ISSN 2616-5880 Vol. 6, Issue 2: 51-56, DOI: 10.25236/AJMC.2025.060207

A New Digital Reconstruction Method of Porous Materials Based on Diffusion Model

Lida Zhang, Wanting Hu, Chenhao Li, Jiawen Cai, Lihong Zhu

College of Science, China Jiliang University, Hangzhou, 310018, China

Abstract: *Porous materials are ubiquitously present in natural environments and engineering applications, where their pore structural characteristics serve as critical factors influencing functional performance. Consequently, quantitative characterization of porous structures have attracted multidisciplinary attention, providing theoretical support for aerospace engineering, electrochemical engineering, functional materials, and biochips. The rapid advancement of computational technologies has enabled digital reconstruction as effective tool for characterizing the microstructures of porous materials. Machine learning-based methodologies have established novel pathways for digital reconstruction of porous materials. Thus, a new structural generation method for porous materials is proposed based on diffusion model. By using the Bentheimer sandstone 2D slices from the digital rocks super-resolution dataset, the image generation is implemented through denoising diffusion models following data preprocessing. Both Fréchet Inception Distance (FID) and Learned Perceptual Image Patch Similarity (LPIPS) metrics are used to evaluate the generated porous images. The results demonstrate the superior performance of diffusion models in generating high-fidelity images, achieving FID and LPIPS scores of 284.6933 and 0.165 respectively. The generated porous images exhibit enhanced structural authenticity compared to conventional methods. The present model provide a comprehensive framework for structural reconstruction of porous materials.*

Keywords: *Diffusion model; Machine Learning; Porous Materials; Digital Reconstruction*

1. Introduction

Porous materials are ubiquitously present in natural environments and engineering applications, where their pore structures and physicochemical properties critically influence transport characteristics and functional performance, attracting multidisciplinary research interest[1-2]. The pore architecture and transport dynamics serve as decisive factors in determining material functionality and process efficiency, driving extensive investigations into quantitative characterization and property prediction of photoelectric porous materials across aerospace engineering, electrochemical systems, functional materials, and biochip technologies.

Advancements in computational methodologies have enabled digital reconstruction and numerical simulation as pivotal tools for studying porous media. Traditional reconstruction approaches primarily rely on experimental data or 2D images to construct two/three-dimensional models through mathematical and statistical techniques. These include stochastic reconstruction based on experimental parameters and numerical methods such as Gaussian field simulation, simulated annealing, process-based modeling, multiple-point statistics, Markov chain Monte Carlo (MCMC) methods, sequential indicator simulation, and hybrid strategies. However, conventional digital core reconstruction techniques suffer from significant hardware burdens and prolonged computational durations, substantially limiting their applicability [3-4].

The emergence of digital rock physics has revolutionized porous media characterization by providing novel structural analysis methods that address the accuracy limitations and time-intensive nature of traditional approaches. As an advanced digital methodology, machine learning-based reconstruction leverages computational algorithms to reconstruct porous media models from limited data, overcoming experimental constraints while delivering precise geometric representations for numerical simulations and performance predictions. For instance, Liqun Shan et al. [5] developed a super-resolution reconstruction algorithm integrating convolutional neural networks (CNN), residual learning, and attention mechanisms to generate high-resolution carbonate and sandstone images with indistinguishable high-frequency details. Yuzhu Wang et al. [6] implemented neighborhood embedding algorithms for micro-CT image enhancement, supplementing low-frequency data from micro-CT with

ISSN 2616-5880 Vol. 6, Issue 2: 51-56, DOI: 10.25236/AJMC.2025.060207

high-frequency features from scanning electron microscopy (SEM) images to achieve state-of-the-art reconstruction performance. Javad Siavashi et al. [7] proposed a CNN-based upscaling framework combined with downsampling techniques to predict macroscopic properties for single- and two-phase flows, demonstrating high consistency in dynamic behavior between coarse and high-resolution models while reducing computational costs. Generative adversarial networks (GAN) [8] currently dominate structural reconstruction methodologies. Wenshu Zha et al. [9] enhanced reconstruction quality and efficiency using Wasserstein GAN with gradient penalty, where CNN serve as both generator and discriminator networks. Mosser et al. [10] developed a GAN-based framework for rapid generation of porous solid-void structures through implicit probability distribution modeling of 3D image datasets. Junxi Feng et al. [11] accelerated multipoint statistical reconstruction via conditional GAN (cGAN) to maintain statistical consistency with target systems. Feng et al. [12] implemented cGAN for full-image reconstruction from subregions using coupled objective functions to constrain training stability. Reza Shams et al. [13] stabilized 3D reconstructions through autoencoder-GAN hybrids optimized via gradient descent. Yang et al. [14] established scale-independent multiscale models using cGANs for improved pore-scale characterization. Rui Xu et al. [15] constructed physics-informed encoder-decoder networks to simulate 3D fluid flow in heterogeneous permeability fields, achieving high agreement with numerical simulations at reduced computational costs. Feng et al. [16] pioneered a bicycleGAN-based framework for direct 2D-to-3D image translation with verified accuracy and efficiency.

