Powder Technology 473 (2026) 122230

Contents lists available at ScienceDirect

Powder Technology

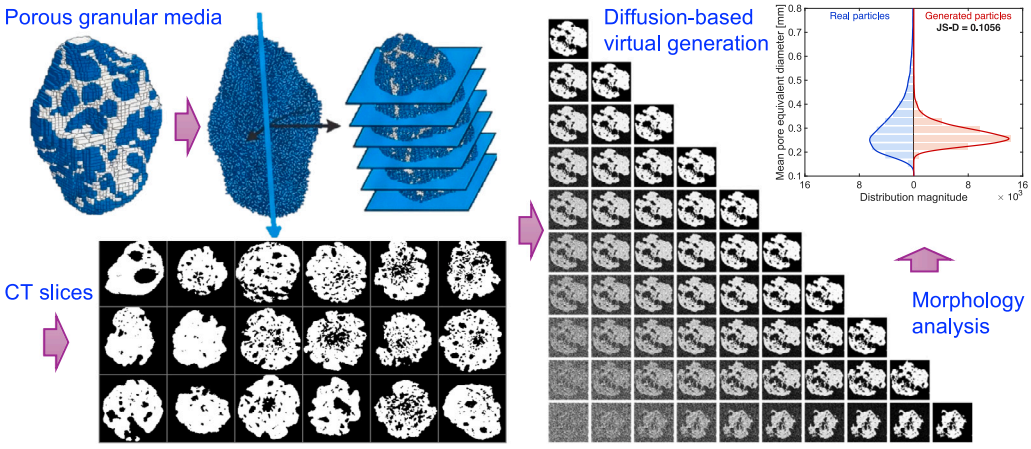
journal homepage: www.journals.elsevier.com/powder-technology

A diffusion-based generative framework for virtual porous granular media generation

Chenghao Li [a,b], Linchong Huang [a,b,c], Zhengshou Lai [a,b,c,*], Shuai Huang [a], Yuexiang Lin [b]

[a] State Key Laboratory of Tunnel Engineering, Zhuhai 519082, China
[b] School of Civil Engineering, Sun Yat-Sen University, Zhuhai 519082, China
[c] Guangdong Key Laboratory of Marine Civil Engineering, Sun Yat-sen University, Zhuhai 519082, China

GRAPHICAL ABSTRACT



HIGHLIGHTS

- A diffusion-based generative workflow is developed for virtual porous granular media generation.
- PCA-guided preprocessing standardizes orientation and enables consistent micro-CT slice datasets for training.
- Stable v-prediction DDPM training achieves high-fidelity reconstruction and unconditional generation.
- A series of morphological descriptors with KDE–JS divergence validation confirm strong statistical agreement.

ARTICLE INFO

Keywords:
Porous granular media
Diffusion model
Generative modeling
Virtual microstructure
X-ray micro-CT

ABSTRACT

The morphology of porous granular media governs transport processes and effective mechanical response in a wide range of particulate and porous materials. X-ray micro-computed tomography (micro-CT) provides direct access to internal architectures, yet experimental sampling remains limited and the resulting datasets are often insufficient for numerical simulations and parametric studies. In this work, an end-to-end diffusion-based generative framework is proposed for virtual porous media generation from image data. Within this framework, a denoising diffusion model is trained to learn the microstructure distribution and to synthesize diverse realizations through controlled sampling. This is a fundamentally different approach in that microstructures are generated by sampling from a learned distribution, rather than being constructed through handcrafted

* Correspondence to: Sun Yat-sen University, Tangjiawan, Xiangzhou District, Zhuhai, 519082, Guangdong, China
E-mail address: laizhengsh@mail.sysu.edu.cn (Z. Lai).

https://doi.org/10.1016/j.powtec.2026.122230
Received 17 December 2025; Received in revised form 21 January 2026; Accepted 1 February 2026
Available online 2 February 2026

reconstruction rules. The framework is demonstrated using coral sand as a representative example of a highly heterogeneous porous granular material. Quantitative evaluation shows that generated samples reproduce key descriptors, including porosity, pore size statistics, spatial correlation measures, and pore topology indicators. These results indicate that diffusion-based generation provides an effective and efficient route to construct statistically consistent virtual porous granular media. The proposed workflow is general and applicable to a broad class of particulate and porous materials, providing statistically consistent digital microstructures for subsequent transport modeling and mechanical analysis.

1. Introduction

Porous granular media are ubiquitous in natural and engineered systems, including soils, sediments, rocks, and biogenic particulate materials [1]. Their macroscopic mechanical and hydraulic responses are controlled by particle-scale morphology, including external grain geometry, internal pore architecture, and surface roughness [2–4]. Coral sand is a representative porous geomaterial in tropical and subtropical marine environments and is widely used in offshore infrastructure, such as reclaimed islands, breakwaters, submarine pipelines, and wind turbine foundations [5–7]. In contrast to terrigenous quartz sand, coral sand originates from fragmented skeletal remains and therefore exhibits irregular boundaries, rough surfaces, intra-granular porosity, and multiscale heterogeneity. These features lead to high compressibility, stress-dependent stiffness, crushability, and complex hydraulic conductivity [8–10]. Accurate characterization that accounts for both external geometry and internal pore structure is thus essential for understanding its micromechanical behavior.

X-ray micro-computed tomography (micro-CT) serves as a robust tool for visualizing the internal architecture of coral sand, allowing for the precise quantification of morphological features such as surface roughness, intra-granular voids, and pore connectivity [11,12]. Despite these capabilities, the widespread application of micro-CT is frequently impeded by prohibitive scanning costs, extensive acquisition and post-processing durations, and restricted scanning volumes [13,14]. Furthermore, the inherent logistical challenges associated with sample collection and high-resolution imaging inevitably limit the number of available particle samples. Consequently, existing micro-CT datasets are often too sparse to support rigorous statistical analyses, large-scale computational particle mechanics simulations, or the calibration of data-driven constitutive models. This prevailing data scarcity underscores the urgent necessity for generating large-scale, statistically representative populations of virtual coral sand particles to augment limited experimental databases and facilitate advanced particle-scale numerical modeling.

To address these challenges, computational reconstruction strategies for porous media have been widely explored [15–17]. Process based models attempt to mimic biological accretion or mechanical fragmentation, but they often depend on idealized rules and struggle to recover the multiscale pore topology typical of bioclastic grains [18,19]. In parallel, data driven generative models have become an active direction for microstructure synthesis, with generative adversarial networks (GANs) being among the most widely adopted [20–23]. For example, Zhang et al. [23] developed a three dimensional multiscale pattern GAN with scaling transformation and multiscale discrimination to enable super resolution reconstruction of porous media from limited training images. Despite these advances, GAN training can be sensitive to hyperparameters and may suffer from instability and mode collapse, which reduces reliability for quantitative engineering use [24]. In addition, adversarial objectives may under resolve fine scale geometric details, such as narrow pore throats, sharp angularities, and highly non convex voids, even though these features strongly control hydromechanical response in porous granular media.

More recently, denoising diffusion probabilistic models (DDPMs) have emerged as a state-of-the-art alternative to adversarial generation [25–27]. They learn the data distribution by reversing a progressive noising process, which leads to stable training and strong representation capacity for complex, multi modal geometries [28,29]. In materials microstructure modeling, Zhang et al. [30] developed a stable diffusion based generator and demonstrated reconstruction, inpainting, and property guided inverse design via ControlNet fine tuning. In porous media reconstruction, Naiff et al. [31] introduced a controlled latent diffusion framework that couples a variational autoencoder with diffusion sampling, and showed that conditioning on porosity can improve consistency across permeability and correlation statistics. While these studies establish diffusion models as a promising direction for microstructure synthesis, porous granular media introduce practical constraints that are rarely addressed in generic generative frameworks. First, individual grains exhibit strong pose variability and geometric anisotropy; consequently, direct learning from raw scans can be confounded by non-physical discrepancies associated with orientation and alignment. Second, the complex morphology of bioclastic grains requires a coupled generation of irregular external contours and internal pore networks, ensuring their spatial consistency. Third, for engineering applications, the generated ensembles must be rigorously validated at the population level using morphology, topology, and correlation-based statistics relevant to mechanics and transport phenomena. To address these challenges, this study presents a specialized, end-to-end framework tailored for porous granular media, rather than proposing a modification to the fundamental diffusion architecture. We leverage the robustness of the standard v-prediction diffusion formulation but integrate it into a domain-aware workflow.

Specifically, this workflow incorporates PCA-guided orientation standardization to eliminate non-physical variability, ensuring the model focuses on learning intrinsic morphological features. This is combined with a comprehensive quantitative evaluation suite that employs descriptor distributions and Jensen–Shannon divergence to verify statistical fidelity. By bridging the gap between advanced generative models and geomechanical requirements, this approach provides a reliable tool for augmenting microstructural datasets. The remainder of this paper is organized as follows. Section 2 presents the proposed framework and implementation details. Section 3 reports quantitative results on a synthetic benchmark and the representative micro CT dataset of real coral sand. Section 4 discusses implications, limitations, and extension pathways. Finally, Section 5 summarizes the main conclusions.

2. Methodologies

This section describes the workflow used in this study. It includes micro-CT dataset construction and preprocessing, diffusion-model formulation, and training and sampling settings. The methodology is designed to support stable learning and generation while preserving pore-scale geometry for subsequent quantitative evaluation and physics-relevant analysis.

2.1. Micro-CT dataset and preprocessing

As aforementioned, this study adopts natural coral sand grains as a representative example of porous granular media. Coral sand is a biogenic, highly porous granular material formed by the fragmentation of marine skeletal remains, and it is characterized by irregular external boundaries, rough surfaces, and complex intra-granular pore networks. These features make coral sand a representative yet challenging example for image-based microstructure learning. Fig. 1 illustrates

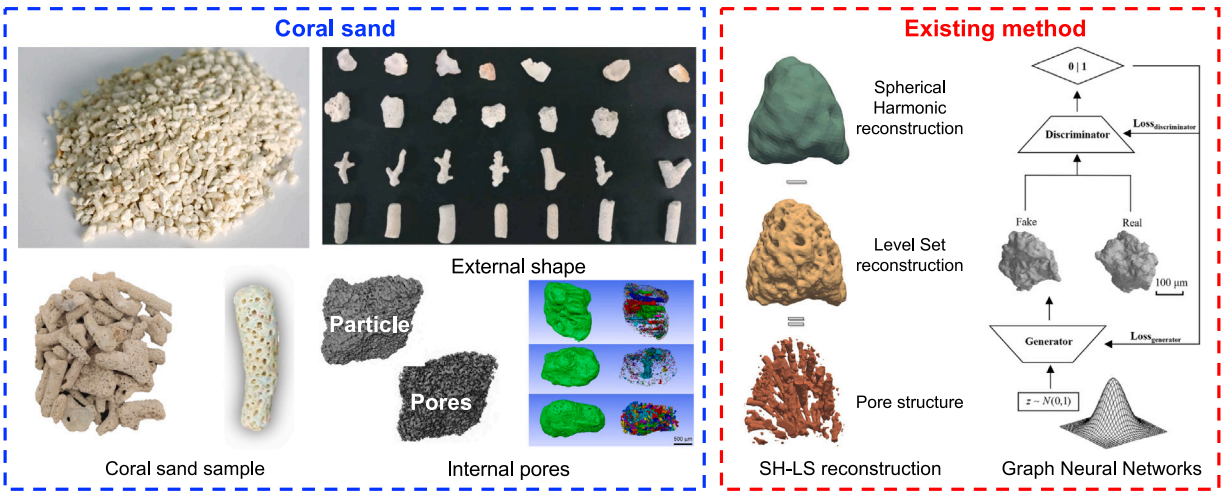


Fig. 1. Representative morphological features of natural coral sand grains and schematic illustration of two classical generation approaches [32–35].

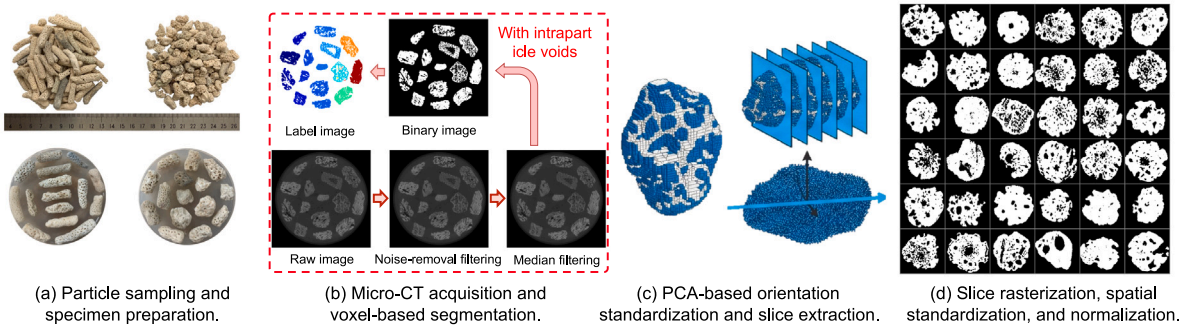


Fig. 2. Workflow for constructing the microstructural dataset from micro-CT scans of coral sand grains.

typical morphological characteristics reported in the literature, highlighting the multiscale heterogeneity that must be retained during data preparation.

The objective of dataset construction and preprocessing is to produce a standardized, orientation-consistent microstructural dataset suitable for diffusion model training. Because coral sand grains exhibit pronounced geometric anisotropy and large particle-to-particle variability, direct use of raw micro-CT volumes may introduce artificial variability associated with pose, cropping, and segmentation inconsistencies. Dedicated preprocessing is therefore performed to reduce non-physical variability while preserving intrinsic morphological information, especially fine-scale pore geometry. The overall preparation procedure is summarized as a workflow consisting of the following steps, as also instructed in Fig. 2.

(1) Micro-CT acquisition and voxel-based segmentation.

