Intelligent Geoengineering 2 (2025) 96–108

Contents lists available at ScienceDirect

Intelligent Geoengineering

journal homepage: www.keaipublishing.com/en/journals/ige/

Full length article

Efficient generation of digital rock CT images using LoRA-enhanced stable diffusion models

Kunyao Li[a], Haijiang Li[a,*], Ali Khudhair[a], Jun Yan[b], Bin Wang[c]

[a] School of Engineering, Cardiff University, Queens Building, the Parade, Cardiff CF24 3AA, UK
[b] State Key Laboratory of Structural Analysis for Industrial Equpment, Department of Engineering Mechanic, Dalian University of Technology Dalian, Dalian 116024, China
[c] Smart Construction AI Studio, China Energy Engineering Group Co., Ltd., Beijing 100025, China

ARTICLE INFO

Keywords:
Digital Rock Analysis
Stable Diffusion
Low-Rank Adaptation (LoRA)
Computed Tomography (CT)
Generative AI
Computational Geoscience

ABSTRACT

Digital rock analysis (DRA) is fundamental for geo-energy research, enabling the characterisation of microstructures for applications like hydrocarbon recovery, carbon storage, and groundwater modelling. Although 2D CT images provide valuable pore-scale data, the scarcity of real-world datasets limits the effectiveness of advanced analysis. Generative AI presents a promising approach for synthesizing high-quality rock images but faces key challenges, including high computational demands, insufficient evaluation metrics, and the trade-off between image fidelity and diversity. To address these limitations, this study proposes the use of Low-Rank Adaptation (LoRA) for fine-tuning stable diffusion models, significantly reducing computational requirements while maintaining image quality. A systematic investigation was conducted to evaluate the influence of LoRA training parameters, including rank and learning rate, on the quality of generated images. Image outputs were assessed using both standard generative metrics, such as Kernel Inception Distance (KID), and domain-specific metrics, including porosity, pore count, and pore area distributions. The optimised LoRA-enhanced diffusion model achieved a 92.6 % reduction in KID relative to baseline models, while also improving inference speed. Building on these advancements, this study demonstrates that the LoRA-enhanced diffusion model significantly improves neural network extrapolation in incomplete data scenarios through statistically consistent synthetic generation. Despite control challenges, this approach reduces costs and enables diverse applications, bridging fundamental rock physics with practical energy research.

1. Introduction

Digital rock analysis (DRA) has emerged as an essential methodology in geo-energy research, facilitating comprehensive characterisation of rock microstructures for applications including hydrocarbon recovery, carbon storage, and groundwater flow modelling. Among the available imaging techniques, two-dimensional (2D) computed tomography (CT) offers a cost-effective means of acquiring essential pore-scale information, facilitating both analysis and modelling of subsurface structures. However, the limited availability of real 2D CT datasets poses a significant barrier to the development of intelligent, data-driven analysis methods.

Recent advancements in generative artificial intelligence (AI), particularly the development of generative adversarial networks (GANs) and diffusion models, present promising opportunities for the synthesis of high-quality two-dimensional rock images. Compared to traditional image generation methods, deep learning approaches are more flexible and capable of capturing the complex granular details found in CT imagery of various rock types. Despite their potential, several critical challenges remain unresolved: (1) State-of-the-art generative models require substantial computational resources for full model training, limiting their practical application in specialised scientific domains like digital rock analysis; and (2) Current assessments of generated images rely on generic metrics, which fail to fully capture the geological and physical properties relevant to domain experts. Moreover, systematic studies of both fidelity and diversity in generated rock CT images are lacking.

To address these challenges, this study introduces a novel approach that leverages Low-Rank Adaptation (LoRA) to fine-tune stable diffusion models for efficient, high-quality generation of 2D CT rock images.

* Corresponding author.
E-mail address: lih@cardiff.ac.uk (H. Li).

https://doi.org/10.1016/j.ige.2025.06.002
Received 13 May 2025; Accepted 9 June 2025
Available online 16 June 2025

The proposed method significantly reduces computational overhead while maintaining or improving generative performance in geoscientific applications. The core contributions of this work are as follows:

- LoRA is employed as a lightweight adaptation technique, injecting low-rank updates into selected layers of the stable diffusion model. This enables rapid fine-tuning without requiring full retraining, thereby reducing computational costs while preserving output quality.
- Critical parameters involved in both training and generation, including LoRA rank and learning rate, are systematically evaluated to determine their influence on image fidelity and diversity. The study provides practical recommendations for parameter selection in real-world applications.
- A novel evaluation framework is implemented, combining standard generative metrics (e.g., Kernel Inception Distance (KID)) with domain-specific properties such as porosity, pore count, and pore area distribution. This ensures that generated images are not only visually plausible but also geologically meaningful.

The primary objective of this research is to demonstrate the effectiveness of LoRA-adapted stable diffusion in producing realistic 2D CT rock images, optimise its configuration for computational efficiency and image quality, and validate a domain-aware evaluation methodology. By bridging the gap between generative AI and domain-specific geological imaging, this study provides a foundation for broader applications in digital rock physics and data-driven geo-energy modelling. The paper is structured as follows: Section 2 reviews related work on digital rock image generation, including traditional statistical methods and modern deep learning approaches, with a focus on diffusion models and LoRA fine-tuning. Section 3 details the research methodology, including model architecture, dataset preparation, and evaluation metrics. Section 4 presents experimental results, analysing the impact of different LoRA configurations on image quality and comparing performance with baseline models. Section 5 discusses the implications of the findings, potential applications, limitations, and directions for future research. Section 6 concludes the study by summarising the key contributions and research significance.

2. Literature review

DRA, as a frontier field in modern geoscience and materials research, leverages advanced imaging technologies to capture the complex microstructures of porous materials. Since its inception in the 1990s, DRA has evolved through the integration of numerical simulation and advanced imaging technologies, becoming widely applied in petroleum engineering, hydrogeology, and carbon storage studies (Yang et al., 2024; Chi et al., 2024; Esmaeili, 2024). Micro-computed tomography (micro-CT), in particular, offers high-resolution imaging capabilities that reveal intricate pore geometries at the microscale. However, despite its utility, micro-CT imaging remains constrained by long acquisition times, limited sample availability, and its inability to fully capture the heterogeneity of rock microstructures (Cnudde and Boone, 2013). These limitations have driven an urgent need for alternative approaches that can generate representative and diverse porous media images, while preserving key geological properties (Okabe and Blunt, 2004; Tahmasebi and Sahimi, 2013; Mosser et al., 2017).

Digital rock image generation methods can be broadly classified into two categories: traditional statistical methods and deep learning methods. Traditional statistical methods include object-centred approaches, process-based methods, and pixel-level information techniques, which primarily rely on statistical and geological theories. In contrast, deep learning methods achieve automatic image generation and reconstruction through artificial neural networks and machine learning algorithms. Additionally, deep learning has been widely applied in digital rock analysis, involving multiple tasks such as image segmentation, super-resolution, and pore-scale modelling. However, obtaining sufficient training data remains a key bottleneck constraining technological development (Li et al., 2023; Karimpouli et al., 2024).

2.1. Traditional methods

Traditional digital rock image generation techniques can be broadly categorised into three classes: object-centred approaches, process-based models, and pixel-level statistical methods. These methods are predominantly grounded in geological theory and statistical modelling.

Object-centred methods treat rock structures as assemblies of discrete entities, such as grains and pores, with spatial configurations determined through heuristics or optimisation algorithms like simulated annealing (Diogenes et al., 2008). While conceptually intuitive and computationally stable, these techniques struggle to capture large-scale pore connectivity and detailed morphological variations due to their reliance on low-order statistical information. Process-based methods aim to simulate geological formation processes, such as sedimentation or diagenesis, resulting in more realistic structures (Biswal et al., 2007; ØREN and Bakke, 2002). However, these simulations often require extensive parameter tuning, are computationally intensive, and are limited in generalisability. Translating complex geological phenomena into efficient and accurate computational models remains a core challenge in this category.

Pixel-level geostatistical approaches model the rock structure on a voxel grid using spatial statistics. Two-point statistics (TPS), such as the Joshi-Quiblier-Adler method, provide porosity-consistent image generation based on correlation functions (Yeong and Torquato, 1998; Torquato, 2002), but fail to capture higher-order spatial patterns. In contrast, multipoint statistics (MPS) techniques, such as Single Normal Equation Simulation (SNESIM), direct sampling, and cross-correlation-based simulation, extract complex spatial features from training images (Tahmasebi and Sahimi, 2013; Okabe and Blunt, 2005; Strebelle, 2002; Mariethoz et al., 2010). While MPS improves global realism, it still struggles to preserve fine structural details and often requires high-quality training datasets.

2.2. Deep learning methods

Deep learning has achieved remarkable success in image generation, particularly in facial recognition, scene generation, and medical imaging. GANs demonstrate unique advantages in digital rock reconstruction. As a data-driven approach, GANs require no prior information, avoid the complexity of manual feature design, and can rapidly generate new structures after training (Goodfellow et al., 2014). Early milestone works successfully generated microstructures of various rock types, including berea packs, Berea sandstone, and Ketton limestone (Mosser et al., 2017, 2018). Subsequently, researchers expanded GANs' applications to innovative areas such as shale digital core reconstruction, image resolution enhancement, and three-dimensional structure reconstruction from two-dimensional fragments (Zha et al., 2020; Zhao et al., 2023; Feng et al., 2020; Kench and Cooper, 2021).