Nevertheless, existing machine learning-based reconstruction models frequently encounter challenges including training instability, mode collapse, and uncontrollable noise artifacts. To address these limitations, this study proposes the adoption of diffusion models for porous material image generation, capitalizing on their enhanced stability and precise controllability in synthetic data production.

2. Denoising Diffusion Models

The theoretical foundation of Denoising Diffusion Probabilistic Models (DDPM) originates from interdisciplinary research in non-equilibrium thermodynamics and stochastic dynamics. Sohl-Dickstein et al. [17] first proposed its fundamental paradigm in 2015, and subsequent advancements by Ho et al. [18] in 2020 established the modern formulation of DDPM through systematic theoretical derivations and algorithmic optimizations. This model achieves reversible mapping between data distributions and Gaussian noise via a Markov chain-driven bidirectional stochastic process. In the field of image generation, DDPM has been widely applied to tasks such as super-resolution reconstruction, medical image restoration, and cross-modal synthesis, demonstrating exceptional fidelity in reconstructing complex textures.

2.1 Noise Addition Process

A denoising diffusion model typically comprises two complementary components: the noise addition process and the structure reconstruction process. The noise addition process incrementally introduces noise to transform data into pure Gaussian noise, while the reconstruction process learns to recover the original data from the noisy input. These mutually opposing processes collectively form the foundational framework of diffusion probabilistic models.

The noise addition process gradually degrades the original data distribution into isotropic Gaussian noise through predefined noise scheduling parameters. The state transition equation is formulated as follows (Equation (1) [18]):

$$q(x_t \mid x_{t-1}) = N(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \quad (1)$$

where β_t denotes a predefined sequence of noise intensity coefficients, and I represents the identity matrix.

As the number of training iterations increases, the data distribution progressively transitions from its original form to a Gaussian noise distribution. After sufficient iterations, the data samples are entirely transformed into Gaussian noise, marking the completion of the forward diffusion process. This degradation mechanism is illustrated in Figure 1.

ISSN 2616-5880 Vol. 6, Issue 2: 51-56, DOI: 10.25236/AJMC.2025.060207

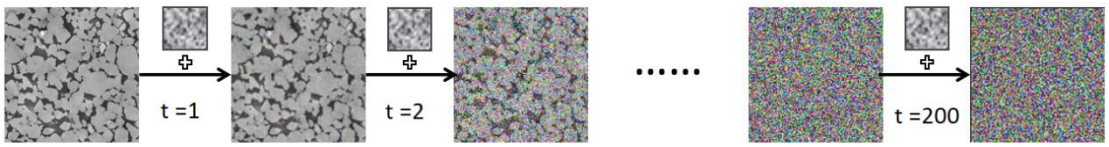

Figure 1: Schematic diagram of the noise addition process.

2.2 Structure Reconstruction Process

Following the completion of the noise addition phase, the generation of novel images requires the inverse transformation from pure noise to porous material microstructure. This critical task is accomplished through the structural regeneration process, which involves training a parameterized neural network to iteratively predict and remove noise perturbations. At each step *t*, the state transition equation is formulated as follows (Equation (2) [18]):

$$p_\theta(\mathrm{x_{t-1}} \mid \mathrm{x_t}) = N(x_{t-1}, \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \tag{2}$$

where θ denotes the model parameter, while μ_θ and Σ_θ are learnable parameters optimized through training.

By optimizing the neural network parameters θ, which govern the reverse process, the Kullback-Leibler (KL) divergence between the forward and reverse diffusion trajectories is minimized. This optimization ensures the generation of high-fidelity data samples from Gaussian noise. The schematic workflow of the reverse diffusion process is illustrated in Figure 2.