The microstructural dataset employed in this study is derived from high-resolution X-ray micro-CT imaging of natural coral sand grains. A total of 152 individual particles were selected from a larger experimental database previously established by the authors [32]. The scans were performed at a voxel resolution of 40 μm, which is sufficient to resolve both the external particle boundaries and the internal pore systems. The raw CT volumes were denoised, thresholded, and segmented to generate binary voxel representations of individual particles, where a voxel value of 1 denotes the solid phase and 0 represents the void space [36]. This binarization procedure effectively removes intensity-related artifacts while preserving the authentic geometry of particle surfaces and intra-granular cavities.

(2) PCA-based orientation standardization and slice extraction.

To ensure rotational consistency across different particles and to avoid artificial variability introduced by random grain orientations, Principal component analysis (PCA) is employed to characterize the dominant geometric orientation of each particle [37]. Considering the voxelized point cloud $\mathbf{p}_i = (x_i, y_i, z_i)$ for $i = 1, \ldots, N$, the covariance matrix is computed as

$$\mathbf{C} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{p}_i - \bar{\mathbf{p}})(\mathbf{p}_i - \bar{\mathbf{p}})^{\mathrm{T}}, \qquad \bar{\mathbf{p}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{p}_i \tag{1}$$

The eigenvector $\mathbf{e}_1$ corresponding to the largest eigenvalue defines the principal geometric axis of the particle. Instead of rotating the entire voxelized volume, a series of slicing planes orthogonal to $\mathbf{e}_1$ is constructed according to

$$(\mathbf{p} - \bar{\mathbf{p}}) \cdot \mathbf{e}_1 = d \tag{2}$$

where d denotes the signed distance along the principal axis. This strategy ensures that all two-dimensional slices are extracted along a consistent direction that maximizes the exposure of internal pore networks while maintaining geometric comparability among different particles.

(3) Slice rasterization, spatial standardization, and normalization.

Each extracted slice is rasterized using nearest-neighbor interpolation to preserve the sharp solid-void interfaces characteristic of coral sand microstructures. As the cross-sectional domain Ω_{slice} varies among particles, all slices are embedded into a fixed 128 × 128 grid using a zero-padding strategy:

$$I_{\mathrm{square}}(x, y) = \begin{cases} I(x, y), & (x, y) \in \Omega_{\mathrm{slice}}, \\ 0, & (x, y) \notin \Omega_{\mathrm{slice}}. \end{cases} \tag{3}$$

This approach avoids geometric distortions associated with image resizing and ensures uniform input dimensions required by the neural

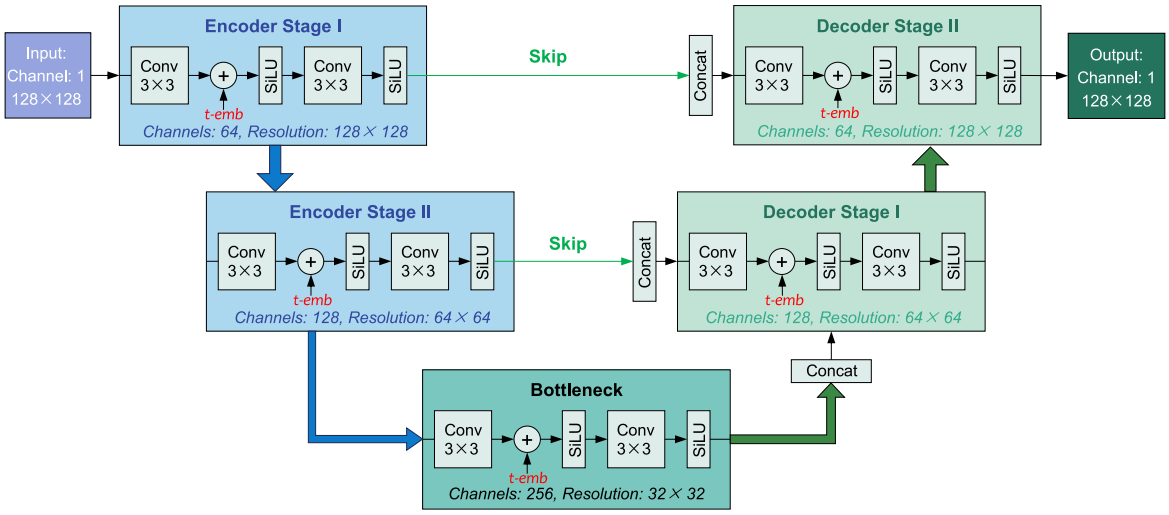


Fig. 3. Illustration of the U-Net architecture, consisting of a multi-level encoder–decoder structure with skip connections for feature fusion.

network. Finally, all voxel values are linearly scaled to the range $[-1, 1]$ to conform to the normalized input space commonly adopted in diffusion-based generative models. The resulting dataset preserves the essential morphological characteristics of coral sand, including irregular particle outlines, multiscale pore clusters, and complex pore connectivity, thereby providing a robust foundation for the diffusion model formulation described in the subsequent section.

2.2. Diffusion model implementation

In this study, we employ the DDPM [26,28] to learn the generative distribution of coral sand microstructures. To ensure numerical stability and accurately capture high-frequency geometric detail, such as sharp particle boundaries and irregular pore edges, we adopt the v-prediction parameterization [38] rather than the standard ϵ-prediction. The network is optimized by minimizing the mean squared error between the predicted and target velocity terms. Detailed mathematical formulations of the forward diffusion process and the reverse denoising kernels are provided in Appendix A.

The denoising network f_θ is parameterized using a U-Net architecture (Fig. 3), which is particularly effective for resolving the multiscale features of porous media [39]. The network employs a multi-level encoder–decoder structure: the encoder pathway extracts coarse-scale morphological context (e.g., particle elongation), while symmetric skip connections transfer high-frequency spatial details directly to the decoder to preserve fine pore boundaries. To incorporate temporal information, sinusoidal timestep embeddings are injected into each convolutional block, enabling the network to adaptively modulate feature extraction according to the noise level t.

2.3. Training and sampling

Model training is governed by the v-prediction objective, which has been shown to provide improved numerical stability and reconstruction accuracy, especially at large diffusion timesteps. Given a noisy microstructural slice $\mathbf{x}_t$ and its corresponding timestep t, the network predicts the velocity variable $\mathbf{v}_t$. Model parameters are optimized by minimizing the mean squared error between the predicted and target velocities,

$$\mathcal{L} = \mathbb{E}\left[\left\| f_\theta(\mathbf{x}_t, t) - \mathbf{v}_t \right\|_2^2\right]. \tag{4}$$

Optimization is performed using the AdamW optimizer, which is widely adopted for training diffusion-based generative models due to its robustness under noisy gradients [40]. In this work, the initial learning rate is set to 1.0×10^{-4}, and a cosine learning-rate decay schedule is applied to achieve smooth and stable convergence. All microstructural slices are normalized to the range $[-1, 1]$. To rigorously assess generalization and monitor overfitting, the complete dataset of 38,651 slices was randomly partitioned into a training set (90%) and a held-out validation set (10%). The model is trained on the training subset with a batch size of 128. The baseline training run spans 100 epochs. The diffusion horizon is fixed at $T = 1000$, consistent with the cosine noise schedule adopted in the forward diffusion process.

Table 1
Training hyperparameters used for the diffusion model.

Category	Parameter	Value/Description
Dataset	Image resolution	128 × 128 (grayscale)
	Input normalization	[−1, 1]
	Data split	90% Training/10% Validation
Optimization	Optimizer	AdamW
	Initial learning rate	1×10^{-4}
	Learning-rate schedule	Cosine decay
	Batch size	128
	Weight decay	Default
	Training epochs	100
Diffusion process	Timesteps (T)	1000
	Noise schedule	Cosine
	Prediction target	v-prediction
	Loss function	MSE ($\mathcal{L}_v$)
Stability	Precision	Mixed precision (AMP)
	Max gradient norm	1.0
	EMA decay	0.9995
Environment	Hardware	NVIDIA RTX 3090 (24 GB)
	Framework	PyTorch 2.x

To improve computational efficiency and numerical robustness, automatic mixed-precision (AMP) training is enabled [41]. Gradient clipping is applied with a maximum norm of 1.0 to mitigate occasional instabilities during optimization. An exponential moving average (EMA) of model parameters is maintained throughout training with a decay factor of 0.9995, and the EMA weights are used for sampling, as they typically produce more stable generation and sharper microstructural details. The main hyperparameters and numerical settings adopted in this work are summarized in Table 1.

After training, synthetic microstructural slices are generated by executing the reverse diffusion process from Gaussian noise. Sampling is carried out using the standard DDPM procedure, and an accelerated DDIM sampler is also employed to reduce the number of reverse steps

Table 2
Morphological and pore-related descriptors used for quantitative evaluation of porous microstructures.

Descriptor	Mathematical definition	Symbol description
Group I: Global shape descriptors		
Area A	$A = \sum_{x,y} I(x,y)$	$I(x,y)$: binary image (1 for solid, 0 for pore).
Perimeter P	$P = \sum_{(x,y)\in\partial\Omega} \Delta s$	$\partial\Omega$: particle boundary.
Roundness C	$C = \frac{4\pi A}{P^2}$	A: area; P: perimeter.
Aspect ratio AR	$\mathrm{AR} = \frac{\lambda_{\max}}{\lambda_{\min}}$	$\lambda_{\max}, \lambda_{\min}$: inertia eigenvalues.
Group II: Boundary and morphological complexity		
Convexity Conv	$\mathrm{Conv} = \frac{P_{\mathrm{convex}}}{P}$	P_{convex}: convex hull perimeter.
Fractal dimension D_f [43]	$N(s) \sim s^{-D_f}$	$N(s)$: box count at scale s.
Skeleton complexity C_s [44]	$C_s = \sum_{x,y} sk(x,y)$	$sk(x,y)$: skeletonized pixels.
Group III: Pore size and abundance		
Porosity ϕ	$\phi = \frac{N_{\mathrm{void}}}{N_{\mathrm{total}}}$	N_{void}: pore pixels.
Pore count N_{pore}	$N_{\mathrm{pore}} = \mathrm{numCC}(\mathrm{imfill}(I) - I)$	Connected pore regions.
Mean pore area $\overline{A_p}$	$\overline{A_p} = \frac{1}{N_{\mathrm{pore}}} \sum A_{p,i}$	$A_{p,i}$: area of ith pore.
Equivalent pore diameter d_{eq}	$d_{\mathrm{eq}} = \sqrt{\frac{4A_p}{\pi}}$	Circular-equivalent pore size.
Group IV: Topology and spatial statistics		
Euler number χ [45]	$\chi = C_{\mathrm{cc}} - H$	C_{cc}: components; H: holes.
Mean power spectrum $\overline{PS}$	$\overline{PS} = \frac{1}{N_{\mathrm{pix}}} \sum \lvert\mathrm{FFT}(I)\rvert$	Frequency-domain complexity.
Two-point correlation $\overline{S_2}$ [46]	$\overline{S_2} = \frac{1}{N_r} \sum_r S_2(r)$	Spatial correlation at distance r.

when needed [42]. Training progression is monitored by synthesizing a fixed panel of 36 samples after each epoch, which provides a consistent qualitative reference for tracking reconstruction fidelity and the emergence of representative pore morphology. These qualitative inspections are complemented by quantitative evaluations using morphology descriptors.

2.4. Quantitative evaluation metrics

Quantitative evaluation is performed at the descriptor level, using a set of morphology- and pore-related measures that are applicable to general porous granular microstructures. Each microstructural slice is represented as a binary image $I(x,y)$, where $I = 1$ denotes the solid phase and $I = 0$ denotes the pore space. The descriptors are computed on each slice to characterize global shape, boundary complexity, pore abundance and size, and topology or spatial statistics. The full set of descriptors and their definitions are summarized in Table 2.

Distribution-level agreement between real and generated samples is assessed by comparing the empirical distributions of each descriptor. For a given descriptor, the real-sample distribution P and the generated-sample distribution Q are estimated using kernel density estimation (KDE) from the corresponding descriptor samples [47]. The Jensen–Shannon divergence (JS-D) is then used as a bounded and symmetric measure of distribution discrepancy [48],

$$\mathrm{JS}(P \parallel Q) = \frac{1}{2}\mathrm{KL}(P \parallel M) + \frac{1}{2}\mathrm{KL}(Q \parallel M), \quad M = \frac{1}{2}(P+Q), \tag{5}$$

where $\mathrm{KL}(\cdot \parallel \cdot)$ denotes the Kullback–Leibler divergence. JS-D is computed for each descriptor, and group-level scores are reported by averaging JS-D over the descriptors within each group. This protocol evaluates both fidelity (matching of descriptor distributions) and diversity (coverage of the descriptor support) in a manner that is directly relevant to subsequent mechanics- or transport-oriented analyses.