Despite their success, GANs face persistent training challenges, notably mode collapse and instability (Liu et al., 2020). In response, diffusion models have emerged as a promising alternative. These models simulate a two-stage process—first adding noise to training data, then learning to reverse this diffusion to synthesise new images (Ho et al., 2020). These models simulate a diffusion process from order to disorder and back to order, gradually adding noise and then systematically removing it to generate data. Their core advantages lie in training stability and high-quality sample generation, effectively avoiding the mode collapse problems common in GANs. Across a diverse range of applications, from synthesising realistic images to performing precise image segmentation and enhancing resolution through super-resolution techniques, diffusion models have demonstrated remarkable success and versatility (Dhariwal and Nichol, 2021; Saharia et al., 2022; Choi et al., 2021; Rombach et al., 2022).

These advancements in diffusion modelling have opened new pathways for addressing longstanding challenges in the field of digital rock analysis, particularly the persistent issue of limited training data. In digital rock research, diffusion models have achieved micro-CT image resolution enhancement and high-quality paired image data augmentation (Ma et al., 2024). Researchers have begun leveraging these powerful generative capabilities to create synthetic datasets that complement real measurements. For example, Esmaeili et al (Esmaeili, 2024). proposed a diffusion-based framework that fuses synthetic and real digital rock images to address data scarcity, improving both image realism and physical property estimation. These developments indicate that diffusion models are well-suited to address the limitations of traditional techniques, particularly the need for flexible, data-driven approaches in geoscientific imaging.

While both traditional and deep learning-based methods have advanced the field of digital rock image generation, several critical limitations remain. First, high computational costs associated with training full-scale diffusion models pose significant barriers to widespread adoption. Second, existing evaluation metrics, such as Fréchet Inception Distance (FID) and Kernel Inception Distance (KID), primarily assess visual similarity and fail to capture domain-specific properties like porosity, connectivity, or pore size distribution. Third, there is a lack of systematic analysis addressing the trade-off between fidelity (preservation of geological realism) and diversity (variation in structural features) in generated outputs.

To address these challenges, this study introduces a novel approach that leverages LoRA to fine-tune stable diffusion models for digital rock CT image generation. The proposed method reduces computational demands, improves physical realism, and enables fine control over output diversity through parameter tuning. Additionally, it incorporates domain-specific evaluation metrics, such as porosity, pore count, and size distribution, to assess the geological plausibility of the generated images. By integrating lightweight fine-tuning, systematic parameter analysis, and application-aware evaluation, this work aims to bridge the gap between state-of-the-art generative methods and practical needs in geo-energy and porous media research.

3. Methodology

This section presents the proposed framework for generating high-fidelity and diverse 2D rock CT images using LoRA-enhanced stable diffusion. The methodology encompasses dataset preparation, model training, image generation, and evaluation. Fig. 1 illustrates the overall workflow, including data annotation, LoRA fine-tuning, and parameter optimisation for guided image synthesis.

3.1. Dataset and preprocessing

This study utilises the Digital Rocks Super Resolution Dataset 1 (DRSRD1) (Wang et al., 2019), specifically the 2D Bentheimer sandstone subset. This dataset comprises high-resolution micro-CT images with a spatial resolution of 3.8 μm. A total of 1000 2D slices were selected, each with a resolution of 800 × 800 pixels. To prepare the data for generative modelling, each image was binarised to separate pore spaces from the rock matrix. Semantic annotations were added to describe properties such as porosity, enabling conditional image generation. These preprocessed binary images served as the foundation for both training and evaluation tasks.

3.2. Model architecture and fine-tuning

3.2.1. Stable diffusion model

Stable diffusion is an advanced diffusion model that generates high-quality images through a series of denoising steps, demonstrating exceptional performance in image generation across numerous domains. Rock CT images typically originate from high-resolution microfocus X-ray tomography scans, featuring complex pore networks, mineral grain boundaries, and micrometre-scale structural details. These characteristics impose high demands on image generation models, requiring them to not only achieve high resolution but also accurately reproduce the physical reality of geological structures.

The core principle of diffusion models is to add Gaussian noise to the original image x_0(i.e., real rock CT slices) through a forward diffusion process. After multiple time steps t, the image is transformed into a pure noise distribution. Then, the reverse diffusion process gradually reconstructs the image by denoising. Compared to traditional GANs, the probabilistic denoising mechanism of stable diffusion is better suited for modelling the complex textures and randomness in rock CT images (Ma et al., 2024). The architecture of stable diffusion is based on U-Net. This is a symmetric convolutional neural network that maintains spatial consistency between the input (e.g., 512 × 512 pixel rock CT slices) and the output. U-Net includes both downsampling and upsampling paths. These paths are composed of Wide ResNet blocks, group normalisation, and self-attention mechanisms, which help capture long-range dependencies between rock pores and particles. The diffusion time step t is embedded into each residual block using sinusoidal positional embedding. This helps simulate how noise changes over time. The forward diffusion process can be mathematically described as:

$$dx = f(x_t)dt + g(t)dw \tag{1}$$

where dx denotes the infinitesimal change in the image state at time t, $f(x_t)$ is the drift term controlling the deterministic evolution, dt is a small time increment, $g(t)$ is the diffusion coefficient determining the noise intensity over time, and dw represents the increment of a Wiener process, introducing Gaussian noise via standard Brownian motion. This equation describes how the structure of the rock CT image is gradually disrupted by adding random noise, leading to a pure noise distribution. The training objective is to infer noise from noisy samples through conditional denoising, optimising the loss function :

$$\min \|\epsilon - f(x + \sqrt{t}\epsilon, t)\|^2 \tag{2}$$

where ϵ represents Gaussian noise, t is the time step, $x + \sqrt{t}\epsilon$ is the image with corresponding noise added at the time step t, and $f(x + \sqrt{t}\epsilon, t)$ is the noise predicted by the model. By minimising this loss, the model learns to recover fine-grained geological textures from noise. However, training the full diffusion model from scratch requires high-end GPUs and substantial resources. Hence, low-rank adaptation can be used to fine-tune it efficiently.

3.2.2. Low-rank adaptation (LoRA)

LoRA is a lightweight fine-tuning technique that adapts the stable diffusion model to the generation of rock CT images by adding a small number of low-rank parameter matrices. This generation process requires not only high fidelity but also accurate reflection of geological properties such as porosity, grain size distribution, and rock type characteristics. Traditional fine-tuning methods adapt to new tasks by adjusting the entire weight matrix (W), resulting in updated weights $W + \Delta W$. LoRA's innovation lies in decomposing the weight update into a product of two low-rank matrices: $\Delta W = A \cdot B$. This decomposition significantly reduces memory and computational requirements while maintaining the ability to adapt to the features of rock CT images.

In this study, LoRA is applied to the U-Net backbone of the stable diffusion model, including convolutional, group normalisation, and self-attention layers. These layers are frozen, and only the LoRA parameters are trained. This approach preserves core denoising capabilities while enabling the model to learn domain-specific features, such as pore geometry and grain boundaries in sandstone CT images. LoRA parameters are applied in parallel to the frozen layers, eliminating inference latency and facilitating rapid adaptation to diverse rock types. To enhance the quality of rock CT image generation, LoRA hyperparameters are fine-tuned, including the rank of low-rank matrices and the scaling factor that controls the magnitude of LoRA updates.

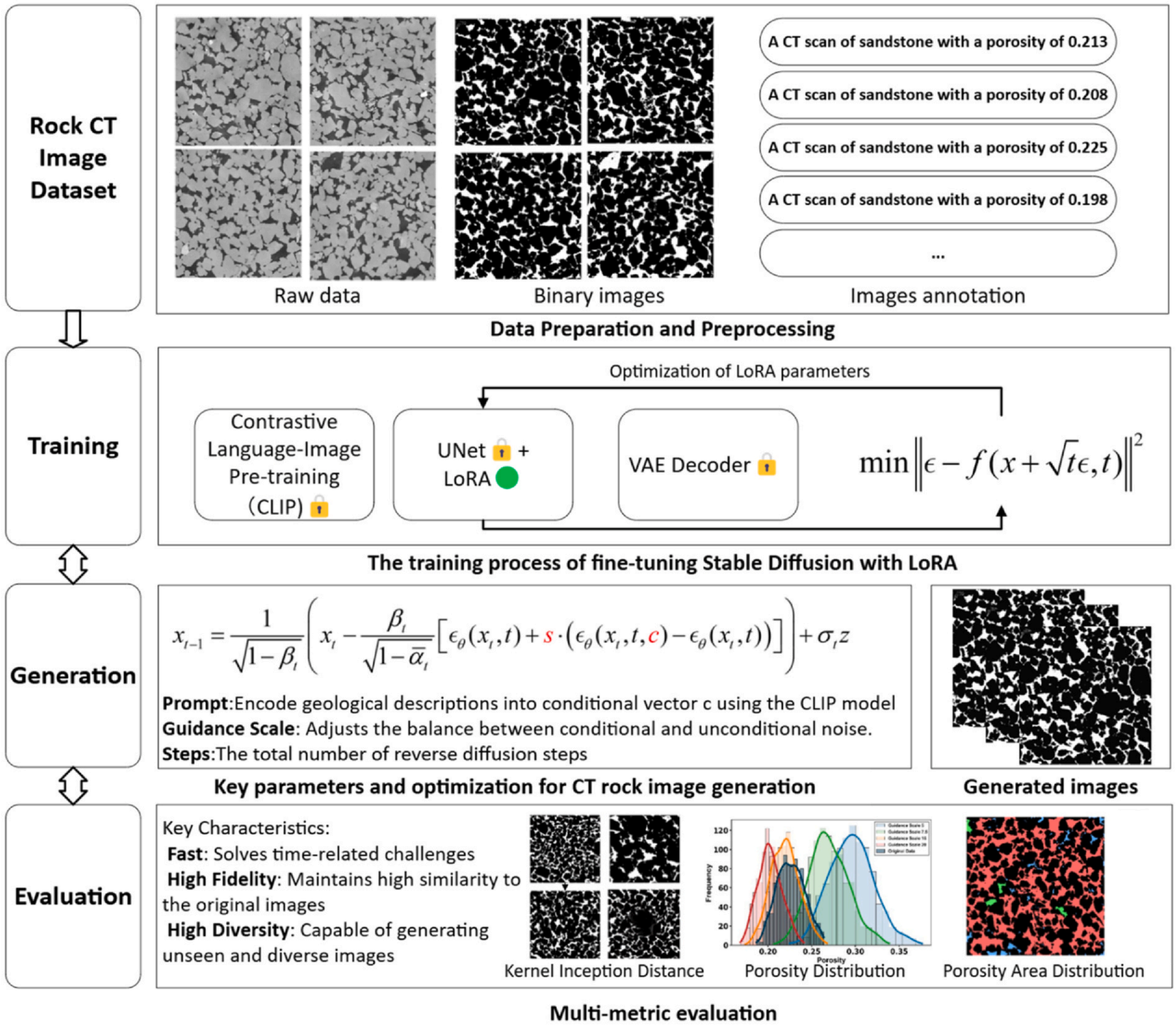


Fig. 1. The overall framework of the proposed method.

