci.* **144**, 181–192 (2018).
- Seibert, P., Raßloff, A., Ambati, M. & Kästner, M. Descriptor-based reconstruction of three-dimensional microstructures through gradient-based optimization. *Acta Mater.* **227**, 117667 (2022).
- Scheunemann, L., Balzani, D., Brands, D. & Schröder, J. Design of 3d statistically similar representative volume elements based on Minkowski functionals. *Mech. Mater.* **90**, 185–201 (2015).
- Lu, B. & Torquato, S. Lineal-path function for random heterogeneous materials. *Phys. Rev. A* **45**(2), 922 (1992).
- Mosser, L., Dubrule, O. & Blunt, M. J. Reconstruction of three-dimensional porous media using generative adversarial neural networks. *Phys. Rev. E* **96**(4), 66 (2017).
- Mosser, L., Dubrule, O. & Blunt, M. J. Stochastic reconstruction of an oolitic limestone by generative adversarial networks. *Transp. Porous Media* **125**(1), 81–103 (2018).
- Volkhonskiy, D., Muravleva, E., Sudakov, O., Orlov, D., Belozero, B., Burnaev, E., & Koroteev, D.: *Reconstruction of 3d porous media from 2d slices*. [arXiv:1901.1023v1](https://arxiv.org/abs/1901.1023v1) (2019)
- Zhao, J., Wang, F. & Cai, J. 3d tight sandstone digital rock reconstruction with deep learning. *J. Petrol. Sci. Eng.* **207**, 109020. <https://doi.org/10.1016/j.petrol.2021.109020> (2021).
- Coiffier, G., Renard, P. & Lefebvre, S. 3d geological image synthesis from 2d examples using generative adversarial networks. *Front. Water* **2**, 30. <https://doi.org/10.3389/frwa.2020.560598> (2020).
- Valsecchi, A., Damas, S., Tubilleja, C. & Arechalde, J. Stochastic reconstruction of 3d porous media from 2d images using generative adversarial networks. *Neurocomputing* **399**, 227–236 (2020).
- Shams, R., Masihi, M., Boozarjomehry, R. B. & Blunt, M. J. A hybrid of statistical and conditional generative adversarial neural network approaches for reconstruction of 3d porous media (st-gan). *Adv. Water Resour.* **158**, 104064 (2021).
- Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A.: Improved training of Wasserstein Gans. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)* 5769–5779 (Curran Associates Inc., 2017)
- Arjovsky, M., Chintala, S., & Bottou, L.: Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning* vol. 70 214–223 (Proceedings of Machine Learning Research, 2017)
- Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., & Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 4681–4690 (2017)
- Thanh-Tung, H., & Tran, T.: Catastrophic forgetting and mode collapse in gans. In *2020 International Joint Conference on Neural Networks (IJCNN)* 1–10 (2020). IEEE.
- Phan, J., Ruspini, L., Kiss, G. & Lindseth, F. Size-invariant 3d reconstruction from a single 2d rock image. *J. Petrol. Sci. Eng. D* **215**, 110648. <https://doi.org/10.1016/j.petrol.2022.110648> (2022).
- 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).
- Xiao, Z., Kreis, K., & Vahdat, A.: *Tackling the generative learning trilemma with denoising diffusion gans*. arXiv preprint [arXiv:2112.07804](https://arxiv.org/abs/2112.07804) (2021)
- Bultreys, T. *Savonnières carbonate*. Digital Rocks Portal. <https://doi.org/10.17612/P7W88K> (2016).
- Kench, S., Squires, I., Dahari, A. & Cooper, S. J. Microlib: A library of 3d microstructures generated from 2d micrographs using sliceGAN. *Sci. Data* **9**(1), 645 (2022).
- Ryan, J., Gerhold, A. R., Boudreau, V., Smith, L. & Maddox, P. S. Introduction to modern methods in light microscopy. *Light Microsc. Methods Protoc.* **66**, 1–15 (2017).
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. & Hochreiter, S. Gans trained by a two time-scale update rule converge to a local Nash equilibrium. *Adv. Neural Inf. Process. Syst.* **30**, 66 (2017).

35. Kim, S. E., Yoon, H. & Lee, J. Fast and scalable earth texture synthesis using spatially assembled generative adversarial neural networks. *J. Contam. Hydrol.* **243**, 103867. <https://doi.org/10.1016/j.jconhyd.2021.103867> (2021).
36. Huang, Y., Xiang, Z. & Qian, M. Deep-learning-based porous media microstructure quantitative characterization and reconstruction method. *Phys. Rev. E* **105**(1), 015308 (2022).
37. Seitzer, M.: pytorch-fid: FID Score for PyTorch. <https://github.com/mseitzer/pytorch-fid>. Version 0.3.0 (2020).
38. Gostick, J. T. *et al.* Porespy: A python toolkit for quantitative analysis of porous media images. *J. Open Source Softw.* **4**(37), 1296 (2019).

Acknowledgements

This work was partially supported by the Norwegian Research Council (Grant Number 296093) and the members of the SmartRocks joint industry project (Repsol AS, and Chevron Corporation).

Author contributions

J.P. and S.M. conducted the experiments and wrote the main manuscript text. All authors contributed to the research idea and reviewed the manuscript.

Funding

Open access funding provided by Norwegian University of Science and Technology.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024