3. Results and analysis

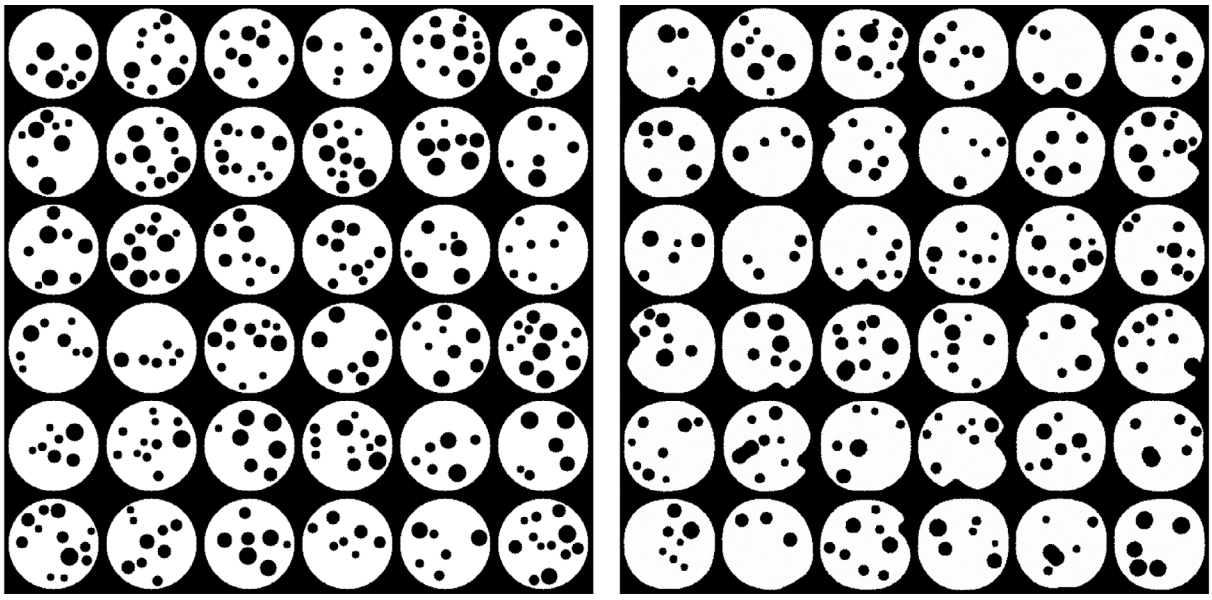
This section presents the performance of the proposed diffusion-based framework. Results are reported for a synthetic benchmark and a representative micro-CT coral sand dataset. The analysis focuses on fidelity and diversity based on morphology descriptors and distribution-level metrics, with emphasis on implications for subsequent mechanics and transport simulations.

3.1. Preliminary validation: synthetic porous disk

Prior to applying the proposed framework to the complex microstructures of coral sand, a preliminary validation study was conducted using synthetic porous disks. This phase was designed to achieve two primary objectives: (i) to verify the numerical implementation of the forward and reverse diffusion kernels; and (ii) to evaluate the model's capacity to capture geometric probability distributions from stochastically generated data under controlled conditions. To establish a controlled ground truth, a synthetic dataset was constructed using a procedural generation algorithm. Each sample comprises a standardized circular domain enclosing a variable number of circular pores (ranging from 6 to 12) with randomized radii and spatial positions, as shown in Fig. 4(a). This configuration maintains a consistent external envelope while enforcing significant internal variability. Consequently, every image constitutes a unique geometric realization, thereby compelling the model to generalize the underlying generative laws rather than merely memorizing specific training exemplars.

The generative fidelity of the trained model is first evaluated through a qualitative visual comparison between the synthetic training dataset and the generated samples, as shown in Fig. 4. The left panel presents representative ground-truth porous disks from the training set, while the right panel displays samples generated by the diffusion model upon convergence. The generated microstructures exhibit a high degree of visual similarity to the training data, accurately reproducing sharp solid–void interfaces and capturing the stochastic variability in pore size, quantity, and spatial arrangement. Particularly, the generated samples do not simply replicate the training instances; instead, they exhibit novel pore configurations, indicating that the model has successfully learned the underlying geometric probability distribution rather than memorizing the dataset.

Beyond the final outputs, the internal dynamics of the diffusion process are illustrated in Fig. 5, which visualizes the trajectories for forward diffusion, deterministic reverse denoising, and unconditional generation. The first row depicts the forward diffusion process, where Gaussian noise is progressively injected into a reference porous disk. As the diffusion steps increases, morphological features are gradually obscured until the image converges to an isotropic noise field, confirming that the forward process effectively transitions the data distribution toward the prescribed Gaussian prior. The second row illustrates the deterministic reverse denoising trajectory, in which the model iteratively reconstructs the underlying structure from a corrupted state. The gradual re-emergence of the circular boundary and



(a) Ground truth samples (b) Generated samples

Fig. 4. Visual comparison between (a) the ground truth training samples and (b) the synthetic samples generated by the diffusion model.

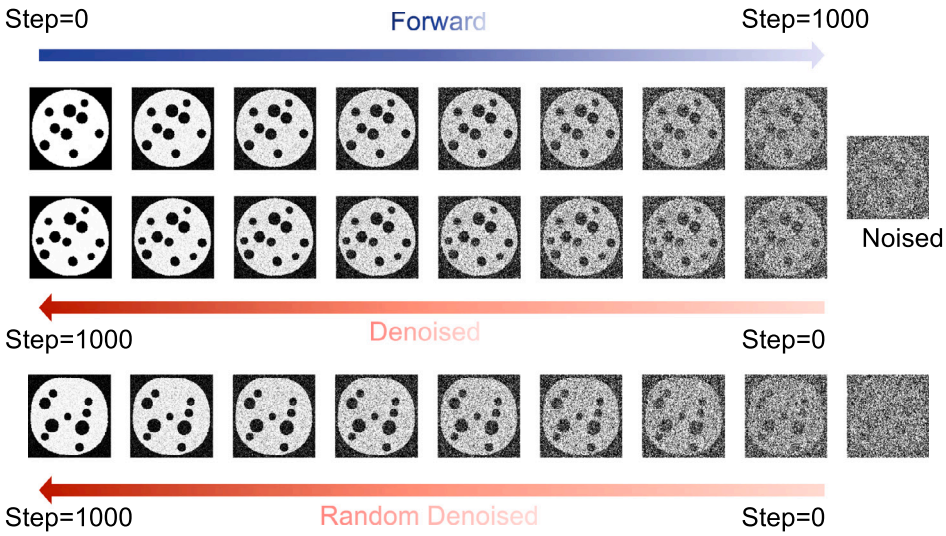
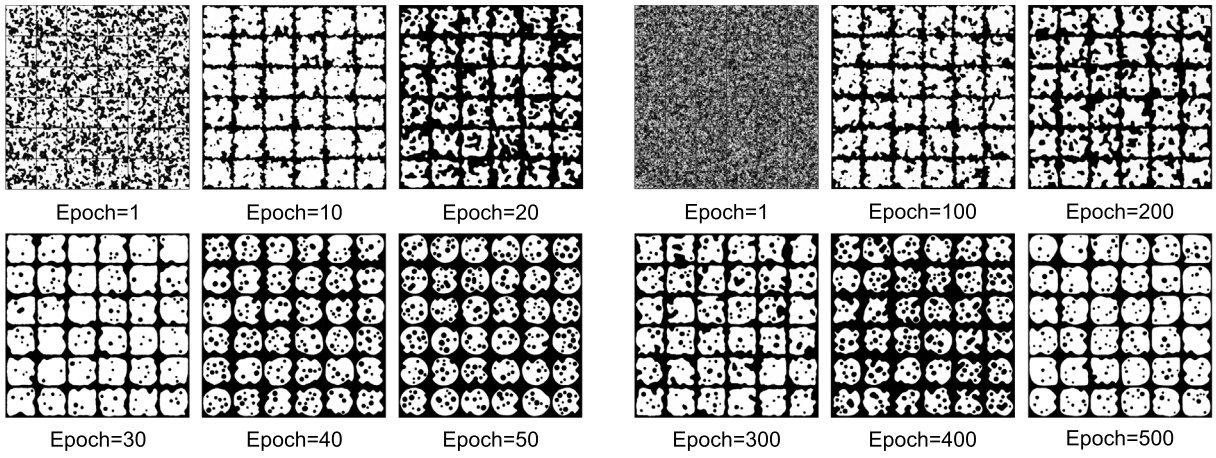


Fig. 5. Forward diffusion, reverse denoising, and unconditional generation trajectories demonstrated using the synthetic porous disk model.

pore locations demonstrates that the network has learned an accurate approximation of the reverse transition probability $p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$. Finally, the third row presents unconditional generation, where the reverse diffusion process is initialized from pure Gaussian noise. Despite the absence of an explicit structural prior, the trajectory converges to a physically plausible porous disk with a well-defined boundary and randomly distributed pores. This result confirms that the model has acquired a robust generative prior over the manifold of porous-disk microstructures.

The influence of dataset size on training stability and convergence behavior was further investigated. Two synthetic datasets were constructed, containing 2000 and 20,000 unique porous-disk images, respectively. Fig. 6 compares the evolution of generated samples across selected training epochs for these two cases. For the smaller dataset (2000 samples), early-stage generations (epochs 1, 100, and 200) are dominated by noise and degenerate artifacts; distinct porous morphologies do not consistently emerge until approximately 300–400 epochs. In contrast, the model trained on the larger dataset (20,000 samples) exhibits accelerated convergence in terms of the number of epochs. Since each epoch represents a full pass through the dataset, the model trained on the larger dataset performs ten times more gradient updates per epoch than the smaller one. Consequently, coherent disk geometries with well-defined, randomly distributed pores appear as early as epochs 30–50. This indicates that while the computational cost per epoch is higher for the larger dataset, the model requires significantly fewer epochs to capture the underlying geometric rules.

Collectively, the results from the synthetic porous-disk experiments confirm that the proposed diffusion-based framework achieves high-fidelity microstructure generation and maintains robust learning stability amidst stochastic geometric variability. This validation serves as



(a) Training samples (b) Generated samples

Fig. 6. Visual comparison between the ground truth training samples and the synthetic samples generated by the diffusion model.

a controlled and interpretable benchmark, establishing a solid methodological foundation for the subsequent application to the more complex and heterogeneous microstructures of coral sand.

3.2. Generative performance on real coral sand microstructures

Following the training phase, the generative capability of the proposed diffusion model to represent complex natural morphologies is systematically investigated using real coral sand microstructures. Prior to evaluating the generative quality, a numerical stability check was performed to verify the inversion consistency of the diffusion process; these verification results are detailed in Appendix B. With the numerical robustness of the sampling process confirmed, we proceed to evaluate the morphological fidelity and diversity of the generated ensembles.

The unconditional generative capability of the model is showcased in Fig. 7. Each row corresponds to an independent reverse diffusion trajectory initialized from pure Gaussian noise, capturing the gradual emergence of structural features throughout the denoising process. Across different realizations, the generated samples exhibit substantial variability in particle morphology, pore density, and spatial arrangement. Yet, despite this diversity, all samples consistently reproduce the defining morphological characteristics of coral sand, such as highly non-convex particle boundaries and complex intra-granular porosity. These results indicate that the model has successfully acquired a robust generative prior over the microstructural manifold of coral sand, generalizing well beyond the memorization of specific training examples.

To further evaluate the visual fidelity of the generated microstructures, a direct qualitative comparison between real and synthetic coral sand slices is presented in Fig. 8. The left panel displays a representative selection of real micro-CT slices, while the right panel presents an equivalent number of samples generated by the trained diffusion model. The synthetic samples faithfully reproduce the defining characteristics of coral sand, including rough, irregular particle boundaries, multiscale pore systems, and intricate internal topologies. Moreover, the observed sample-to-sample variability in the generated dataset closely mirrors that of the physical micro-CT data, indicating that the diffusion model has successfully captured the essential geometric distributions and pore-scale statistics. These visual results provide compelling qualitative evidence that the proposed framework is capable of synthesizing realistic and diverse coral sand microstructures, thereby establishing a robust foundation for the quantitative morphological analysis in the subsequent section.

3.3. Training convergence and parameter analysis

To assess training convergence and numerical stability, model performance is monitored using a set of complementary metrics during training. The diffusion model is optimized with the v-prediction objective, and the corresponding loss $\mathcal{L}_v$ is therefore the primary indicator of convergence. In addition, two auxiliary metrics are tracked for diagnostic purposes rather than optimization. Specifically, an ℓ_1 reconstruction error is used to quantify pixel-level recovery of the clean image, and an SSIM-based loss is used to evaluate structural fidelity that is sensitive to pore geometry and connectivity [49]. The definitions and interpretations of these metrics are summarized in Table 3.

The qualitative evolution of generated samples during the early training stage (Epochs 2–20) is shown in Fig. 9. At Epochs 2–6, the outputs are dominated by high-frequency noise, indicating that the denoiser has not yet learned a coherent solid-void representation. Between Epochs 8 and 14, continuous grain boundaries begin to emerge and the overall granular morphology becomes recognizable, whereas the intra-granular pore structure remains blurred. By Epochs 16–20, particle outlines become sharp and pore cavities become increasingly well defined, suggesting that the model has learned to represent both global geometry and fine-scale pore patterns. This visual transition is consistent with stable learning under the v-prediction parameterization for geometrically complex porous microstructures.

The quantitative evolution of the model is reported in Fig. 10, which tracks the metrics for both the training and validation datasets to assess generalization capability. The primary training objective, v-prediction loss (Fig. 10(a)), exhibits a sharp decline within the first 10 epochs and stabilizes after approximately 40 epochs. Particularly, the validation loss tracks the training loss closely with a minimal gap, indicating that the model generalizes well to unseen microstructures without suffering from significant overfitting. A similar trend is observed in the noise consistency loss (Fig. 10(b)), which serves as a proxy for the stability of the reverse diffusion trajectory. The convergence of the validation curve suggests that the learned denoising dynamics are statistically robust across the dataset. Finally, the SSIM-based loss (Fig. 10(c)) exhibits a monotonic decrease, stabilizing at approximately 0.3 for both datasets. The strong alignment between the training and validation curves confirms that the model effectively minimizes structural dissimilarity, thereby capturing the morphological fidelity of coral sand—such as pore connectivity and grain boundaries—rather than merely memorizing pixel-level noise patterns.