3.3. Generation process

After completing LoRA fine-tuning, the generation process of rock CT images is implemented through reverse diffusion sampling, aiming to create images with geological characteristics from random noise. The generation process starts from pure noise and reconstructs the rock CT images through iterative denoising. The update formula for single-step reverse diffusion, combined with text prompts (through condition c) and guidance scale (s), can be expressed as :

$$x_{t-1} = \frac{1}{\sqrt{1-\beta_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}[\epsilon_\theta(x_t, t) + s\cdot(\epsilon_\theta(x_t, t, c) - \epsilon_\theta(x_t, t))]\right) + \sigma_t z \tag{3}$$

where $x_{t-1}$is the image state from the time step t to $t-1$, the noise scheduling parameter at time step t, controlling the amount of noise added or removed in each step, β_t is the cumulative denoising factor, representing the degree of denoising from the initial time to t, and $\bar{\alpha}_t$ is the noise predicted by the U-Net controlled by condition c (text prompt).

This formula combines unconditional and conditional noise predictions through the Classifier-Free Guidance mechanism. $\epsilon_\theta(x_t, t)$ provides a denoising baseline for random rock structures, while $\epsilon_\theta(x_t, t, c)$ adjusts the denoising direction according to text prompts (such as porosity descriptions), and s controls the balance between the two. This mechanism balances geological control and structural randomness. Based on this, three key parameters were studied:

- **Text prompt (c):** Text prompts describing geological properties (e.g., "A CT scan of sandstone with a porosity of 0.213") are embedded via the CLIP model into a condition vector. This vector directly modulates the denoising trajectory, guiding the model to produce images that exhibit the specified porosity and microstructure.
- **Guidance scale (s):** This scalar controls the emphasis on the text condition. Higher values (e.g., 15–20) increase adherence to the prompt but may introduce artefacts such as over-sharpened grain boundaries. Lower values (e.g., 7.5) allow more structural randomness but may reduce prompt fidelity.
- **Iteration steps (T):** The number of reverse diffusion steps directly influences image clarity. Higher values ($100-200$) yield better pore edge definition and mineral detail but increase computational load. Lower values may speed up inference at the cost of structural fidelity.

These parameters were systematically varied and analysed to understand their impact on the geological validity and visual realism of the generated images.

3.4. Evaluation matrics

To assess the quality and utility of generated CT images, both fidelity and diversity are evaluated using domain-appropriate metrics. Regarding fidelity, the focus is on visual authenticity, accuracy of geological features, and consistency of statistical distributions. The core value of rock CT images lies in their microstructural features, such as porosity. Therefore, the evaluation metrics should reflect the similarity between generated images and real CT data in these aspects. As for diversity, the examination focuses on whether the generated images can not only follow the patterns of the original data but also extend beyond, producing porous images with porosity values not present in the original dataset. Consequently, this study employs two main categories of metrics: one category is Kernel Inception Distance (KID), which uses deep learning algorithms to analyse the fidelity between generated images and original images. The other category is pore-related features associated with the physical properties of rock CT images, including porosity distribution, quantified using KL divergence, and statistical analysis of pore count and pore area distribution.

3.4.1. Kernel inception distance (KID)

KID is a metric that measures the similarity between the distributions of generated images and real images, particularly suitable for evaluating the visual authenticity of rock CT images (Bińkowski et al., 2021). Unlike the Inception Score (IS), KID uses feature embeddings from Inception V3 and kernel methods to measure maximum mean discrepancy (MMD), offering greater robustness and reduced bias. For rock CT images, KID assesses whether generated images are close to real data in aspects such as pore networks and grain boundaries by extracting high-level features from images. The calculation formula for KID is :

$$\mathrm{KID}=\| E_{x\sim P_r}[k(x,\cdot)] - E_{x\sim P_g}[k(x,\cdot)] \|^2_{MMD} \tag{4}$$

where P_r is Distribution of real rock CT images, P_g is the Distribution of generated rock CT images, $k(x,\cdot)$ is Kernel function, used to map images to feature space, E is expectation operation, and $\|\cdot\|^2_{\mathrm{MMD}}$ is square of maximum mean discrepancy, measuring the distance between two distributions.

In this study, 512 × 512 pixel images were input into a pretrained Inception V3 model, and KID scores were computed from the extracted features. Lower KID values indicate that the generated images are visually closer to real CT data, suggesting successful replication of geological features such as pore connectivity and mineral grain morphology.

3.4.2. Porosity and kullback–leibler (KL) divergence

Porosity is a fundamental physical property of porous media, directly influencing reservoir permeability and subsurface flow behaviour. To evaluate the geological fidelity of the generated images, porosity and pore area were calculated using a binary segmentation method. For various generation conditions (such as different prompts or Guidance Scale values), 1000 images were generated under each condition to calculate their porosity distribution match with the real rock CT dataset. The real dataset (denoted as $\boldsymbol{P_r}$) and each set of generated images (denoted as $\boldsymbol{P_g}$, 1000 images per set) underwent binarisation to calculate the porosity $\boldsymbol{\phi}$ of each image. For the construction of porosity distribution, porosity values were divided into bins, and probability distributions of the real dataset and generated image sets in each bin were calculated, denoted as $P_r(\phi)$ and $P_g(\phi)$ respectively. The KL Divergence between these two distributions is defined as:

$$D_{\mathrm{KL}}(P_r \| P_g) = \sum_{\phi} P_r(\phi) \log\left(\frac{P_r(\phi)}{P_g(\phi)}\right) \tag{5}$$

Where $P_r(\phi)$ is probability of porosity ϕ in the real rock CT dataset. $P_g(\phi)$ is probability of porosity ϕ in the generated image set. $\sum_{\phi}$ is the summation over all porosity bins.

A lower KL divergence value indicates a stronger alignment between the porosity distributions of the generated and real datasets, validating the model's ability to generate physically realistic and geologically meaningful structures.

4. Experimental results

Diffusion models, particularly stable diffusion, have achieved remarkable success in image generation due to their denoising-based generative capabilities and architectural robustness. However, training a complete diffusion model from scratch typically requires substantial computational resources and time constraints that are often present in practical applications. LoRA offers an efficient fine-tuning solution by introducing low-rank decomposition matrices into pre-trained models. This significantly reduces the number of trainable parameters, thereby decreasing both computational and memory costs while maintaining model performance. LoRA thus provides a promising path towards efficient optimisation in resource-constrained environments.

To comprehensively evaluate the effectiveness of LoRA-enhanced stable diffusion in generating rock CT images, the experimental procedure is divided into two main phases: training and generation. During the training phase, the focus is on assessing the impact of LoRA-specific parameters such as Network Dim and Network Alpha, with the Kernel Inception Distance (KID) used as the primary evaluation metric. Meanwhile, the generation phase investigates how generation parameters, such as text prompt, guidance scale, and sampling steps, affect image quality, diversity, and geological realism, particularly in relation to pore-scale features.

4.1. Training process parameters

This experiment evaluates the impact of different LoRA configuration parameters, Network Dim and Network Alpha, on generation quality, and to conduct a comprehensive comparison with diffusion model. Network Dim determines the rank of the LoRA network, which reflects the complexity of low-rank features that the LoRA module can learn. A higher value allows the model to capture more complex features, but it also increases computational load and memory usage. Network Alpha is a scaling factor for LoRA weights, influencing both the learning rate and training stability. If the Alpha value is too high, it may lead to instability or overfitting; if too low, it can degrade generation quality. Therefore, selecting appropriate values for these two parameters is critical for effective LoRA fine-tuning.

Fig. 2 shows the generated data from the trained models, including original data, outputs from a diffusion model trained from scratch (Esmaeili, 2024) and images generated under different parameter settings after 2000 training steps. The goal is to analyse model performance under limited training iterations. The tested parameter configurations are summarised in Table 1. For all four LoRA settings, Network Dim is set to twice the value of Network Alpha, except for the baseline model.

The results show that the LoRA-enhanced stable diffusion model consistently outperforms the baseline model across all tested configurations. As shown in Fig. 2, the baseline model produces porous structures with blurred boundaries, missing small-scale pore details, uneven spatial distribution, and excessive local aggregation. The optimal configuration, Parameter 3 (Network Dim = 32, Alpha = 16), achieves a KID value of 0.0427, representing a 92.6 % reduction compared to the baseline (0.5795). This highlights the model's capacity to accurately replicate geological textures under limited training.