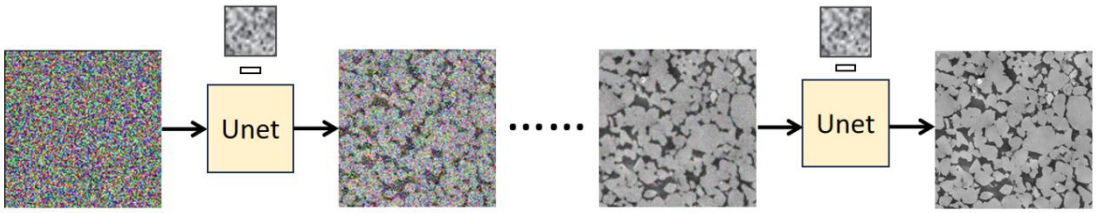

Figure 2: Schematic diagram of the structural regeneration process.

3. Model Training

3.1 Network model Architecture

In diffusion probabilistic models, the U-Net network [19] plays a crucial role. This network structure not only inherits the symmetrical encoder-decoder architecture of the traditional U-Net but also undergoes optimization and improvement in line with the characteristics of diffusion models. Firstly, this U-Net network still retains its classic "U" shape, meaning it consists of a gradually downsampling encoder part and a gradually upsampling decoder part. During the encoder stage, the network gradually extracts features from the input data through convolutional layers and pooling layers, while reducing the size of the feature maps. This helps the model capture broader contextual information. The encoder stage is composed of an initial convolutional layer with a kernel size of 7*7 and padding of 3atte, and two downsampling residual modules. Each downsampling residual module consists of two Block_klass convolutional layers, a residual layer with an attention mechanism[20], and a downsampling layer with a kernel size of 4*4, stride of 2, and padding of 1. The Block_klass convolutional layer contains two convolutional layers with a kernel size of 3*3 and padding of 1, a group normalization layer, and a SiLU activation function. In the decoder stage, the network gradually restores the size and structure of the original data through upsampling and convolution operations, thereby achieving the reconstruction or generation of the input data. The decoder stage is composed of two upsampling residual blocks, a convolutional layer with a kernel size of 3*3 and padding of 1, and a final convolutional layer with a kernel size of 1*1. Each upsampling residual module consists of two Block_klass convolutional layers similar to those in the encoder, a residual layer with an attention mechanism, and an upsampling layer with a kernel size of 4*4, stride of 2, and padding of 1. There is also a bottleneck section between the encoder and decoder. The bottleneck section is composed of two convolutional layers with a kernel size of 3*3 and padding of 1, sandwiching a residual layer with an

ISSN 2616-5880 Vol. 6, Issue 2: 51-56, DOI: 10.25236/AJMC.2025.060207

attention mechanism. The input images are 800 sandstone images with a pixel size of 64*64. They first pass through a 7*7 initial convolutional layer, where the spatial dimensions remain unchanged, but the number of channels changes from 1 to 66. Then, local features are extracted through 3*3 convolutional layers, followed by group normalization and SiLU function activation to increase the nonlinearity of the network and enable it to learn complex mapping relationships. Residual connections are introduced to help the network learn identity mappings and alleviate the vanishing gradient problem in deep networks. The attention mechanism is added to enhance the network's ability to focus on important features and improve the model's expressiveness. The spatial dimensions are reduced and the number of channels is increased through downsampling blocks. After passing through two downsampling residual blocks, the feature maps reach the bottleneck section, where the image is processed at the minimum spatial dimension, with the number of channels remaining unchanged but features being further extracted. Then, through two upsampling residual blocks, each upsampling block increases the spatial dimensions through transposed convolution (the reverse operation of downsampling) while reducing the number of channels. Other operations are the same as those in downsampling until the size is restored to be close to that of the input image. Then, the feature maps enter another Block_klass convolutional layer, which performs nonlinear transformation of the upsampling feature maps, cross-channel information exchange, and further feature fusion to help the model better extract information from high-level feature maps and then pass it to the final convolutional layer. The image passes through the final convolutional layer, where the number of channels changes from dim to out_dim, and the spatial dimensions remain unchanged. This layer maps the deep features learned by the model to the final output space. However, unlike the traditional U-Net, this U-Net also incorporates information from the time step. The network receives the time step as an additional input at each layer of the encoder and decoder, enabling the model to adjust its feature extraction and reconstruction strategies based on the current noise level. Additionally, the U-Net in the diffusion probabilistic model employs advanced techniques such as residual connections and attention mechanisms to further enhance its performance.

3.2 Training Dataset

The experimental framework employs 800 Bentheimer sandstone slices from the Digital Rocks database (800×800 pixels, 3.8 μm resolution) as the training dataset[21]. As depicted in Figure 3, these images exhibit a biphasic structure, with white regions corresponding to the solid matrix and black areas representing pore spaces. Prior to model training, a comprehensive preprocessing protocol is executed: (1) Spatial Standardization: Random cropping is applied to extract 64×64 pixel sub-images. (2) Phase Contrast Optimization: Histogram stretching enhances grayscale differentiation between pores and matrix. (3) Augmentation Strategy: Implementation of stochastic horizontal flips and constrained rotational transformations (±15°). (4) Value Normalization: Pixel intensities are linearly rescaled to the [-1, 1] interval.