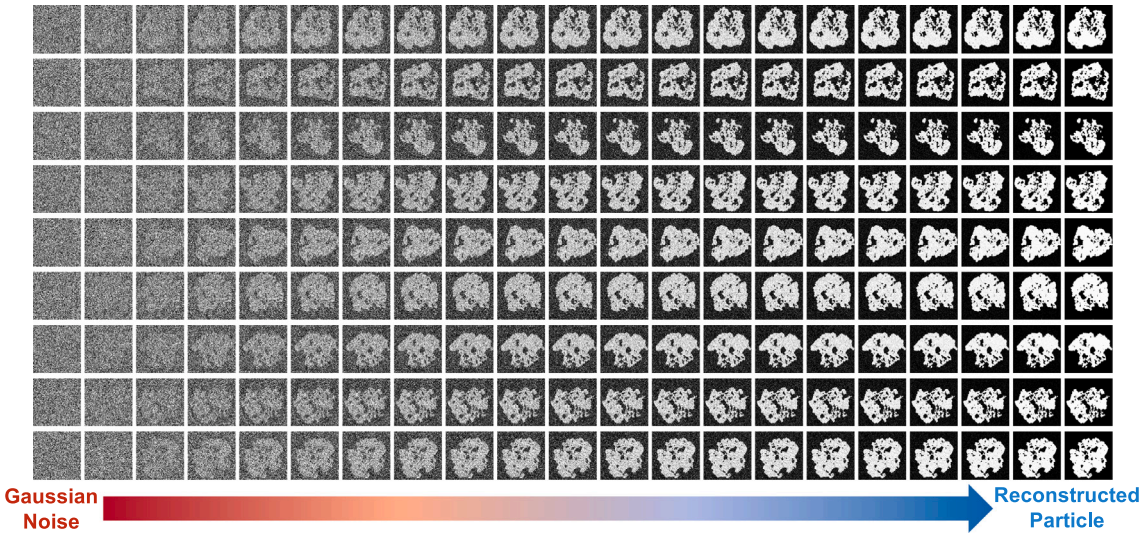
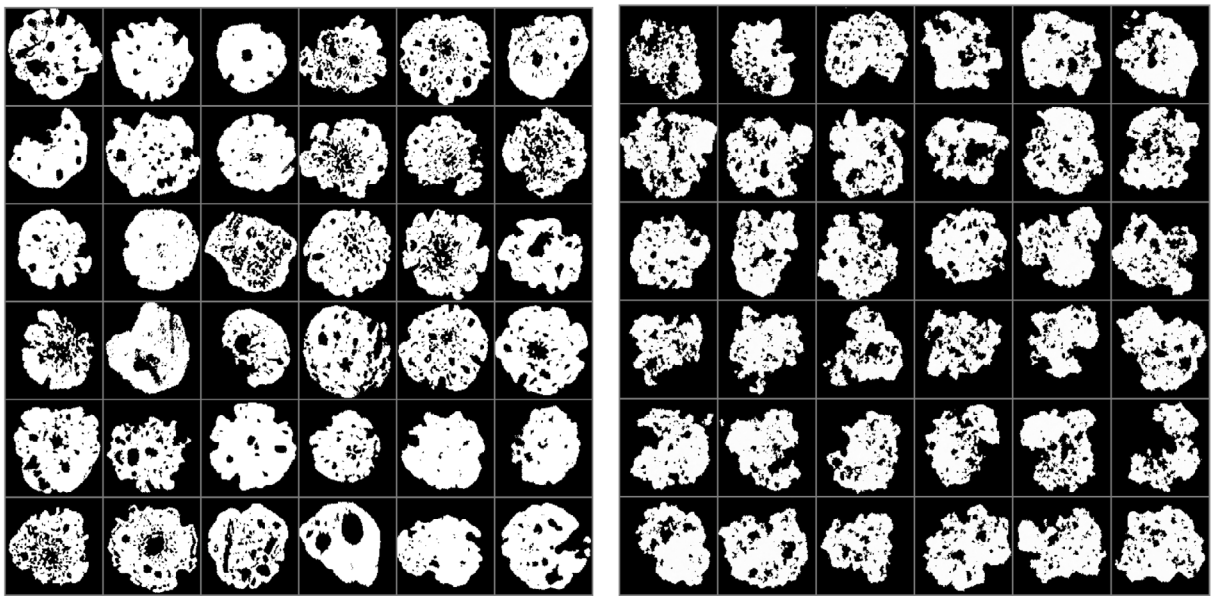


Fig. 7. Visualization of unconditional generation trajectories for coral sand microstructures. Each row depicts an independent reverse diffusion process initialized from standard Gaussian noise (left). As the denoising steps progress from left to right, the random noise is iteratively refined into structured features.



(a) Training samples (b) Generated samples

Fig. 8. Visual comparison between real micro-CT slices and randomly generated synthetic slices.

Table 3
Loss functions and monitoring metrics used for training and evaluation of the diffusion model.

Metric	Mathematical definition	Interpretation
v-prediction loss	$\mathcal{L}_v = \left\|f_\theta(\mathbf{x}_t, t) - \mathbf{v}_t\right\|_2^2$, $\mathbf{v}_t = \sqrt{\bar{\alpha}_t}\,\boldsymbol{\epsilon}_t - \sqrt{1-\bar{\alpha}_t}\,\mathbf{x}_0$	Primary optimization objective; measures accuracy of the learned reverse dynamics under the v-parameterization.
Inversion consistency loss	$\mathcal{L}_{\text{inv}} = \left\|\boldsymbol{\epsilon}_t - \hat{\boldsymbol{\epsilon}}_t\right\|_2^2$ (noise domain) $\mathcal{L}_{\text{inv}} = \left\|\mathbf{x}_0 - \hat{\mathbf{x}}_0^{(\text{inv})}\right\|_1$ (image domain)	Diagnostic measure of forward-reverse compatibility; lower values indicate that denoising predictions remain consistent with the imposed corruption statistics.
SSIM-based loss	$\mathcal{L}_{\text{SSIM}} = 1 - \text{SSIM}(\mathbf{x}, \mathbf{y})$ $\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$	Structural fidelity measure based on local statistics; sensitive to pore boundary sharpness and connectivity patterns.

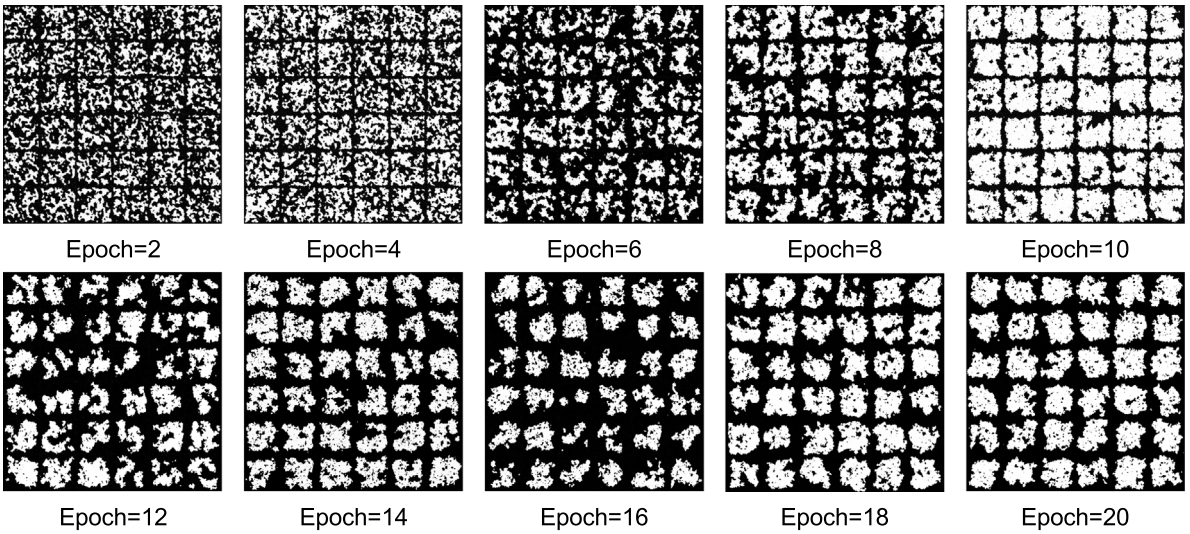
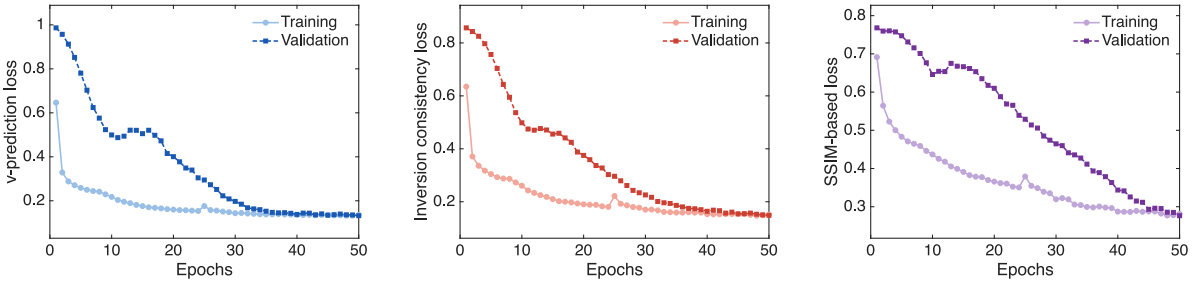


Fig. 9. Visual evolution of generated microstructures during the early stages of training (Epochs 2–20).



(a) v-prediction loss (b) Inversion consistency loss (c) SSIM-based loss

Fig. 10. Evolution of training and reconstruction metrics during diffusion model optimization.

3.4. Statistical comparison of morphological descriptors

Quantitative similarity between real and generated microstructures is assessed through statistical comparisons of morphological and pore-related descriptors. To provide a definitive answer regarding the fidelity of the generated geometry, we established a rigorous validation framework encompassing a total of 14 descriptors. These descriptors cover global particle geometry, boundary complexity, pore-scale statistics, and higher-order spatial organization. The complete descriptor set and mathematical definitions are provided in Table 2. The analysis includes 38,651 two-dimensional slices extracted from the real micro-CT dataset and an equal number of 38,651 slices synthesized by the diffusion model, which ensures a balanced and unbiased comparison representative of the population statistics.

Group I focuses on global shape descriptors, including area A, perimeter P, roundness C, and aspect ratio AR. As shown in Fig. 11, the generated distributions reproduce the overall range and the dominant modes of the real data, but non-negligible shifts are also observed. In particular, A and P exhibit moderate divergence (JS-D = 0.0665 and 0.2464), and C shows the largest discrepancy within this group (JS-D = 0.2565). The generated slices tend to populate lower roundness values and exhibit a distributional shift that is consistent with more tortuous particle outlines. In contrast, AR is matched more closely (JS-D = 0.0257), indicating that the model captures the first-order elongation statistics while still exhibiting bias in compactness-related measures.

Group II characterizes boundary roughness and morphological complexity using convexity Conv, fractal dimension D_f, and skeleton complexity C_s. These descriptors probe complementary aspects of boundary irregularity, including non-convexity, multiscale roughness, and topological branching. As shown in Fig. 12, C_s displays a close match (JS-D = 0.1065), suggesting that the model preserves the overall level of boundary branching complexity. However, both Conv and D_f show larger discrepancies (JS-D = 0.1958 and 0.1065). The generated samples shift toward lower convexity, indicating an increased prevalence of concave features, while the shift in D_f suggests that the multiscale roughness spectrum is not reproduced with the same fidelity. Taken together, the results indicate that the diffusion model captures boundary complexity in a global sense, but still exhibits bias in how irregularity is distributed across scales.

Group III evaluates pore-scale statistics, including porosity ϕ, pore count N_{pore}, mean pore area $\overline{A_p}$, and equivalent pore diameter d_{eq}. As shown in Fig. 13, the agreement is strong for $\overline{A_p}$ (JS-D = 0.0885), whereas ϕ and N_{pore} exhibit more pronounced divergence (JS-D = 0.0687 and 0.1455). The generated slices tend to shift toward higher porosity and higher pore counts, suggesting that the model slightly over-represents the prevalence of voids. At the same time, d_{eq} shows only a moderate discrepancy (JS-D = 0.1056), which indicates that the characteristic pore-size scale is largely retained. This combination implies that the discrepancy is driven more by pore multiplicity and void allocation than by a systematic inflation of pore size.

Group IV examines higher-order spatial characteristics through the Euler number χ, mean power spectrum $\overline{PS}$, and two-point correlation function $\overline{S_2}$. As shown in Fig. 14, χ is reproduced with close agreement (JS-D = 0.1195), supporting that connectivity-related topology is broadly consistent at the slice level. The match for $\overline{PS}$ is also reasonable (JS-D = 0.1360), indicating comparable frequency-domain content. In contrast, $\overline{S_2}$ exhibits a larger discrepancy (JS-D = 0.0733),

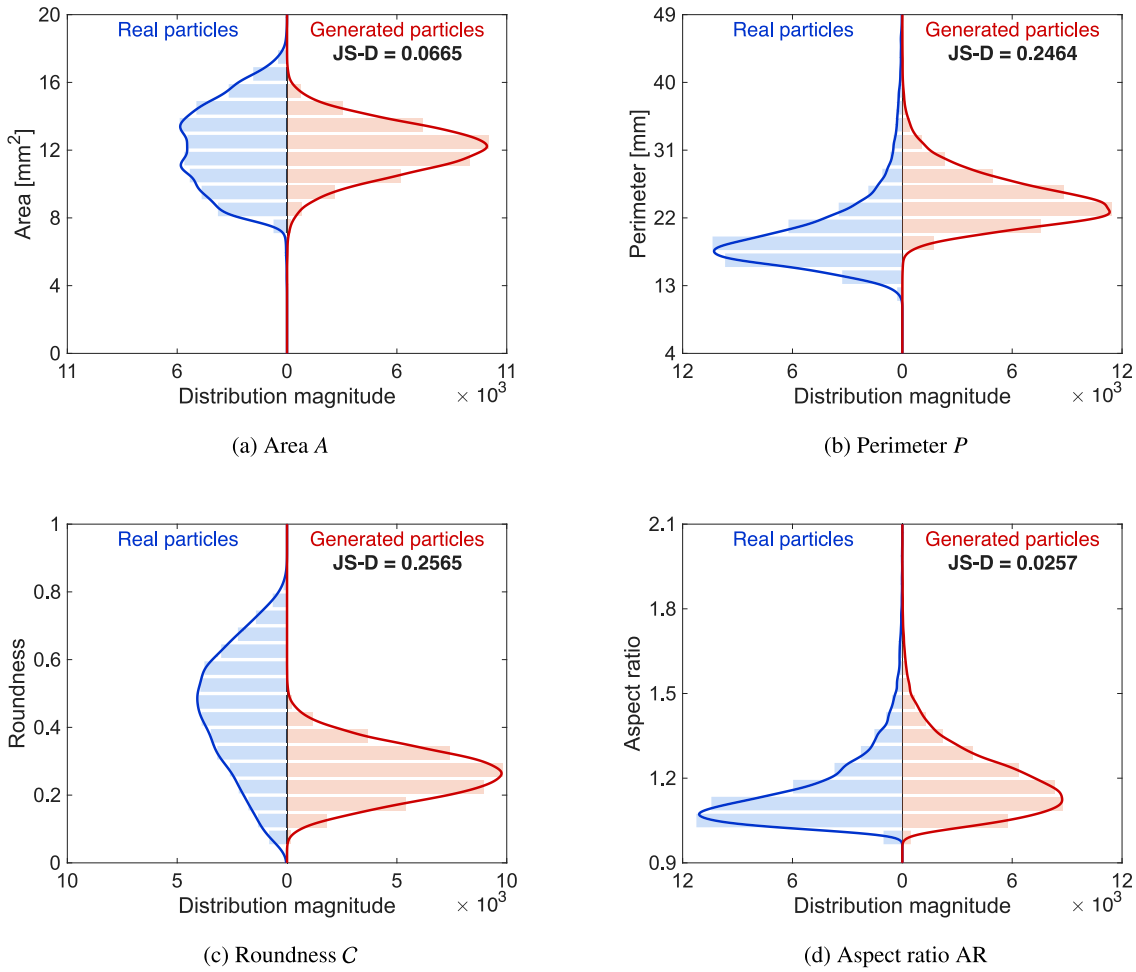


Fig. 11. Comparison of global shape descriptors (Group I).

reflecting residual differences in long-range spatial organization. This gap suggests that, while local geometry and topology are captured effectively, the model has a weaker constraint on correlation structure over larger spatial separations, which is particularly relevant for transport-representative reconstruction.