The optimal configuration (r = 32) achieves a KID value of 0.0427, which is a 92.6 % reduction compared to the diffusion model (0.5795), indicating a substantial improvement in generation quality. As the Network Dim value increases from 8 to 32, the KID value consistently decreases (0.2236 → 0.1949 → 0.0427), demonstrating that higher rank parameters allow the model to better capture data features. In Fig. 2, it can be observed that when Network Dim is relatively small, the

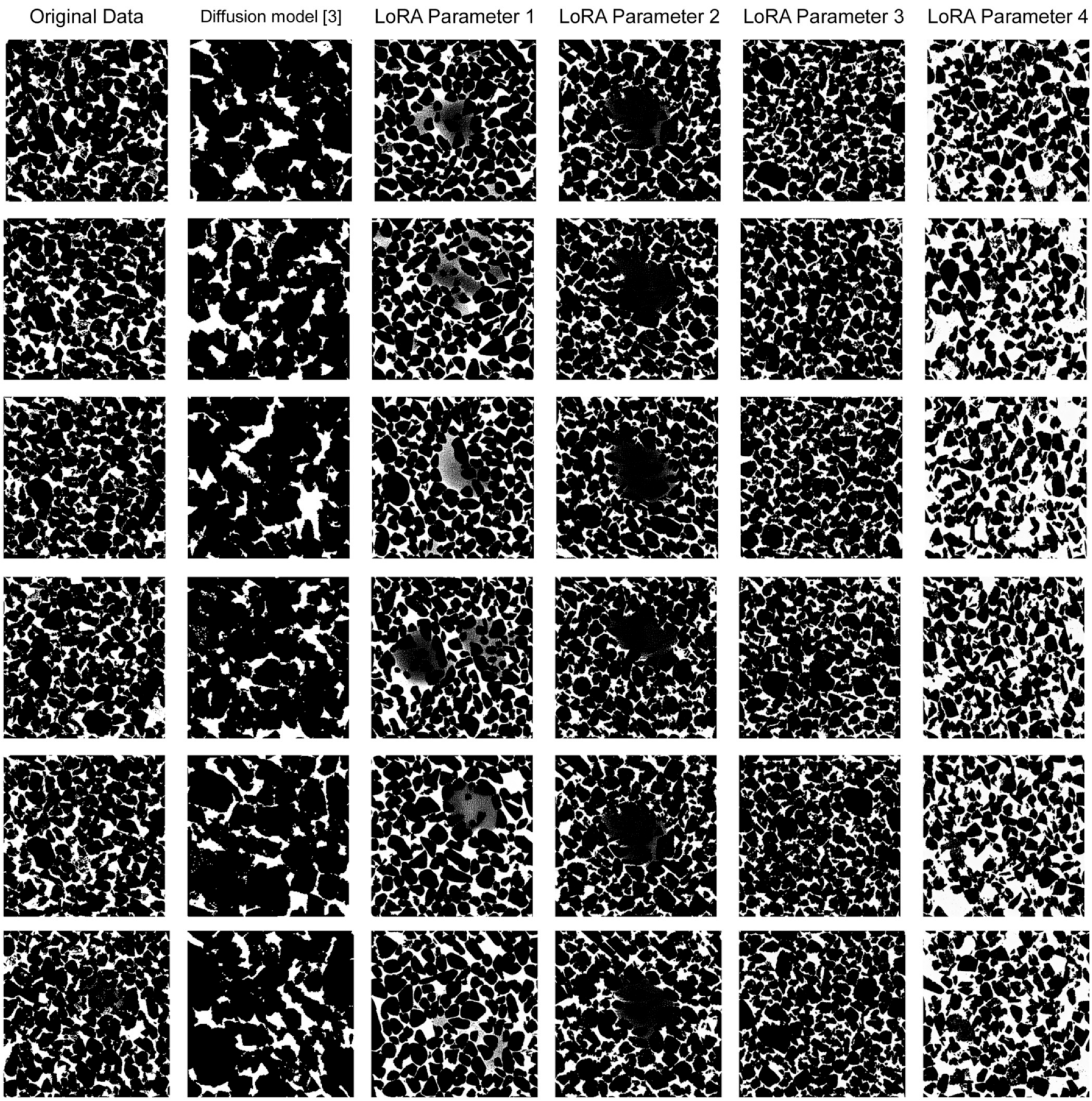


Fig. 2. Original and generated data under different model trainings.

model learns the noise areas of the original image, leading to underfitting. However, when Network Dim increases further to 64, the KID value drops to 0.0842, and the image shows a typical "U-shaped curve" feature. This indicates that a higher r is not always better in the generation of rock CT images. In Parameter 4 of Fig. 2, significant loss of particle information occurs. A very high Network Dim value may cause the model to overfit, limiting its generalisation ability and reducing its ability to express fine details in the image. Furthermore, the KID variance for the $r = 32$ configuration (1.5816e-07) is the lowest, further confirming the stability of its generation quality.

Table 1
Model training parameters and results.

Model Parameters	Network Dim	Network Alpha	KID Value	Variance
Diffusion Model (Esmaeili, 2024)	-	-	0.5795	1.5994e − 07
Parameter 1	8	4	0.2236	2.3724e − 07
Parameter 2	16	8	0.1949	1.7257e − 07
Parameter 3	32	16	0.0427	1.5816e − 07
Parameter 4	64	32	0.0842	2.1204e − 07

Table 2
Porosity statistics under different generation conditions.

Label	Prompt	Guidance Scale	Avg. Porosity	Std. Dev.	KL Divergence
Original	-	-	0.2260	0.0159	0.0000
P0.20-GS5	0.20	5.0	0.2959	0.0256	16.2073
P0.20-GS7.5	0.20	7.5	0.2700	0.0225	6.6871
P0.20-GS15	0.20	15.0	0.2194	0.0177	0.2390
P0.20-GS20	0.20	20.0	0.2023	0.0138	4.4085
P0.25-GS5	0.25	5.0	0.2958	0.0266	13.3591
P0.25-GS7.5	0.25	7.5	0.2701	0.0230	5.7343
P0.25-GS15	0.25	15.0	0.2194	0.0179	0.2749
P0.25-GS20	0.25	20.0	0.2009	0.0149	4.9716
P0.30-GS5	0.30	5.0	0.2958	0.0274	14.4281
P0.30-GS7.5	0.30	7.5	0.2695	0.0228	5.3943
P0.30-GS15	0.30	15.0	0.2201	0.0163	0.2499
P0.30-GS20	0.30	20.0	0.2003	0.0158	5.4240

4.2. Porosity distribution

Stable diffusion involves not only the impact of training parameters on generation quality but also the effect of different parameter settings during the generation process, particularly in the case of rock CT images. As a type of porous material, rock CT images have porosity as one of their key structural parameters, directly affecting critical properties such as material density, strength, permeability, and adsorption.

In this experiment, the impact of stable diffusion parameters on the porosity of generated images is analysed, focusing on the relationship between the porosity values in the prompts and the guidance scale. A structured prompt template is used: "A CT scan of sandstone with a porosity of {porosity}", testing three porosity values (0.20, 0.25, 0.30). For each value, four guidance scale parameters (5, 7.5, 15, 20) are assessed to examine their effects on the results. The goal is twofold: (1) to assess whether stable diffusion could generate images with porosity values matching the prompt; (2) to evaluate how the guidance scale influences the distribution and realism of the generated porosity.

The results show that the generated images' porosity values are mainly influenced by the guidance scale, rather than the prompt. This occurs because porosity is not part of the loss function during training but rather serves as a guiding prompt. As such, the model learns only from the original image data, making the guidance scale's impact more prominent. Table 2 shows the porosity statistics of images generated under various experimental conditions. "P" represents different porosity values in the prompts, while "GS" indicates the guidance scale. The average porosity of the generated images remain similar across different prompt settings when the guidance scale is the same, indicating that the generation process relies more on the guidance scale than on the prompt itself. Fig. 3(e) further illustrates this, where the box maps demonstrate consistent patterns across different prompt settings at the same guidance scale.

Fig. 3(a)–(c) illustrate how the porosity distributions shift under different prompt values. At low guidance scales (GS = 5), the model produces higher porosity values ($\sim$0.295), regardless of the prompt, as seen in P0.20-GS5, P0.25-GS5, and P0.30-GS5 samples. These values exceed the real sample value of 0.2260 and do not follow prompt values. The underlying mechanism here involves the model's tendency to favor creativity over constraint when guidance is minimal - essentially giving the generative process more room to explore diverse possibilities rather than strictly adhering to prompts. This is further confirmed by KL Divergence measurements, which reached their highest values ($13-16$) at GS= 5, indicating significant differences from the original distribution. In contrast, at high guidance scales (GS = 15 or 20), the model produces lower porosity values ($\sim$0.20) that converge toward the mean of the original data, even when prompted for higher porosities. For instance, P0.20-GS20, P0.25-GS20, and P0.30-GS20 yielded values of 0.2023, 0.2009, and 0.2003 respectively. Fig. 3(d) shows that at GS= 15, KL divergence reached its lowest point ($\sim$0.25), suggesting optimal similarity to the original distribution, while increasing to GS= 20 caused KL divergence to rise again ($4-5$), following a "U-shaped" trend. Fig. 3(e) further confirms that at the same GS value, output porosity remains nearly constant across different prompts, indicating alignment with dominant training data patterns rather than prompts. This analysis demonstrates that GS $\approx$ 15 represents the optimal balance, producing porosity distributions that most closely resemble real rock CT data.

Overall, the results of the analysis demonstrate that the generation of rock CT images with specific porosity using stable diffusion is predominantly controlled by the guidance scale rather than the porosity values specified in prompts. The P0.20-GS15 combination produces images with porosity distributions closest to the original data, exhibiting minimal KL divergence and thus representing the most realistic rock CT images. For greater diversity in generated images, lower guidance scale values are preferable, as they offer more varied results while maintaining reasonable distribution characteristics. These findings highlight that when using stable diffusion to generate porous material images, adjusting the guidance scale should be prioritised over modifying porosity values in prompts. The optimal parameter selection depends on whether the priority is authenticity (higher GS) or diversity (lower GS).