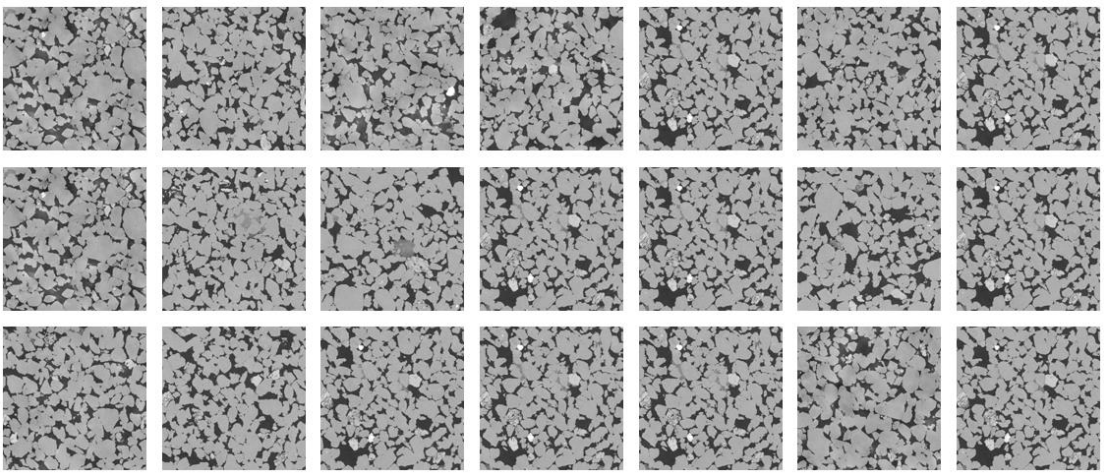

Figure 3: The microstructures of sandstone images.

ISSN 2616-5880 Vol. 6, Issue 2: 51-56, DOI: 10.25236/AJMC.2025.060207

4. Results and Analysis

4.1 Fréchet Inception Distance (FID)

The Fréchet Inception Distance (FID) serves as a robust metric for assessing the quality and diversity of generative models. Rooted in Fréchet distance theory, it quantifies the similarity between two image sets by comparing their statistical distributions in the feature space of the Inception v3 classification network. Specifically, FID computes the distance between feature vectors extracted from real and generated images, where lower scores indicate higher similarity. An ideal FID score of 0 signifies identical image distributions.

While FID demonstrates sensitivity to fine-grained details and alignment with human visual perception—making it particularly suitable for evaluating GAN-generated imagery—it exhibits notable limitations. These include dependency on input data scale and distribution, computational intensity, and inherent reliance on the pretrained Inception network's feature representation. In this study, the diffusion model achieves an FID score of 284.6933, reflecting its generative performance relative to the training data.

4.2 Learned Perceptual Image Patch Similarity (LPIPS)

LPIPS emerges as a deep learning-driven perceptual similarity metric that diverges from traditional error-based evaluations. By leveraging pretrained neural networks, it models human visual perception through hierarchical feature comparisons, capturing both global structural coherence and local textural patterns. Lower LPIPS values (ranging 0–1) denote smaller perceptual discrepancies, with applications spanning image restoration, super-resolution, medical imaging, and multimedia compression.

However, LPIPS introduces computational overhead due to its dependency on deep networks and sensitivity to training data quality. For evaluation, generated images undergo debinarization and normalization before LPIPS calculation. Random sampling of 10 synthesized images yields a mean LPIPS value of 0.165, which falls below the human perceptual difference threshold. This quantitatively confirms the high fidelity of generated microstructures in replicating authentic pore-solid textural characteristics.

5. Conclusion

This study systematically investigates diffusion model-based image generation technology, providing a comprehensive exposition of its theoretical foundations, architectural design, training methodologies, and evaluation metrics. A dedicated diffusion framework is implemented for microstructure synthesis, with rigorous assessment of generation performance. The principal findings are summarized as follows:

(1) Feasibility Validation: The experimental results substantiate the viability of diffusion models in optoelectronic porous material image generation. A machine learning-driven technical framework is established, enabling rapid synthesis of diverse photoelectric porous media with customizable pore architectures.