Overall, the descriptor-level comparisons indicate that the diffusion model reproduces the principal statistical characteristics of the real microstructures across geometry, boundary complexity, pore statistics, and spatial organization. The remaining discrepancies are not uniform across descriptors and are mainly concentrated in compactness- and non-convexity-related measures (e.g., C and Conv) and in long-range spatial correlation (e.g., $\overline{S_2}$). These deviations suggest that local pore geometry and slice-level topology are captured with higher fidelity than correlation structure at larger spatial separations, which motivates further physics-oriented validation and targeted model refinement.

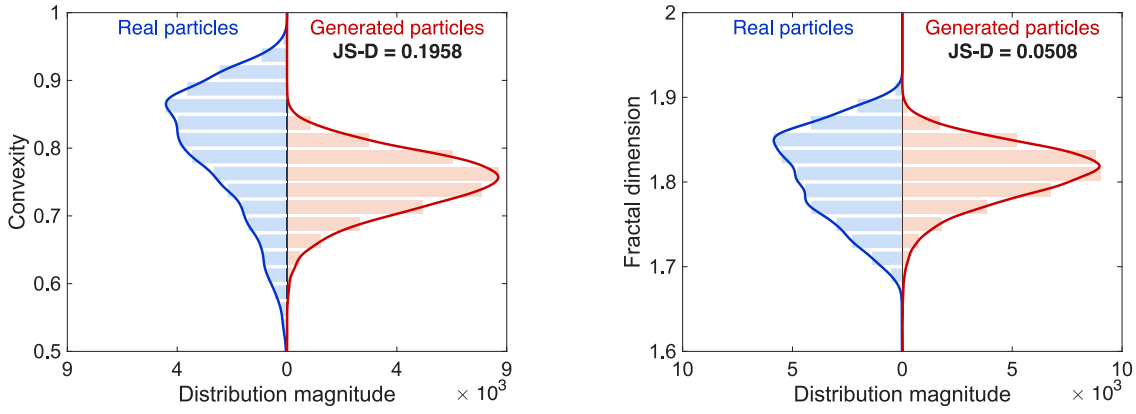
4. Discussions

4.1. Generative mechanism and comparative superiority

To better interpret the performance of the proposed framework, it is instructive to examine its underlying generative mechanism in relation to established alternatives, particularly GANs. GAN-based models formulate generation as a competitive zero-sum game between a generator and a discriminator. While conceptually appealing, this adversarial optimization is well known to be challenging in practice, frequently suffering from training instability, vanishing gradients, and mode collapse, where only a limited subset of the data distribution is reproduced [50,51]. In contrast, DDPMs recast the generation task as a sequential likelihood-based learning problem. By explicitly learning to reverse a gradual noise corruption process, diffusion models avoid the saddle-point optimization intrinsic to GANs and typically achieve more stable training behavior together with improved coverage of distribution modes [26,52].

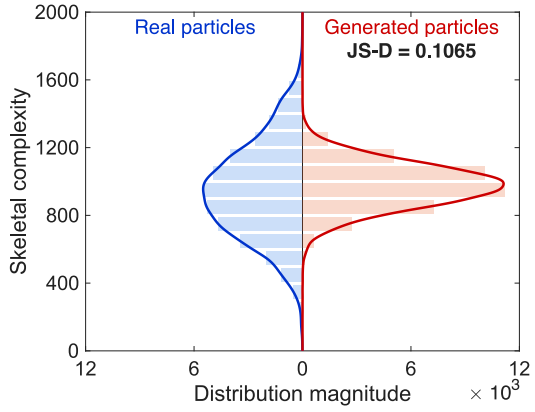
To empirically assess these conceptual differences in the context of porous granular media, a standard deep convolutional GAN (DCGAN) [53] was implemented as a baseline and trained on the same coral sand dataset. The model was intentionally configured without additional stabilization strategies in order to provide a direct and transparent comparison. Input images were binarized and linearly normalized to the range $[-1, 1]$ to match the hyperbolic tangent output of the generator. The generator mapped a 128-dimensional Gaussian latent vector to 128×128 single-channel images through a sequence of upsampling deconvolutional blocks with Batch Normalization and ReLU activations. The discriminator adopted a symmetric architecture composed of strided convolutions and LeakyReLU activations, producing scalar validity scores. Training was performed using the Adam optimizer ($\alpha = 2 \times 10^{-4}$, $\beta_1 = 0.5$, $\beta_2 = 0.999$) with a batch size of 64 over 50 epochs.

The results of the adversarial baseline are shown in Fig. 15. Although the DCGAN is capable of generating granular-like morphologies that visually resemble the training samples, the training process exhibits pronounced oscillations in discriminator confidence, reflecting the inherent instability of adversarial optimization. Moreover, while



(a) Convexity

(b) Fractal dimension D_f



(c) Skeleton complexity C_s

Fig. 12. Comparison of boundary and morphological complexity descriptors (Group II).

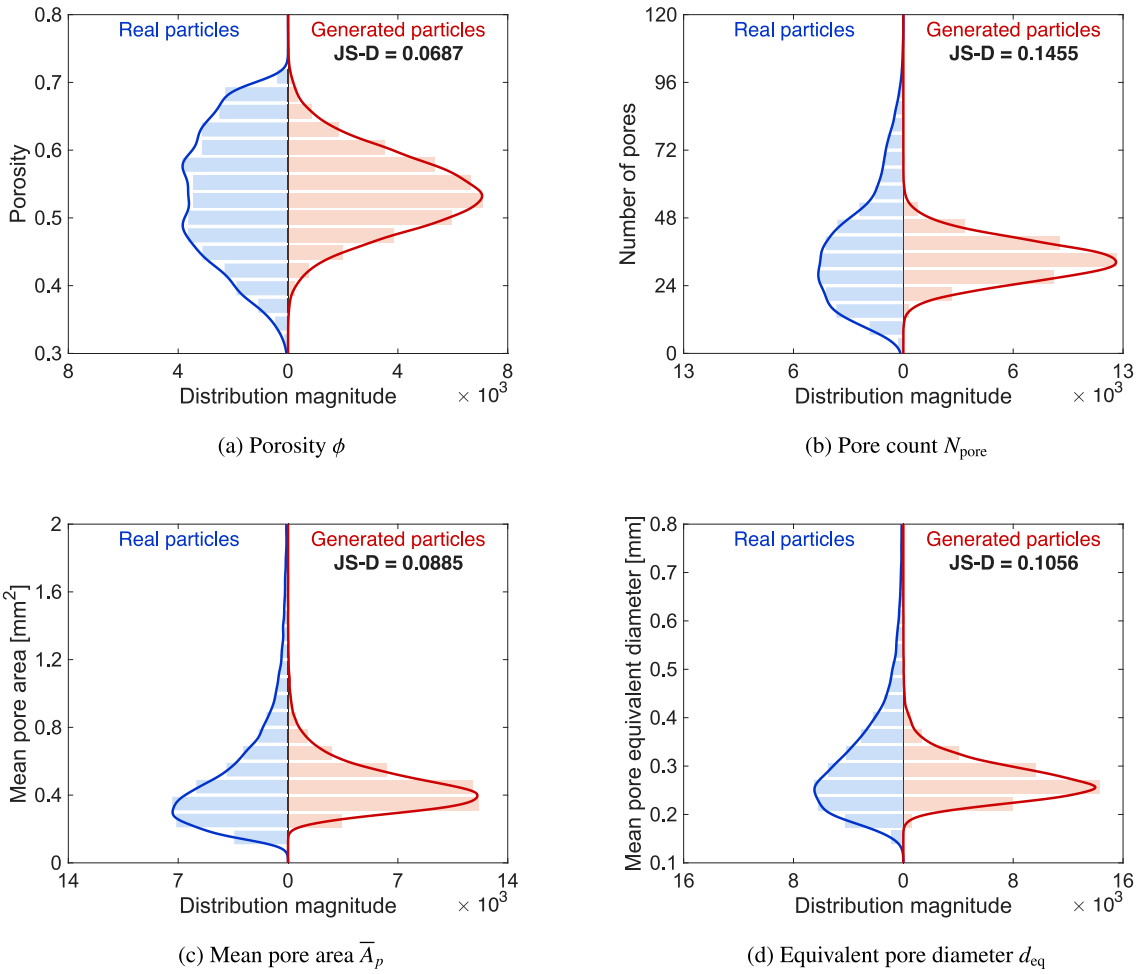
individual generated slices may appear plausible, the ensemble displays noticeable pattern repetition and limited ability to resolve the highly irregular and non-convex grain boundaries characteristic of coral sand. This behavior is largely attributable to the one-step generation paradigm of GANs, in which the complex mapping from latent space to image space must be learned directly, in contrast to the progressive, multistep refinement employed by diffusion-based models.

This behavior contrasts with the diffusion model adopted in this study, which is trained using a stable regression-based objective and naturally induces a multiscale, hierarchical generation process. Unlike the direct latent-to-image mapping employed by GANs, the forward diffusion process progressively attenuates high-frequency information, such as thin pore throats and sharp boundaries, while low-frequency components associated with the global grain geometry persist over longer noise horizons. As a result, the reverse sampling process follows a complementary coarse-to-fine trajectory. As illustrated in Fig. 16, the reconstruction of global geometry dominates the early denoising stages (large t), ensuring that the overall particle shape is established before finer-scale details emerge. Subsequent denoising steps progressively refine pore-scale textures and connectivity as the noise level approaches zero. This hierarchical dynamic contributes to the robustness of the diffusion model: under severe corruption, sampling is guided by the learned prior to generate plausible global morphologies without collapsing to a single mode, while at higher signal-to-noise ratios the model effectively resolves complex local and topological features. By combining data-driven learning with statistically consistent synthesis across multiple scales, the diffusion framework mitigates several structural limitations commonly encountered in adversarial generative approaches.

4.2. Numerical stability and parameter analysis

The choice of prediction target (ϵ-prediction v.s. v-prediction) and noise schedule (linear v.s. cosine) is critical for learning the complex topology of porous granular media. To investigate the stability of the proposed framework, we conducted comparative ablation studies focusing on the parameterization objective and the model's sensitivity to noise schedules. Initial experiments utilizing the standard noise prediction (ϵ-prediction) objective revealed significant instability. As shown in Fig. 17, models trained with this objective fail to converge to meaningful microstructures, producing only unstructured high-frequency noise artifacts even after 100 epochs. This failure persists regardless of whether a linear or cosine noise schedule is employed. These empirical results corroborate the theoretical derivation provided in Appendix A.3, which indicates that the singularity of the scaling factor $1/\sqrt{\bar{\alpha}_t}$ as $t \to T$ leads to severe variance. Consequently, the ϵ-prediction formulation proves unsuitable for the binary-like, high-contrast nature of the porous media data used in this study.

In contrast, adopting the v-prediction parameterization resolves this instability effectively. As illustrated in Fig. 18, the model successfully synthesizes coherent coral sand particles with sharp boundaries and complex internal voids under the same training duration. This confirms that the geometric rotation formulation (Eq. (A.6)) provides a numerically stable learning gradient throughout the entire diffusion horizon,



(a) Porosity ϕ

(b) Pore count N_{pore}

(c) Mean pore area $\overline{A}_p$

(d) Equivalent pore diameter d_{eq}

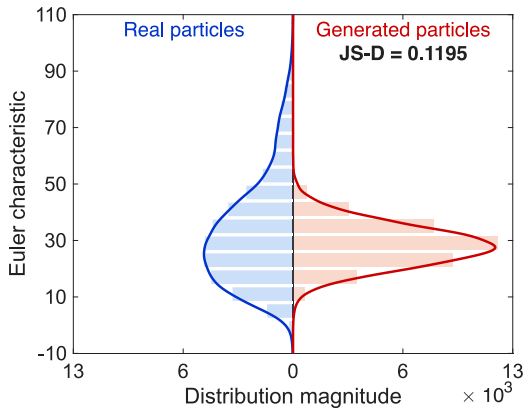
Fig. 13. Comparison of pore-scale descriptors (Group III).

allowing the network to recover structural details that are otherwise lost due to numerical precision issues in the standard formulation. With the stability of the prediction target established, we further compared the efficacy of linear versus cosine noise schedules. Fig. 18(c) presents a radar chart of the JS-D for 14 morphological descriptors. The performance curves for the linear (blue) and cosine (orange) schedules exhibit substantial overlap and closely track the center, indicating comparable generative fidelity. Unlike findings in natural image synthesis where schedule choice can drastically alter texture quality, our results suggest that for porous granular media, the generative performance is largely insensitive to the specific noise variance progression, provided that the stable v-prediction objective is used. While the linear schedule shows a marginal advantage in specific shape descriptors such as roundness, both schedules achieve statistically satisfactory alignment with real data. This robustness implies that the proposed framework is not hyperparameter-dependent and can reliably reconstruct complex pore topologies under standard settings.