4.3. Porous area distribution

This section investigates the model's capability to simulate not just porosity values but also the distribution and size of pores, which directly affect the microstructural realism of generated porous media. The study evaluates how well the model reproduces pore area distributions within the training data range and whether it can extrapolate to unseen porosity levels. To assess this, the pore quantity, spatial distribution, and microstructural characteristics were compared between original and generated CT images. Fig. 4 shows the porosity area distribution for original, in-range, and out-of-range generated data, with pore sizes categorised into 0–500 pixels (blue), 500–1000 pixels (green), and 1000 + pixels (red), highlighting the diversity of CT rock images. This colour-coded segmentation highlights how pore size composition varies with porosity.

For data with the same porosity, the pore count and distribution between original and generated images were compared. While overall distribution patterns remained similar, notable diversity in local features was observed. In Fig. 4, comparing the Original Image and Generated within Data Range, although images with the same porosity were generated, there were differences in the area distribution of pores of varying sizes. Specifically, as shown in Fig. 5(a) and (b), for a porosity of 0.1884, small pores (0–500 pixels) decreased from 93.46 % in the original image to 87.11 % in the generated image, while medium-sized

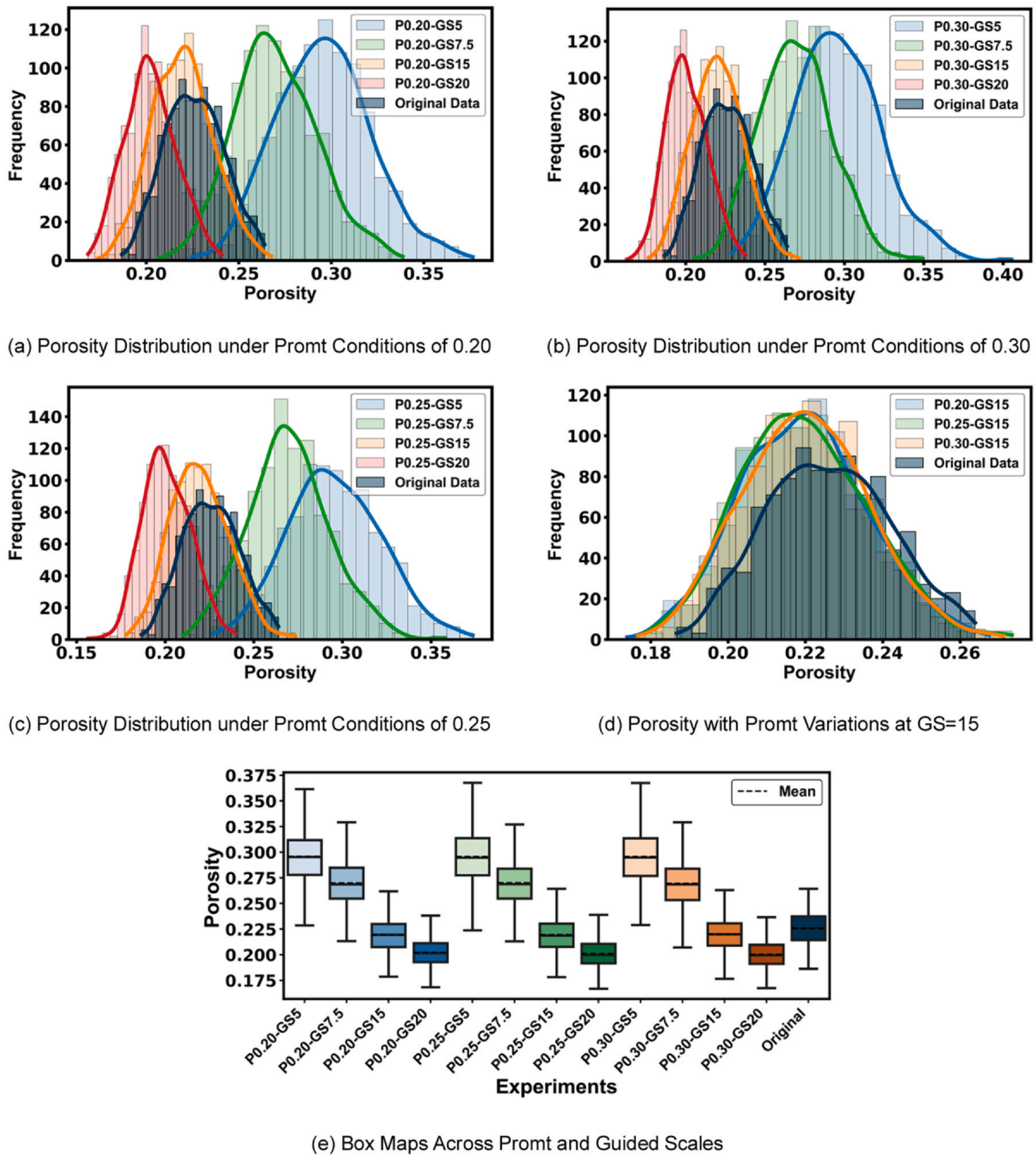


Fig. 3. Comparative study of porosity distribution with varying promt and guided scale.

pores (500–1000 pixels) increased from 3.39 % to 8.36 %. Similarly, for a porosity of 0.2312, large pores (1000 + pixels) significantly increased from 50.65 % to 70.33 % in the generated images. These differences suggest that while the global porosity metric remains stable, the local pore morphology becomes more varied in generated images. This diversity likely results from the stochastic design of the model and the influence of data augmentation during training, allowing the generation of structurally diverse yet physically plausible samples.

To evaluate the model's extrapolation capabilities, images with porosity values outside the training data range—such as 0.1636, 0.2972, and 0.3584—are selected from the existing generated dataset. Fig. 4 shows that Lower porosity (0.1636) generated samples with fewer small pores and relatively increased medium/large pores compared to in-range samples like 0.1884. Higher porosities (0.2972, 0.3584) lead to a significant increase in large pore areas, especially at 0.3584, where the red regions dominate. As shown in Fig. 5(c) and (d), the extrapolated porosity data (marked with red asterisks) has a similar pore count distribution to the original data. In terms of area, the 0.1636 sample has fewer small pore regions compared to 0.1884, while the large pore area is relatively higher. For porosities 0.2972 and 0.3584, the large pore areas increase significantly, especially at 0.3584, where the proportion is notably higher than in the 0.2639 sample. The pore feature variations in the extrapolated images closely align with the original image data. This consistency confirms the physical validity and practical usability of the extrapolated results. These results strongly indicate that the generation model can not only accurately replicate the pore distribution characteristics within the original data range but also reasonably predict pore distribution beyond this range. This confirms that the model has learned the underlying relationship between porosity and pore geometry, and is capable of generalising beyond the training set while maintaining geological plausibility.

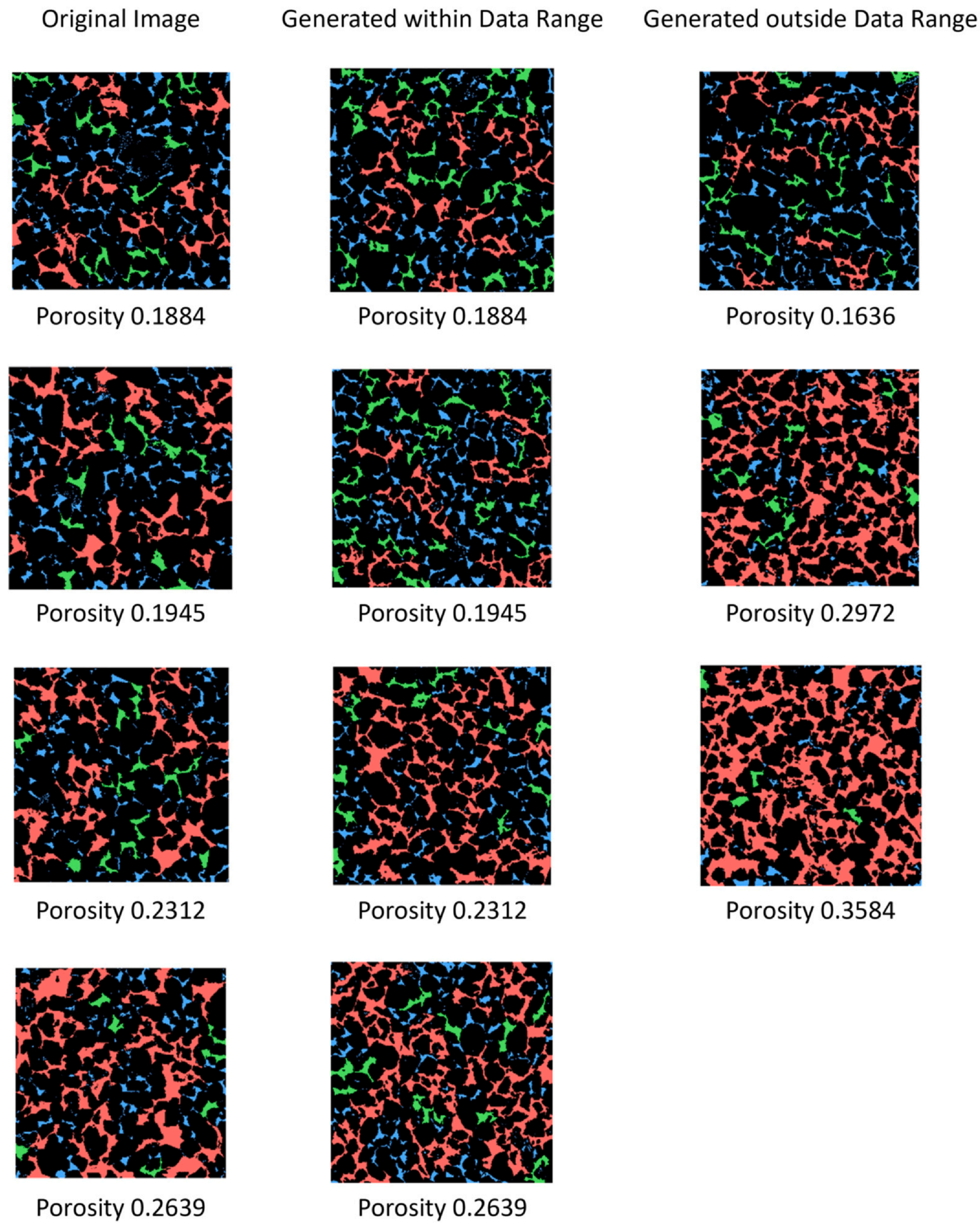


Fig. 4. Comparative porosity area distribution: original, in-range generated, and out-of-range generated data.