(2) High-Fidelity Generation: Leveraging progressive denoising mechanisms, the diffusion model demonstrates exceptional capability in producing high-resolution images that exhibit remarkable structural congruence with ground-truth samples. This approach provides an efficient data generation paradigm particularly suited for applications requiring micron-scale texture accuracy.

The presented methodology establishes a robust theoretical foundation for subsequent experimental investigations while offering critical technical references for porous media image synthesis and related interdisciplinary applications, including energy material design and biomedical scaffold optimization.

References

[1] Liu P, Chen G-F. Porous materials: processing and applications[M]. Oxford, UK: Elsevier, 2014.
[2] Guddati S, Kiran A S K, Leavy M, et al. Recent advancements in additive manufacturing technologies for porous material applications[J]. The International Journal of Advanced Manufacturing Technology, 2019, 105: 193-215.

ISSN 2616-5880 Vol. 6, Issue 2: 51-56, DOI: 10.25236/AJMC.2025.060207

[3] Ma Z C, He X H, Yan P C, et al. A fast and flexible algorithm for microstructure reconstruction combining simulated annealing and deep learning[J]. Computers and Geotechnics, 2023, 164: 105755.
[4] Feng J X, Teng Q Z, He X H, et al. Accelerating multi-point statistics reconstruction method for porous media via deep learning[J]. Acta Materialia, 2018, 159: 296 - 308.
[5] Shan L, Liu C, Liu Y, et al. Rock CT Image Super-Resolution Using Residual Dual-Channel Attention Generative Adversarial Network[J]. Energies, 2022, 15(14): 5115.
[6] Wang Y, Rahman S S, Arns C H. Super resolution reconstruction of μ-CT image of rock sample using neighbour embedding algorithm[J]. Physica A: Statistical Mechanics and its Applications, 2018, 493: 177 - 188.
[7] Siavashi J, Najafi A, Ebadi M, et al. A CNN-based approach for upscaling multiphase flow in digital sandstones[J]. Fuel, 2022, 308: 122047.
[8] Goodfellow I, Pouget-Abadie J, Mirza M, et al, Warde-Farley D, Ozair S, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems. Montreal: MIT Press, 2014: 2672-2680.
[9] Zha W, Li X, Xing Y, et al. Reconstruction of shale image based on Wasserstein Generative Adversarial Networks with gradient penalty[J]. Advances in Geo-Energy Research, 2020, 4(1): 107 - 114.
[10] Mosser L, Dubrule O, Blunt M J. Reconstruction of Three-Dimensional Porous Media Using Generative Adversarial Neural Networks[J]. Physical Review E, 2017, 96(4): 043309.
[11] Feng J, He X, Teng Q, et al. Accelerating multi-point statistics reconstruction method for porous media via deep learning[J]. Acta Materialia, 2018, 159: 296 - 308. ISSN 1359-6454.
[12] Feng J, He X, Teng Q, et al. Reconstruction of porous media from extremely limited information using conditional generative adversarial networks[J]. Physical Review E, 2019, 100(3): 033308.
[13] Shams R, Masihi M, Boozarjomehry R B, et al. Coupled generative adversarial and auto-encoder neural networks to reconstruct three-dimensional multi-scale porous media[J]. Journal of Petroleum Science and Engineering, 2020, 186: 106794. ISSN 0920-4105.
[14] Yang Y, Liu F, Yao J, et al. Multi-scale reconstruction of porous media from low-resolution core images using conditional generative adversarial networks[J]. Journal of Natural Gas Science and Engineering, 2022, 99: 104411.
[15] Xu R, Zhang D, Wang N. Uncertainty quantification and inverse modeling for subsurface flow in 3D heterogeneous formations using a theory-guided convolutional encoder-decoder network[J]. Journal of Hydrology, 2022, 613(Pt A): 128321.
[16] Feng J, Teng Q, Li B, et al. An end-to-end three-dimensional reconstruction framework of porous media from a single two-dimensional image based on deep learning[J]. Computer Methods in Applied Mechanics and Engineering, 2020, 368: 113043.
[17] Sohl-Dickstein J, Weiss E A, Maheswaranathan N, et al. Deep unsupervised learning using nonequilibrium thermodynamics[C]. International conference on machine learning. pmlr, 2015.
[18] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[J]. Journal of Machine Learning Research, 2020, 21(1): 1-68.
[19] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[EB/OL]. (2015-05-18)[2025-03-05]. https://arxiv.org/abs/1505.04597.
[20] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48 - 62.
[21] Wang Y D, Mostaghimi P, Armstrong R. A Super Resolution Dataset of Digital Rocks (DRSRD1): Sandstone and Carbonate[J]. Digital Rocks Portal, April 2019.