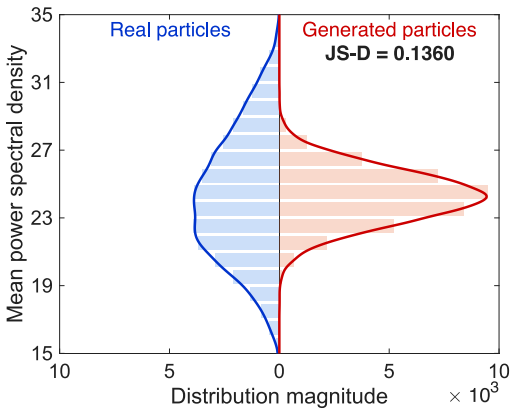
4.3. Sources of discrepancy, limitations, and future directions

Descriptor-level evaluation indicates good agreement for most geometry- and pore-size-related statistics, while the remaining discrepancies are concentrated in measures that are sensitive to compactness, non-convexity, and long-range spatial organization. In particular, deviations in roundness and convexity suggest that the generated slices exhibit a modest bias in boundary tortuosity, whereas the larger discrepancy in the two-point correlation function indicates residual differences in spatial correlation at larger separations. These gaps are consistent with the fact that long-range correlation and topology are weakly constrained by local pixel-wise denoising objectives and can be sensitive to small perturbations in binary interfaces. In addition, connectivity-related measures computed on 2D slices may not fully reflect the underlying three-dimensional pore network, and slice-to-slice variability can amplify apparent differences even when local pore-size statistics are matched.

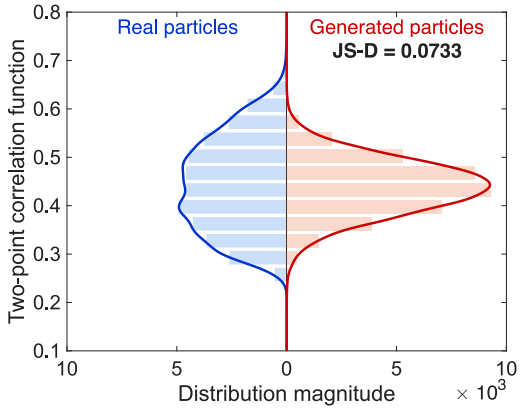
These limitations also clarify practical directions for extending the present framework. First, topology- and correlation-sensitive fidelity can be strengthened by incorporating additional physics- or descriptor-informed constraints during training or sampling, for example by augmenting the objective with correlation-aware regularization or by adopting conditional generation that controls selected global statistics. Second, the transition from two-dimensional slices to three-dimensional volumes is necessary for applications where permeability, tortuosity, and connectivity govern macroscopic behavior. A natural extension is to employ three-dimensional diffusion models directly or latent diffusion strategies to reduce computational cost while preserving three-dimensional structure. Finally, the most stringent validation for engineering usage is physics-relevant verification. Coupling the generated microstructures with downstream mechanics or transport simulations would quantify whether descriptor-level agreement translates into consistent effective properties, and would provide guidance on which descriptors require tighter control for specific applications.



(a) Euler number χ



(b) Mean power spectrum $\overline{PS}$



(c) Two-point correlation function $\overline{S_2}$

Fig. 14. Comparison of higher-order spatial descriptors (Group IV).

5. Summary

This study developed a diffusion-based generative framework for reconstructing and synthesizing porous granular media, using coral sand micro-CT images as a representative case. The workflow combines PCA-guided orientation standardization with a v-prediction diffusion model to learn reverse denoising dynamics from high-resolution data, enabling both reconstruction from corrupted inputs and unconditional generation from Gaussian noise. The results show stable training and effective recovery of multiscale features, including irregular grain boundaries and intra-granular pore networks, across different noise levels. Particularly, a rigorous statistical evaluation based on 14 morphological and pore-related descriptors confirms that the generated geometries accurately mimic the real-world samples. Good agreement is observed between real and generated datasets in terms of global geometry, pore-size characteristics, and solid-void organization, while the remaining discrepancies are mainly associated with topology- and correlation-sensitive measures. These outcomes support diffusion models as a robust and scalable tool for augmenting limited microstructural datasets and facilitating subsequent geomechanical analyses. Future work will extend the framework to three-dimensional generation, introduce conditioning for property-guided synthesis, and integrate the generated microstructures into discrete element method- and computational fluid dynamics-based simulations to assess mechanical and hydraulic responses.

CRediT authorship contribution statement

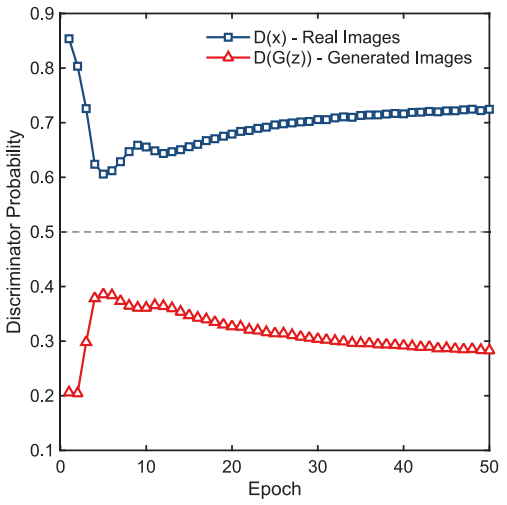
Chenghao Li: Writing – original draft, Software, Formal analysis, Data curation. **Linchong Huang:** Project administration, Funding acquisition. **Zhengshou Lai:** Writing – review & editing, Methodology, Investigation. **Shuai Huang:** Resources, Investigation, Data curation. **Yuexiang Lin:** Resources, Investigation.

Declaration of competing interest

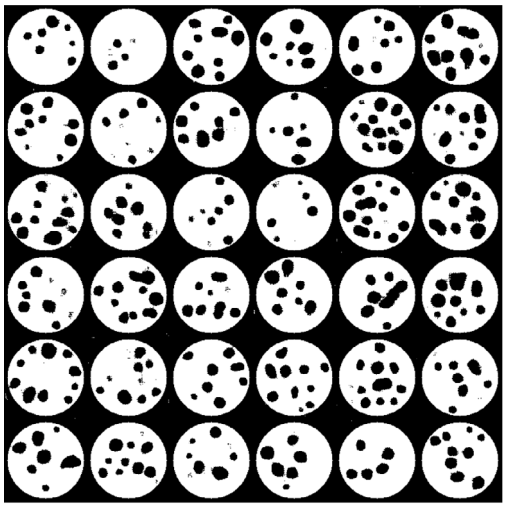
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

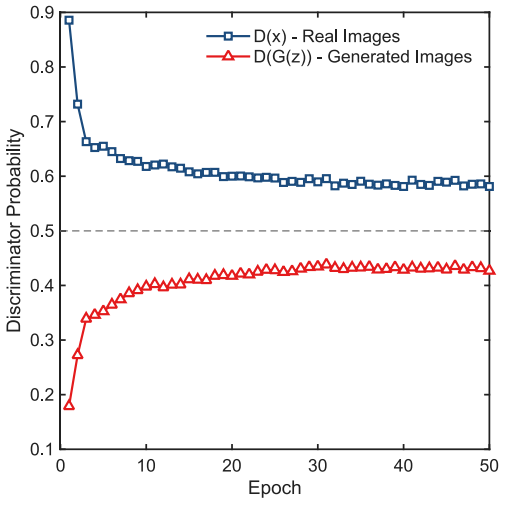
This work was supported by the Guangdong Major Project of Basic Research (2025B0303000011), the National Natural Science Foundation of China (52478428), the Guangdong Basic and Applied Basic Research Foundation (2024A1515012179), and the Shenzhen Kunpeng Young Innovative and Entrepreneurial Project (KPGR-2025073118 3203033).



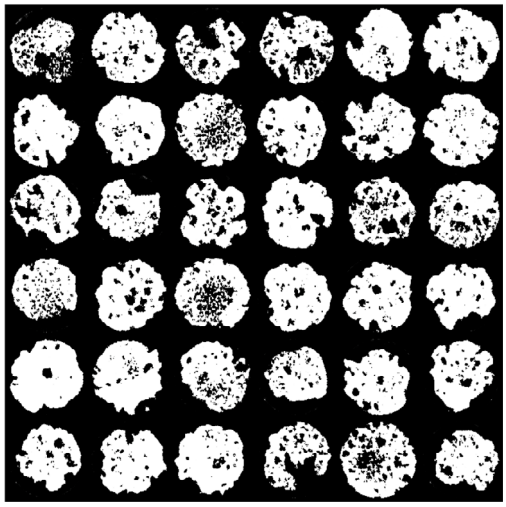
(a) Discriminator probability (Circle dataset)



(b) Generated slices (Circle dataset)



(c) Discriminator probability (Sand dataset)



(d) Generated slices (Sand dataset)

Fig. 15. Comparison between real micro-CT slices and randomly generated synthetic slices using DCGAN: (a)–(b) results for the idealized circular dataset, and, (c)–(d) results for the real coral sand dataset.

Appendix A. Formulations of the diffusion framework

This appendix provides the detailed mathematical formulation of the diffusion framework employed in this study. It covers the definitions of the forward and reverse processes, the mathematical specification of the noise schedules, and the algebraic derivation of the reconstruction formulas that justify the stability of the v-prediction parameterization.

A.1. Forward diffusion process and noise schedules

As illustrated in Fig. A.19, the forward diffusion process is defined as a fixed Markov chain that gradually corrupts the clean data $\mathbf{x}_0$ by adding Gaussian noise. The transition kernel at each timestep t is given by [54]:

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{1-\beta_t}\,\mathbf{x}_{t-1}, \beta_t \mathbf{I}\right) \tag{A.1}$$

where $\mathbf{x}_t \in \mathbb{R}^d$ denotes the vectorized image state, $\mathbf{I}$ is the identity matrix, and $\beta_t \in (0,1)$ represents the variance schedule. A key property of this process is that the marginal distribution at an arbitrary timestep t can be sampled in closed form:

$$q(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\,\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I}\right), \quad \text{with} \quad \bar{\alpha}_t = \prod_{i=1}^{t}(1-\beta_i) \tag{A.2}$$

The choice of the noise schedule β_t governs the rate of information destruction. In this study, we compare the standard linear schedule with the cosine schedule [29] to analyze their impact on microstructural preservation. The linear schedule defines β_t via linear interpolation:

$$\beta_t = \beta_{\min} + \frac{t}{T}(\beta_{\max} - \beta_{\min}) \tag{A.3}$$

where the parameters are typically set with $\beta_{\min} = 10^{-4}$ and $\beta_{\max} = 0.02$. While standard for natural images, this schedule tends to destroy fine-scale geometric details too rapidly in the early stages. In contrast, the cosine schedule is defined directly on the cumulative signal variance $\bar{\alpha}_t$ to ensure a smoother degradation profile:

$$\bar{\alpha}_t = \frac{f(t)}{f(0)}, \quad f(t) = \cos^2\left(\frac{t/T + s}{1+s} \cdot \frac{\pi}{2}\right) \tag{A.4}$$

where $s = 0.008$ is a small offset to prevent singularities near $t = 0$. While our ablation study (discussed in Section 4.2) suggests that

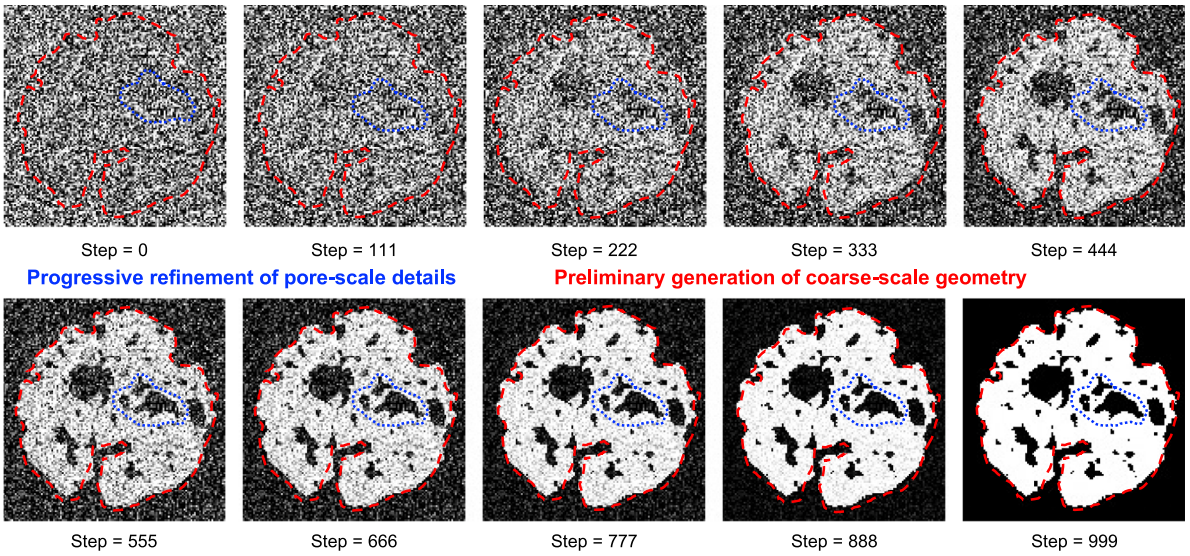
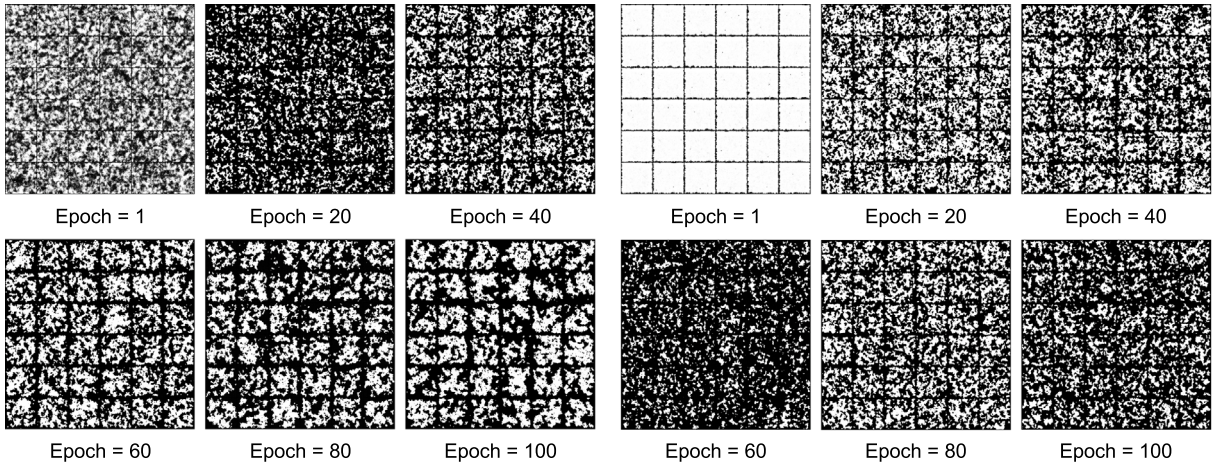


Fig. 16. Multiscale coarse-to-fine evolution of porous grain morphology during diffusion sampling.