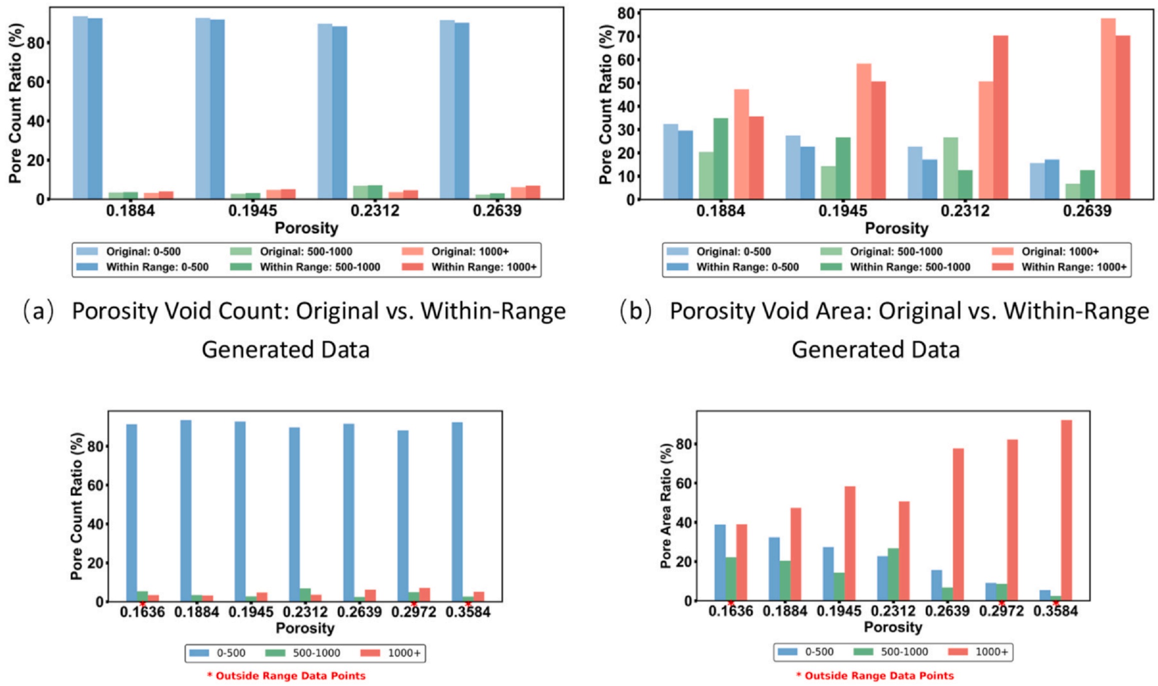
4.4. Generation iterations

The number of sampling steps in diffusion models is a critical hyperparameter that directly affects both the quality of generated images and the computational cost associated with the generation process. This section investigates the trade-off between image fidelity and inference time by analysing how different step counts influence the KID and the time required to generate a batch of 50 images. Fig. 6 presents the results of this analysis, highlighting the relationship between sampling steps, image quality, and computational efficiency.

As the number of sampling steps increases, the KID value decreases substantially, indicating improved alignment between the distribution of generated images and that of the real data. For instance, increasing the number of steps from 10 to 100 results in a significant drop in KID, from 0.0972 to 0.0291. This improvement demonstrates that additional diffusion steps enable the model to better refine image features and reduce generative noise. However, the benefits of increasing the step count diminish beyond a certain threshold. When the number of steps is increased to 500, the KID value changes only marginally to 0.0305. This suggests that while early increments in step count notably enhance image realism, further increases yield only negligible improvements.

In contrast, the computational cost escalates sharply with higher step counts. The time required to generate 50 images rises from 92.37 s at 10 steps to 3395.44 s at 500 steps. This exponential increase in inference time imposes a significant burden on computational resources. Notably, at 50 steps, the generation process requires 350.14 s and achieves a KID of 0.0427. This result reflects a desirable balance between image quality and computational efficiency. While higher step counts continue to improve image fidelity slightly, the marginal gains are outweighed by the substantial increase in time cost.

In conclusion, the findings suggest that 50 sampling steps represent an effective compromise between realism and efficiency. This configuration is capable of producing high-quality porous media images with a relatively low computational burden. Consequently, for practical



(a) Porosity Void Count: Original vs. Within-Range Generated Data

(b) Porosity Void Area: Original vs. Within-Range Generated Data

(c) Porosity Void Count: Original vs. Out-of-Range Generated Data by Porosity

(d) Porosity Void Area: Original vs. Out-of-Range Generated Data by Porosity

Fig. 5. Comparative porosity void metrics among original, within-range, and out-of-range data.

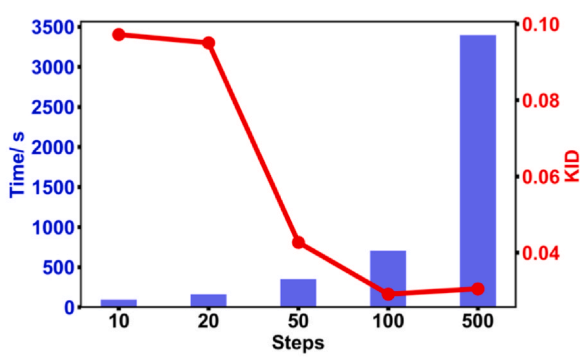


Fig. 6. Effects of sampling steps on inference time and KID score.

applications in which both generation speed and image fidelity are important, selecting a moderate number of diffusion steps, such as 50, is recommended. This choice ensures that the model remains efficient without significantly sacrificing the structural realism of the generated outputs.

5. Discussion

5.1. Impact of data distribution incompleteness

The performance of neural network models is highly dependent on the completeness and representativeness of their training data distributions. This dependency is particularly critical in fields such as geological image analysis, medical diagnostics, and industrial inspection, where data diversity directly influences model reliability. However, in real-world scenarios, achieving complete data coverage is often infeasible due to constraints such as data acquisition costs, equipment limitations, safety concerns, or the natural rarity of certain phenomena. As a result, models are frequently exposed to out-of-distribution (OOD) data during deployment, leading to a significant reduction in prediction reliability. This challenge has prompted a re-assessment of deep learning's generalisation capacity and highlighted data distribution incompleteness as a key bottleneck in high-reliability applications.

This study explores the potential of generative AI, specifically LoRA fine-tuned Stable Diffusion, to mitigate the effects of data distribution incompleteness. The approach aims to improve neural networks' capacity for OOD prediction by supplementing incomplete datasets with statistically consistent synthetic samples. As illustrated in Fig. 7**(a)**, an 80 % − 20 % training-testing split is adopted to emulate real-world data scarcity conditions. During training, both the baseline model trained on incomplete data and the enhanced model using generative augmentation exhibit similar convergence behaviours (Figs. 7**(b) and** 7**(c)**). However, their performance on the test set diverges significantly.

The model trained solely on incomplete data perform poorly on OOD samples, returning high error metrics: MSE = 0.0021, RMSE = 0.0455, MAE = 0.0359, and a negative $R^2 = -1.6824$. As shown in Fig. 7**(d)**, the model exhibites large prediction errors, particularly in regions outside the training distribution. This reflects the fundamental limitation of conventional neural networks in extrapolation tasks, they tend to learn statistical associations within observed data rather than capturing the underlying generative mechanisms.

In contrast, the model enhanced with LoRA fine-tuned Stable Diffusion data achieves significantly better results, with MSE approaching zero, RMSE = 0.0051, MAE = 0.0042, and an $R^2 = 0.9660$. The visualisation in Fig. 7**(e)** clearly illustrates this improvement, with reduced error evident across the results. These results confirm that data generated by generative models can effectively bridge gaps in the original dataset, enabling neural networks to generalise beyond their initial training distribution. Unlike traditional data augmentation methods, such as image flipping, cropping, or rotation, generative models learn the underlying distributional patterns, offering a statistically grounded method to improve robustness in neural networks. This approach is particularly valuable in safety-critical systems, resource-constrained environments, and domains requiring high predictive reliability under data scarcity.

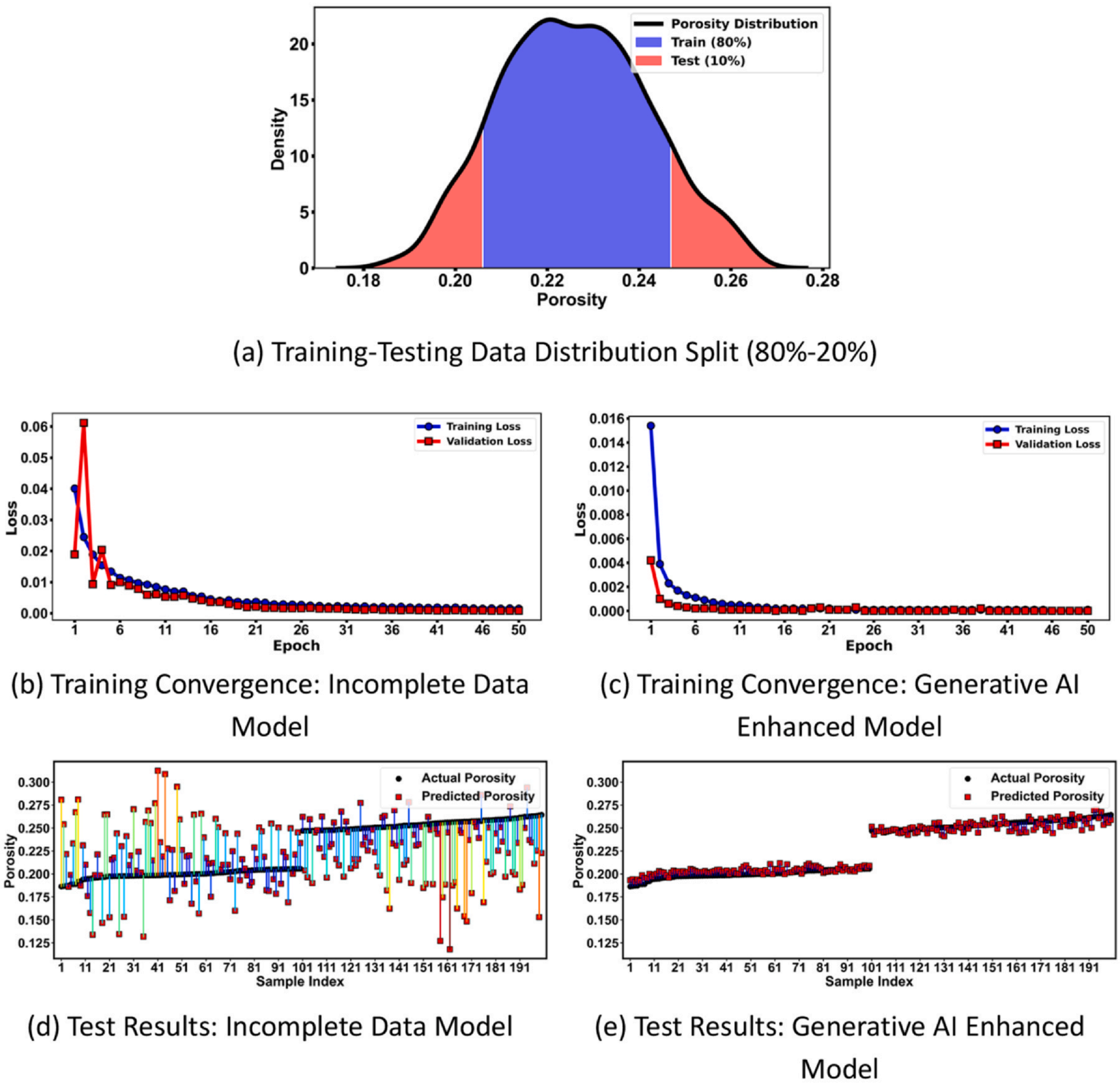


Fig. 7. Comparative analysis of incomplete data model and generative AI enhanced model.

In summary, integrating generative AI into the data pipeline significantly enhances the generalisation capability of neural networks. By learning and replicating the broader data distribution, LoRA-enhanced Stable Diffusion offers a practical and scalable solution to the long-standing challenge of incomplete training data, supporting the development of more robust and deployable machine learning systems.

5.2. Future directions and broader implications

To further illustrate the methodological framework and its broader implications, Fig. 8 presents a conceptual overview of the LoRA fine-tuning stable diffusion pipeline, along with its key advantages, limitations, and future research directions. The pipeline is structured into four primary stages. The first stage, data preparation, involves image collection, preprocessing, and prompt engineering to ensure the input data is suitable for model training and generation. The second stage, LoRA fine-tuning, includes low-rank adaptation and hyperparameter optimisation, enabling efficient fine-tuning with only a small fraction (1–4 %) of the total model parameters. This significantly reduces computational costs while preserving the model's expressive capacity. The third stage, image generation, leverages the stable diffusion model alongside guidance scale tuning and sampling step optimisation to enhance control over image fidelity, realism, and structural accuracy. Finally, the fourth stage, evaluation, assesses the generated outputs using fidelity and diversity metrics to determine their quality and utility for downstream applications.

This approach offers several key advantages. It enables the generation of images with high structural diversity, supports rapid generation due to lightweight fine-tuning, and maintains high fidelity in reproducing geological textures and features. However, the framework also presents notable challenges. The stochastic nature of diffusion models introduces strong randomness, making it difficult to consistently control specific features such as porosity or grain orientation. Additionally, the model is parameter sensitive, requiring careful calibration of tuning parameters (e.g., Network Dim, guidance scale) for different datasets or material types, which may complicate deployment in diverse operational environments.

Despite these limitations, the framework shows immense potential across a broad range of scientific imaging applications. As illustrated in the lower section of Fig. 8, eight promising research directions highlight the versatility of this approach. Data augmentation allows for the expansion of limited datasets by synthesising realistic and diverse porous media images. Multi-scale image simulation supports the generation of consistent images across spatial scales, facilitating pore-scale analysis from nanometres to centimetres. Conditional image generation enables the synthesis of images based on physical parameters such as porosity,

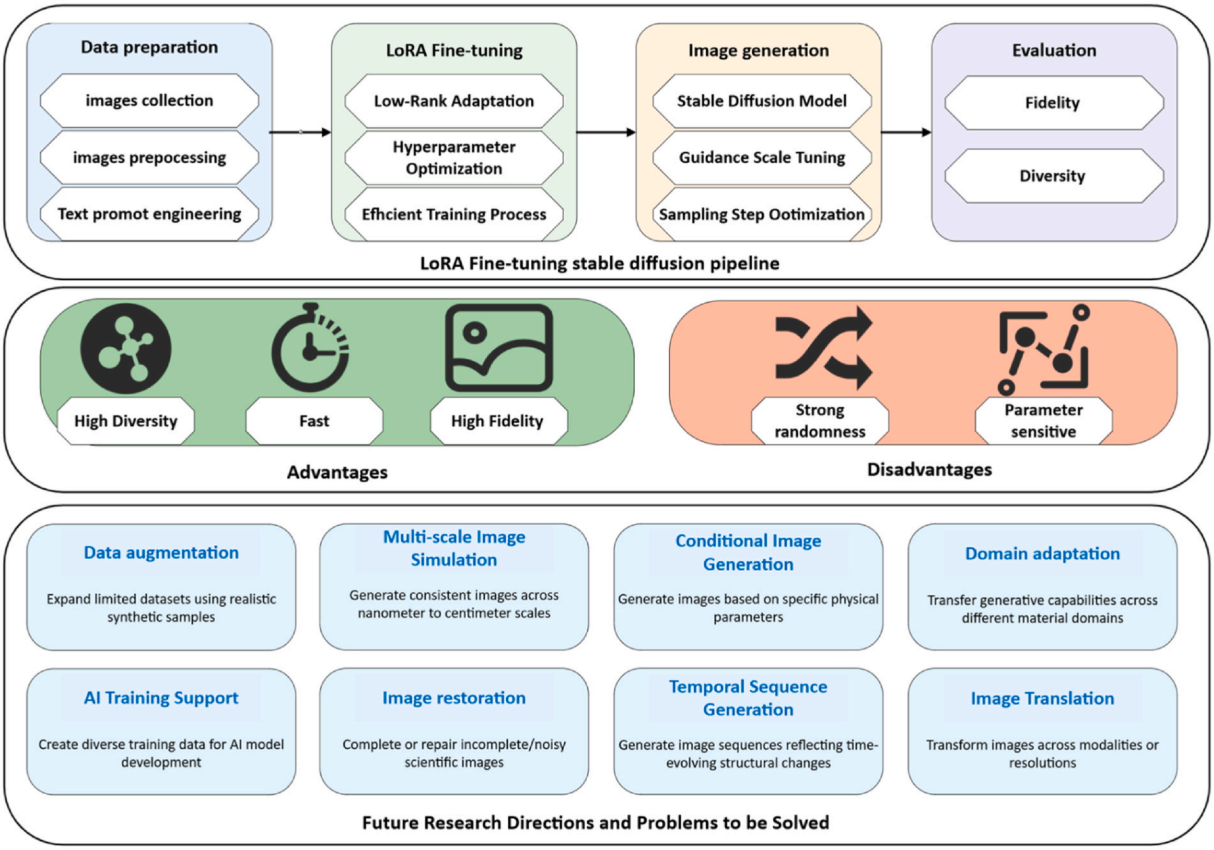


Fig. 8. LoRA-fine-tuned stable diffusion method for porous media image generation: workflow, benefits, and challenges.

aiding in property-specific dataset creation. Domain adaptation allows the model to be transferred across different geological or material domains with minimal retraining. Moreover, the framework can support AI training by generating diverse datasets for model development, and improve image restoration by completing or repairing noisy or incomplete scientific images. It can also produce temporal image sequences, enabling the simulation of time-evolving phenomena such as crack propagation, and perform image translation between different imaging modalities (e.g., CT to SEM) or resolutions, enhancing imaging workflows.

In summary, the LoRA-enhanced stable diffusion pipeline provides a scalable, efficient, and highly adaptable solution for scientific image generation. With continued improvements in algorithmic design, incorporation of physics-based constraints, and integration across imaging modalities, this approach is poised to play a transformative role in digital rock physics, materials science, and related fields.

6. Conclusion

This study has demonstrated that integrating LoRA with a stable diffusion framework offers an efficient and high-fidelity solution for the generation of digital rock CT images. By leveraging a lightweight fine-tuning strategy, the proposed approach significantly reduces computational and memory demands while maintaining high generative quality. Through systematic experimentation, several key findings have emerged that highlight the effectiveness and versatility of this method.