(a) Linear schedule ($\boldsymbol{\varepsilon}$-prediction) (b) Cosine schedule ($\boldsymbol{\varepsilon}$-prediction)

Fig. 17. Visual demonstration of training instability using the standard noise prediction ($\boldsymbol{\varepsilon}$-prediction). Regardless of the noise schedule, the model fails to converge, producing only high-frequency noise artifacts.

the linear schedule is also effective for this specific dataset, the cosine schedule provides a smoother degradation of information in the theoretical limit and remains the standard choice in the literature. Therefore, to ensure broad applicability and align with established baselines, the cosine schedule is adopted as the default configuration for the primary results in this study.

A.2. Reverse denoising process

The reverse process aims to invert the corruption chain and is modeled as a parameterized Gaussian transition:

$$p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t) = \mathcal{N}\left(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_t\right) \tag{A.5}$$

To enhance numerical stability, particularly for geometries with sharp interfaces, we adopt the v-prediction parameterization [38]. The training target $\mathbf{v}_t$ is defined as a rotation of the clean signal and the noise in the spherical latent space:

$$\mathbf{v}_t \equiv \sqrt{\bar{\alpha}_t}\,\boldsymbol{\varepsilon}_t - \sqrt{1-\bar{\alpha}_t}\,\mathbf{x}_0 \tag{A.6}$$

The neural network f_θ is trained to predict this velocity vector, i.e., $\hat{\mathbf{v}}_t = f_\theta(\mathbf{x}_t, t)$, by minimizing the mean squared error.

A.3. Derivation of reconstruction formulas

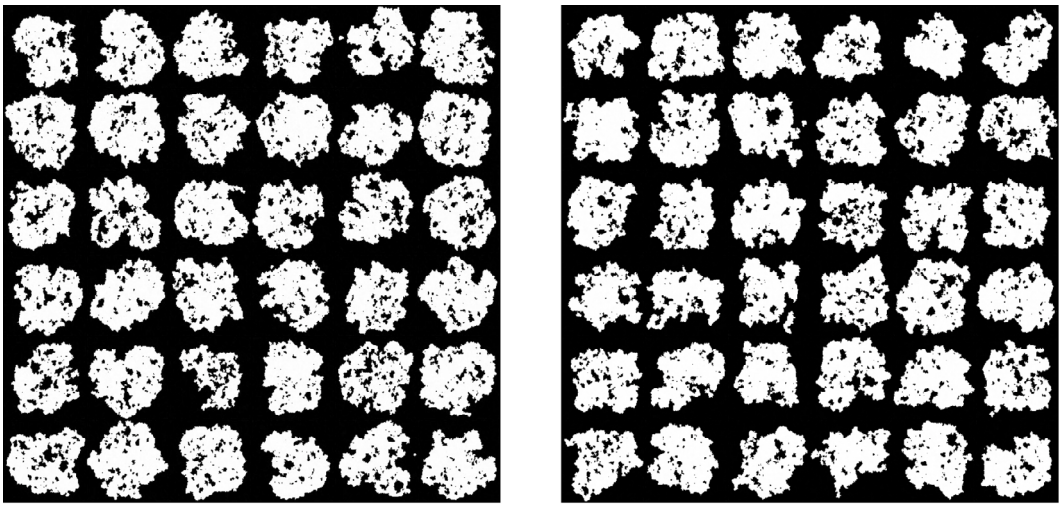
This subsection provides the algebraic derivations for recovering the clean image estimate $\hat{\mathbf{x}}_0$ from the network output, highlighting the numerical characteristics of different parameterizations.

Instability of noise prediction (ε-prediction)

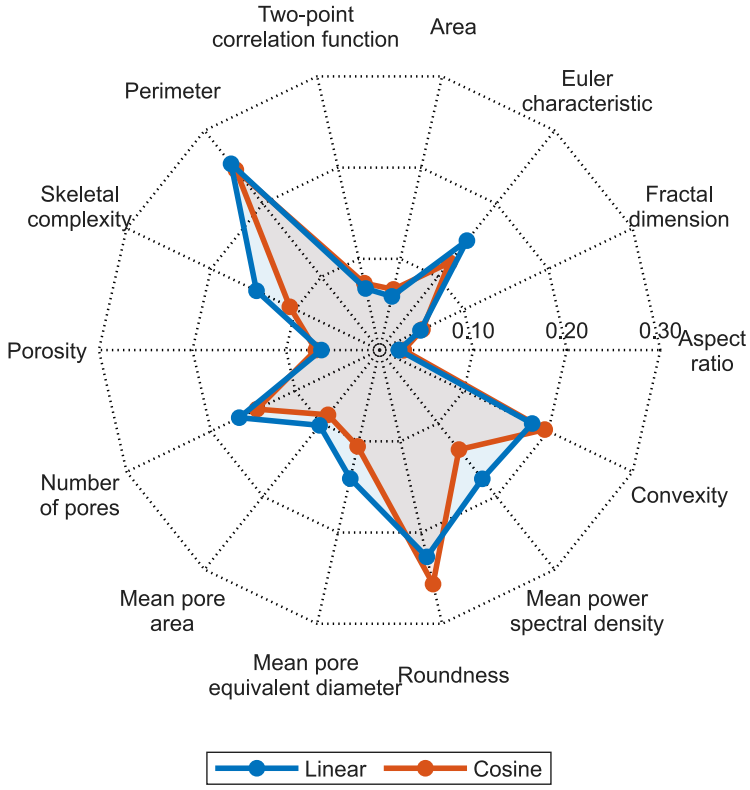
In the standard formulation [26], the forward state is expressed as $\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\,\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\varepsilon}_t$. Rearranging this to solve for $\mathbf{x}_0$ given a noise prediction $\hat{\boldsymbol{\varepsilon}}_t$ yields:

$$\hat{\mathbf{x}}_0 = \frac{\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\,\hat{\boldsymbol{\varepsilon}}_t}{\sqrt{\bar{\alpha}_t}} \tag{A.7}$$

From a numerical perspective, as $t \to T$, the signal-to-noise ratio approaches zero and $\bar{\alpha}_t \to 0$. Consequently, the scaling factor $1/\sqrt{\bar{\alpha}_t}$ in Eq. (A.7) becomes unbounded. This singularity amplifies any small prediction errors in $\hat{\boldsymbol{\varepsilon}}_t$, leading to significant variance and training instability in the terminal diffusion stages.



(a) Linear schedule (v-prediction) (b) Cosine schedule (v-prediction)



(c) JS-Divergence of morphological descriptors

Fig. 18. Quantitative and qualitative comparison of noise schedules under the stable v-prediction: (a) linear schedule and (b) cosine schedule; and (c) radar chart of JS-D (lower is better).

Stability of velocity prediction (v-prediction)

For the v-prediction parameterization, let $c_t = \sqrt{\bar{\alpha}_t}$ and $s_t = \sqrt{1-\bar{\alpha}_t}$. The relationship between the observation $\mathbf{x}_t$ and the target $\mathbf{v}_t$ forms a linear system:

$$\begin{cases} \mathbf{x}_t = c_t \mathbf{x}_0 + s_t \boldsymbol{\epsilon}_t & \text{(Observation)} \\ \mathbf{v}_t = c_t \boldsymbol{\epsilon}_t - s_t \mathbf{x}_0 & \text{(Definition)} \end{cases} \tag{A.8}$$

To solve for $\mathbf{x}_0$, we eliminate $\boldsymbol{\epsilon}_t$ by multiplying the first equation by c_t and the second by s_t, then subtracting:

$$c_t \mathbf{x}_t - s_t \mathbf{v}_t = (c_t^2 + s_t^2)\mathbf{x}_0 = \mathbf{x}_0 \tag{A.9}$$

Thus, the reconstruction formula is:

$$\hat{\mathbf{x}}_0 = \sqrt{\bar{\alpha}_t}\,\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\,\hat{\mathbf{v}}_t \tag{A.10}$$

Similarly, the noise component is recovered as:

$$\hat{\boldsymbol{\epsilon}}_t = \sqrt{1-\bar{\alpha}_t}\,\mathbf{x}_t + \sqrt{\bar{\alpha}_t}\,\hat{\mathbf{v}}_t \tag{A.11}$$

These derivations confirm that v-prediction allows for the recovery of the clean signal via a pure geometric rotation (linear combination), avoiding the division by zero inherent in ϵ-prediction. This property ensures consistent numerical behavior across all noise levels, making it particularly suitable for learning the complex, high-frequency boundaries of porous granular media.

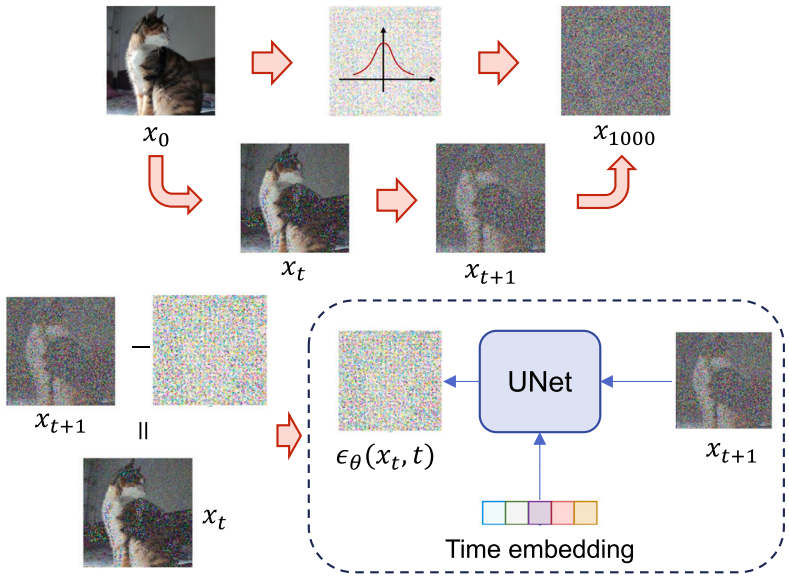


Fig. A.19. Schematic illustration of the forward (noise addition) and reverse (denoising) processes in the diffusion model.

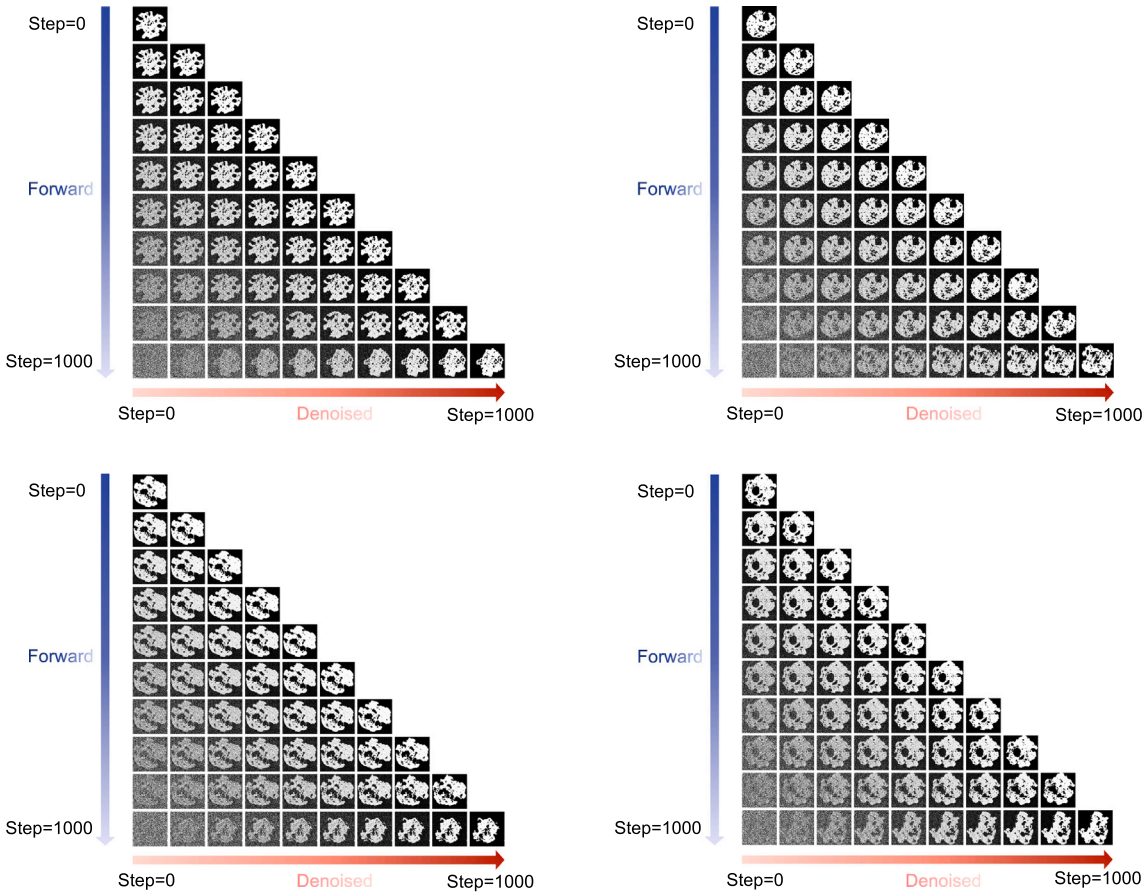


Fig. B.20. Visualization of forward noise corruption and conditional reverse reconstruction trajectories. The vertical axis represents increasing noise levels, while the horizontal axis shows the denoising progress.