First, LoRA fine-tuning provides a resource-efficient optimisation strategy for adapting diffusion models to domain-specific tasks. The optimal configuration, using a network dimension of 32 and a scaling factor of 16, achieved a 92.6 % improvement in KID compared to a full diffusion model trained from scratch, while modifying only a small fraction of the model's parameters. Second, the guidance scale was found to play a critical role in determining the porosity characteristics of generated images. A guidance scale of approximately 15 yielded porosity distributions most closely aligned with the original CT data. This highlights the importance of prompt conditioning in controlling structural realism during generation. Third, the number of sampling steps introduces a trade-off between image quality and computational efficiency. While increased steps improve fidelity, the gains diminish beyond 50 steps, making it a practical balance point for generation tasks. Fourth, the model exhibits a strong capacity to learn and generalise the underlying relationship between porosity and pore-scale morphology. Notably, when applied to out-of-distribution prediction tasks, the use of LoRA-generated images markedly improved neural network performance, from an R^2 value of –1.6824 to 0.9660, demonstrating the model's utility in addressing data distribution incompleteness and enhancing robustness in extrapolation scenarios.

Collectively, these findings confirm that LoRA-enhanced stable diffusion offers a powerful and flexible generative framework for digital rock analysis. It supports a wide range of downstream applications, including data augmentation, multi-scale simulation, conditional image generation, and domain adaptation, with significant implications for geo-energy research and beyond. While challenges such as parameter sensitivity and output randomness persist, ongoing advances in diffusion modelling, physical constraint integration, and cross-modal learning are expected to further improve controllability and generalisation. In this context, the proposed framework not only bridges critical data gaps in digital rock physics but also lays the groundwork for innovative applications across geosciences, materials engineering, and scientific imaging. As such, it presents a promising direction for future research and practical deployment in both academic and industrial settings.

CRediT authorship contribution statement

Kunyao Li: Methodology, Software, Writing – original draft. **Haijiang Li:** Conceptualization, Funding acquisition, Supervision. **Ali**

Khudhair: Writing – review & editing. **Jun Yan:** Resources, Supervision, Writing – review & editing. **Bin Wang:** Validation, Writing – review & editing.

Data availability

The data utilized in this study were obtained from the Digital Rock Portal (https://www.digitalrocksportal.org/projects/211).

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.

Acknowledgment

This work is part of the Knowledge Transfer Partnership (KTP) project – DIGIBRIDGE: BIM and Digital Twins in support of Smart Bridge Structural Surveying. The project receives funding from Innovate UK (reference number: 10003208) and the China Scholarship Council (Grant No. CSC 202408420030). We would like to express our sincere gratitude to Mohammad Esmaeili for providing the open-source implementation that served as the baseline model in our research.

References

Bińkowski, M., Sutherland, D.J., Arbel, M., Gretton A., 2021 Demystifying MMD GANs 2021. https://doi.org/10.48550/arXiv.1801.01401.

Biswal, B., Øren, P.-E., Held, R.J., Bakke, S., Hilfer, R., 2007. Stochastic multiscale model for carbonate rocks. Phys. Rev. E 75, 061303. https://doi.org/10.1103/PhysRevE.75.061303

Chi, P., Sun, J., Zhang, R., Yan, W., Cui, L., 2024. Digital rock reconstruction enhanced by a novel GAN-based 2D-3D image fusion framework. Adv. Water Resour. 193, 104813. https://doi.org/10.1016/j.advwatres.2024.104813

Choi, J., Kim, S., Jeong, Y., Gwon, Y., Yoon, S., 2021 ILVR: Conditioning method for denoising diffusion probabilistic models 2021. https://doi.org/10.48550/arXiv.2108.02938.

Cnudde, V., Boone, M.N., 2013. High-resolution X-ray computed tomography in geosciences: a review of the current technology and applications. Earth-Sci. Rev. 123, 1–17. https://doi.org/10.1016/j.earscirev.2013.04.003

Dhariwal, P., Nichol, A., 2021. Diffusion models beat GANs on image synthesis. Curran Associates Inc., Red, Hook, NY, USA, pp. 8780–8794.

Diogenes, A.N., dos Santos, L.O.E., Fernandes, C.P., 2008. C.R. Appoloni, Porous media microstructure reconstruction using pixel-based and object-based simulated annealing: comparison with other reconstruction methods. In: Rio de Janeiro, RJ (Brazil).

Esmaeili, M., 2024. Enhancing digital rock analysis through generative artificial intelligence: diffusion models. Neurocomputing 587, 127676. https://doi.org/10.1016/j.neucom.2024.127676

Feng, J., Teng, Q., Li, B., He, X., Chen, H., Li, Y., 2020. An end-to-end three-dimensional reconstruction framework of porous media from a single two-dimensional image based on deep learning. Comput. Methods Appl. Mech. Eng. 368, 113043. https://doi.org/10.1016/j.cma.2020.113043

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Adv. Neural Inf. Process. Syst., MIT Press, pp. 2672–2680. https://doi.org/10.5555/2969033.2969125

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models 2020. https://doi.org/10.48550/arXiv.2006.11239.

Karimpouli, S., Kadyrov, R., Siegert, M., Saenger, E.H., 2024. Applicability of 2D algorithms for 3D characterization in digital rocks physics: an example of a machine learning-based super resolution image generation. Acta Geophys 72, 861–874. https://doi.org/10.1007/s11600-023-01149-7

Kench, S., Cooper, S.J., 2021. Generating three-dimensional structures from a two-dimensional slice with generative adversarial network-based dimensionality expansion. Nat. Mach. Intell. 3, 299–305. https://doi.org/10.1038/s42256-021-00322-1

Li, X., Li, B., Liu, F., Li, T., Nie, X., 2023. Advances in the application of deep learning methods to digital rock technology. Adv. Geo-Energy Res. 8, 5–18. https://doi.org/10.46690/ager.2023.04.02

Liu, K., Qiu, G., Tang, W., Zhou, F., 2020. Spectral regularization for combating mode collapse in GANs. Image Vis. Comput. 104, 104005. https://doi.org/10.1016/j.imavis.2020.104005

Ma, Y., Liao, Q., Yan, Z., You, S., Song, X., Tian, S., Li, G., 2024. Stable diffusion for high-quality image reconstruction in digital rock analysis. Adv. Geo-Energy Res. 12, 168–182. https://doi.org/10.46690/ager.2024.06.02

Ma, Y., Liao, Q., Yan, Z., You, S., Song, X., Tian, S., Li, G., 2024. Stable diffusion for high-quality image reconstruction in digital rock analysis. Adv. Geo-Energy Res. 12, 168–182. https://doi.org/10.46690/ager.2024.06.02

Mariethoz, G., Renard, P., Straubhaar, J., 2010. The direct sampling method to perform multiple-point geostatistical simulations. Water Resour. Res. 46. https://doi.org/10.1029/2008WR007621

Mosser, L., Dubrule, O., Blunt, M.J., 2017. Reconstruction of three-dimensional porous media using generative adversarial neural networks. Phys. Rev. E 96, 043309. https://doi.org/10.1103/PhysRevE.96.043309

Mosser, L., Dubrule, O., Blunt, M.J., 2018. Stochastic reconstruction of an oolitic limestone by generative adversarial networks. Transp. Porous Media 125, 81–103. https://doi.org/10.1007/s11242-018-1039-9

Okabe, H., Blunt, M.J., 2004. Prediction of permeability for porous media reconstructed using multiple-point statistics. Phys. Rev. E 70, 066135. https://doi.org/10.1103/PhysRevE.70.066135

Okabe, H., Blunt, M.J., 2005. Pore space reconstruction using multiple-point statistics. J. Pet. Sci. Eng. 46, 121–137. https://doi.org/10.1016/j.petrol.2004.08.002

ØREN, P.E., Bakke, S., 2002. Process based reconstruction of sandstones and prediction of transport properties. Transp. Porous Media 46, 311–343. https://doi.org/10.1023/A:1015031122338

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-resolution image synthesis with latent diffusion models. IEEE, New Orleans, LA, USA, pp. 10674–10685. https://doi.org/10.1109/CVPR52688.2022.01042

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E.L., Ghasemipour, K., Gontijo Lopes, R., Karagol Ayan, B., Salimans, T., Ho, J., Fleet, D.J., Norouzi, M., 2022. Photorealistic text-to-image diffusion models with deep language understanding. Adv. Neural Inf. Process. Syst. 35, 36479–36494.

Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics. Math. Geol. 34, 1–21. https://doi.org/10.1023/A:1014009426274

Tahmasebi, P., Sahimi, M., 2013. Cross-correlation function for accurate reconstruction of heterogeneous media. Phys. Rev. Lett. 110, 078002. https://doi.org/10.1103/PhysRevLett.110.078002

Torquato, S., 2002. Random heterogeneous materials. Springer,, New York, NY. https://doi.org/10.1007/978-1-4757-6355-3

Wang, Y.D., Armstrong, R.T., Mostaghimi, P., 2019. Enhancing resolution of digital rock images with super resolution convolutional neural networks. J. Pet. Sci. Eng. 182, 106261. https://doi.org/10.1016/j.petrol.2019.106261

Yang, Y., Wang, J., Liu, F., Zhang, L., Sun, H., Zhong, J., Zhang, K., Yao, J., 2024. Super-resolution of digital rock images with hybrid attention multi-branch neural network. Gas. Sci. Eng. 128, 205395. https://doi.org/10.1016/j.jgsce.2024.205395

Yeong, C.L.Y., Torquato, S., 1998. Reconstructing random media. Phys. Rev. E 57, 495–506. https://doi.org/10.1103/PhysRevE.57.495

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

Zhao, B., Saxena, N., Hofmann, R., Pradhan, C., Hows, A., 2023. Enhancing resolution of micro-CT images of reservoir rocks using super resolution. Comput. Geosci. 170, 105265. https://doi.org/10.1016/j.cageo.2022.105265