Appendix B. Inversion consistency check

This appendix presents the validation of the diffusion model's forward-reverse consistency. Fig. B.20 presents a matrix visualization illustrating the relationship between forward noise corruption and conditional reverse reconstruction for a representative two-dimensional coral sand slice. The first column depicts the forward diffusion process. Starting from the original microstructure (upper-left corner), Gaussian noise is progressively added according to the cosine noise schedule, resulting in the gradual obscuration of particle boundaries and intra-granular pore networks. For each corruption level shown in the vertical axis, the corresponding reverse denoising trajectory is computed and visualized along the horizontal direction.

At low-to-moderate corruption levels, the conditional reverse trajectories exhibit high-fidelity reconstruction. In these cases, irregular particle outlines and major intra-particle voids are recovered with high accuracy, closely matching the reference slice. Intermediate denoising steps reveal a hierarchical reconstruction behavior, where large-scale geometric features are first re-established, followed by the gradual refinement of fine-scale pore structures. As the initial corruption level increases, the reconstruction behavior shifts. While the reverse trajectories still converge toward physically realistic coral sand particles, the final pore configurations do not exactly replicate the reference microstructure. This deviation reflects the intrinsic stochastic nature of diffusion models: when the original signal is dominated by noise, multiple microstructural realizations become statistically compatible with the corrupted state. In such instances, the model relies on its learned generative prior to synthesize plausible morphological details rather than performing an exact inversion. Overall, the result demonstrates that the proposed model achieves stable and accurate conditional reconstruction under moderate noise levels while maintaining physically meaningful generative behavior under severe corruption.

Data availability

Data will be made available on request.

References

[1] H.M. Jaeger, S.R. Nagel, R.P. Behringer, Granular solids, liquids, and gases, Rev. Modern Phys. 68 (4) (1996) 1259.

[2] S. Rui, Z. Guo, T. Si, Y. Li, Effect of particle shape on the liquefaction resistance of calcareous sands, Soil Dyn. Earthq. Eng. 137 (2020) 106302.

[3] G. Cho, J. Dodds, J.C. Santamarina, Particle shape effects on packing density, stiffness, and strength: natural and crushed sands, J. Geotech. Geoenvironmental Eng. 132 (5) (2006) 591–602.

[4] K. Tao, H. Konietzky, F. Gao, R. Wu, W. Dang, C. Li, Y. Liu, Stick–slip friction and surface contact density of bare granite joint affected by cyclic normal stress, Rock Mech. Rock Eng. (2025) 1–21.

[5] X. Wang, Y. Jiao, R. Wang, M. Hu, Q. Meng, F. Tan, Engineering characteristics of the calcareous sand in Nansha Islands, south China sea, Eng. Geol. 120 (1–4) (2011) 40–47.

[6] X. Wang, X. Wang, J. Shen, H. Ding, D. Wen, C. Zhu, S. Lv, Foundation filling performance of calcareous soil on coral reefs in the south China sea, Appl. Ocean Res. 129 (2022) 103386.

[7] Y. Wu, N. Li, X. Wang, J. Cui, Y. Chen, Y. Wu, H. Yamamoto, Experimental investigation on mechanical behavior and particle crushing of calcareous sand retrieved from south China sea, Eng. Geol. 280 (2021) 105932.

[8] M. Coop, The mechanics of uncemented carbonate sands, Géotechnique 40 (4) (1990) 607–626.

[9] H. Wei, T. Zhao, J. He, Q. Meng, X. Wang, Evolution of particle breakage for calcareous sands during ring shear tests, Int. J. Geomech. 18 (2) (2018) 04017153.

[10] Y. Xiao, H. Liu, X. Ding, Y. Chen, J. Jiang, W. Zhang, Influence of particle breakage on critical state line of rockfill material, Int. J. Geomech. 16 (1) (2016) 04015031.

[11] R.A. Ketcham, W.D. Carlson, Acquisition, optimization and interpretation of X-ray computed tomographic imagery: applications to the geosciences, Comput. Geosci. 27 (4) (2001) 381–400.

[12] J. Desrues, G. Viggiani, P. Besuelle, Advances in X-ray Tomography for Geomaterials, vol. 118, John Wiley & Sons, 2010.

[13] V. Cnudde, M.N. Boone, High-resolution X-ray computed tomography in geosciences: A review of the current technology and applications, Earth-Sci. Rev. 123 (2013) 1–17.

[14] S. Schlüter, A. Sheppard, K. Brown, D. Wildenschild, Image processing of multiphase images obtained via X-ray microtomography: a review, Water Resour. Res. 50 (4) (2014) 3615–3639.

[15] M.J. Blunt, B. Bijeljic, H. Dong, O. Gharbi, S. Iglauer, P. Mostaghimi, A. Paluszny, C. Pentland, Pore-scale imaging and modelling, Adv. Water Resour. 51 (2013) 197–216.

[16] H. Andrä, N. Combaret, J. Dvorkin, E. Glatt, J. Han, M. Kabel, Y. Keehm, F. Krzikalla, M. Lee, C. Madonna, et al., Digital rock physics benchmarks—Part I: Imaging and segmentation, Comput. Geosci. 50 (2013) 25–32.

[17] T. Zhai, Z. Zhao, X. Zhou, A review of 3D reconstruction methods for complex microstructures of porous media: Imaging devices, techniques and applications, Rock Mech. Rock Eng. (2025) 1–44.

[18] P.-E. Øren, S. Bakke, Process based reconstruction of sandstones and prediction of transport properties, Transp. Porous Media 46 (2) (2002) 311–343.

[19] N. Xiao, F. Berto, X. Zhou, Three-dimensional stochastic reconstruction of porous media: A systematic review, J. Build. Eng. 91 (2024) 109642.

[20] L. Mosser, O. Dubrule, M.J. Blunt, Reconstruction of three-dimensional porous media using generative adversarial neural networks, Phys. Rev. E 96 (4) (2017) 043309.

[21] J. Shi, W. Zhang, W. Wang, Y.-h. Sun, C. Xu, H. Zhu, Z. Sun, Randomly generating three-dimensional realistic schistous sand particles using deep learning: Variational autoencoder implementation, Eng. Geol. 291 (2021) 106235.

[22] T. Zhang, D. Li, F. Lu, A pore space reconstruction method of shale based on autoencoders and generative adversarial networks, Comput. Geosci. 25 (6) (2021) 2149–2165.

[23] T. Zhang, X. Ji, F. Lu, 3D reconstruction of porous media by combining scaling transformation and multi-scale discrimination using generative adversarial networks, J. Pet. Sci. Eng. 209 (2022) 109815.

[24] H. Mirzaee, S. Kamrava, P. Tahmasebi, Minireview on porous media and microstructure reconstruction using machine learning techniques: Recent advances and outlook, Energy Fuels 37 (20) (2023) 15348–15372.

[25] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, S. Ganguli, Deep unsupervised learning using nonequilibrium thermodynamics, in: International Conference on Machine Learning, pmlr, 2015, pp. 2256–2265.

[26] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, Adv. Neural Inf. Process. Syst. 33 (2020) 6840–6851.

[27] P. Dhariwal, A. Nichol, Diffusion models beat gans on image synthesis, Adv. Neural Inf. Process. Syst. 34 (2021) 8780–8794.

[28] Y. Song, J. Sohl-Dickstein, D.P. Kingma, A. Kumar, S. Ermon, B. Poole, Score-based generative modeling through stochastic differential equations, 2020, arXiv preprint arXiv:2011.13456.

[29] A.Q. Nichol, P. Dhariwal, Improved denoising diffusion probabilistic models, in: International Conference on Machine Learning, PMLR, 2021, pp. 8162–8171.

[30] Y. Zhang, T. Long, H. Zhang, Stable diffusion for the inverse design of microstructures, 2024, arXiv preprint arXiv:2409.19133.

[31] D. Naiff, B.P. Schaeffer, G. Pires, D. Stojkovic, T. Rapstine, F. Ramos, Controlled latent diffusion models for 3D porous media reconstruction, Comput. Geosci. 206 (2026) 106038.

[32] S. Huang, L. Huang, Z. Lai, J. Zhao, Morphology characterization and discrete element modeling of coral sand with intraparticle voids, Eng. Geol. 315 (2023) 107023.

[33] D. Chen, W. Zhang, C. Li, L. Ma, X. Shi, H. Li, H. Zhu, Randomly generating realistic calcareous sand for directional seepage simulation using deep convolutional generative adversarial networks, J. Rock Mech. Geotech. Eng. (2025).

[34] B. Zhou, J. Wang, Generation of a realistic 3D sand assembly using X-ray micro-computed tomography and spherical harmonic-based principal component analysis, Int. J. Numer. Anal. Methods Geomech. 41 (1) (2017) 93–109.

[35] X. Wang, Y. Wu, J. Cui, C.-Q. Zhu, X.-Z. Wang, Shape characteristics of coral sand from the south China sea, J. Mar. Sci. Eng. 8 (10) (2020) 803.

[36] B. Zhao, J. Wang, 3D quantitative shape analysis on form, roundness, and compactness with μCT, Powder Technol. 291 (2016) 262–275.

[37] D.V. Vranic, D. Saupe, J. Richter, Tools for 3D-object retrieval: Karhunen-loeve transform and spherical harmonics, in: 2001 IEEE Fourth Workshop on Multimedia Signal Processing (Cat. No. 01TH8564), IEEE, 2001, pp. 293–298.

[38] J. Ho, W. Chan, C. Saharia, J. Whang, R. Gao, A. Gritsenko, D.P. Kingma, B. Poole, M. Norouzi, D.J. Fleet, et al., Imagen video: High definition video generation with diffusion models, 2022, arXiv preprint arXiv:2210.02303.

[39] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[40] R. Llugsi, S. El Yacoubi, A. Fontaine, P. Lupera, Comparison between adam, AdaMax and adam w optimizers to implement a weather forecast based on neural networks for the andean city of quito, in: 2021 IEEE Fifth Ecuador Technical Chapters Meeting, ETCM, IEEE, 2021, pp. 1–6.

[41] L. Ding, W. Fei, Y. Huang, S. Ding, W. Dai, C. Li, J. Zou, H. Xiong, AMPA: Adaptive mixed precision allocation for low-bit integer training, in: Forty-First International Conference on Machine Learning, 2024.

[42] Z. Tang, J. Tang, H. Luo, F. Wang, T.-H. Chang, Accelerating parallel sampling of diffusion models, in: Forty-First International Conference on Machine Learning, 2024.
[43] H. Liang, M. Tsuei, N. Abbott, F. You, AI framework with computational box counting and integer programming removes quantization error in fractal dimension analysis of optical images, Chem. Eng. J. 446 (2022) 137058.
[44] C. Yang, B. Indurkhya, J. See, B. Gao, Y. Ke, Z. Boukhers, Z. Yang, M. Grzegorzek, Skeleton ground truth extraction: methodology, annotation tool and benchmarks, Int. J. Comput. Vis. 132 (4) (2024) 1219–1241.
[45] B. Yao, H. He, S. Kang, Y. Chao, L. He, A review for the Euler number computing problem, Electronics 12 (21) (2023) 4406.
[46] Y. Jiao, F. Stillinger, S. Torquato, Modeling heterogeneous materials via two-point correlation functions: Basic principles, Phys. Rev. E—Statistical, Nonlinear, Soft Matter Phys. 76 (3) (2007) 031110.
[47] Y.-C. Chen, A tutorial on kernel density estimation and recent advances, Biostat. Epidemiology 1 (1) (2017) 161–187.
[48] M.L. Menéndez, J.A. Pardo, L. Pardo, M.d.C. Pardo, The jensen-shannon divergence, J. Franklin Inst. 334 (2) (1997) 307–318.
[49] J.-F. Pambrun, R. Noumeir, Limitations of the SSIM quality metric in the context of diagnostic imaging, in: 2015 IEEE International Conference on Image Processing, ICIP, IEEE, 2015, pp. 2960–2963.
[50] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014).
[51] M. Arjovsky, L. Bottou, Towards principled methods for training generative adversarial networks, 2017, arXiv preprint arXiv:1701.04862.
[52] P. Dhariwal, A. Nichol, Diffusion models beat gans on image synthesis, Adv. Neural Inf. Process. Syst. 34 (2021) 8780–8794.
[53] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, 2015, arXiv preprint arXiv:1511.06434.
[54] F.-A. Croitoru, V. Hondru, R.T. Ionescu, M. Shah, Diffusion models in vision: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 45 (9) (2023) 10850–10869.