Advances in Geo-Energy Research

Vol. 12, No. 3, p. 168-182, 2024

Original article

Stable diffusion for high-quality image reconstruction in digital rock analysis

Yutian Ma[1], Qinzhuo Liao[2]*, Zhengting Yan[2], Shaohua You[2], Xianzhi Song[2], Shouceng Tian[1], Gensheng Li[2]

[1]*College of Petroleum, China University of Petroleum (Beijing) at Karamay, Karamay 834000, P. R. China*

[2]*National Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing), Beijing 102249, P. R. China*

Keywords:
Stable diffusion
digital rock analysis
image generation
artificial intelligence
deep learning

Cited as:
Ma, Y., Liao, Q., Yan, Z., You, S., Song, X., Tian, S., Li, G. Stable diffusion for high-quality image reconstruction in digital rock analysis. Advances in Geo-Energy Research, 2024, 12(3): 168-182.
https://doi.org/10.46690/ager.2024.06.02

Abstract:
Digital rock analysis is a promising approach for visualizing geological microstructures and understanding transport mechanisms for underground geo-energy resources exploitation. Accurate image reconstruction methods are vital for capturing the diverse features and variability in digital rock samples. Stable diffusion, a cutting-edge artificial intelligence model, has revolutionized computer vision by creating realistic images. However, its application in digital rock analysis is still emerging. This study explores the applications of stable diffusion in digital rock analysis, including enhancing image resolution, improving quality with denoising and deblurring, segmenting images, filling missing sections, extending images with outpainting, and reconstructing three-dimensional rocks from two-dimensional images. The powerful image generation capability of diffusion models shed light on digital rock analysis, showing potential in filling missing parts of rock images, lithologic discrimination, and generating network parameters. In addition, limitations in existing stable diffusion models are also discussed, including the lack of real digital rock images, and not being able to fully understand the mechanisms behind physical processes. Therefore, it is suggested to develop new models tailored to digital rock images for further progress. In sum, the integration of stable diffusion into digital core analysis presents immense research opportunities and holds the potential to transform the field, ushering in groundbreaking advances.

1. Introduction

Digital rock analysis presents the microstructure of underground rocks by digital scanning and three dimensions (3D) reconstruction of core samples, to achieve accurate description and prediction of reservoir properties (Li et al., 2023). It plays an irreplaceable role in the study of unconventional reservoirs with complexity and heterogeneity, such as tight sand and shale (Goral et al., 2020). The essence of digital core technology lies in its image generation and processing capabilities (Zha et al., 2021). In order to accurately capture the pore network and particle distribution inside the rock, advanced image reconstruction techniques are required, to provide a 3D visualization of the rock samples and data support for analyzing the geometric characteristics (Liao et al., 2022).

Enhancing image resolution, identifying lithology, and reconstructing 3D digital rocks are critical for advancing digital rock analysis beyond traditional methods' limitations (Li et al., 2023). Recent advancements in deep learning have significantly contributed to these areas. Due to limitations inherent in imaging technologies like micro-computed tomography (micro-CT), high-resolution images often feature a smaller field of view, whereas images with larger fields of view tend to have lower resolution (Ahuja et al., 2022). To address this challenge, super-resolution convolutional neural networks in deep learning, alongside derivative algorithms such as enhanced deep super-resolution network and widely activated super-resolution network, have proven effective in enhancing

Yandy Scientific Press

*Corresponding author.
E-mail address: 2021016092@st.cupk.edu.cn (Y. Ma); liaoqz@cup.edu.cn (Q. Liao); yanzhengting22@163.com (Z. Yan); youshaohua01@163.com (S. You); songxz@cup.edu.cn (X. Song); tscsydx@163.com (S. Tian); ligs@cup.edu.cn (G. Li).

Received February 27, 2024; revised March 14, 2024; accepted April 5, 2024; available online April 9, 2024.

micro-CT image resolution for sandstone and carbonate rocks, improving quality by 3-5 dB and reducing errors by 50%-70% compared to bicubic interpolation methods (Wang et al., 2019). For lithological identification, Convolutional Neural Network (CNN) models have been utilized for geological facies classification in well logging, suitable for conventional well logging data processing, providing a robust method for lithology identification (Imamverdiyev and Sukhostat, 2019). Li et al. (2024) proposed a hybrid model integrating CNN with Support Vector Machine for shale reservoir lithology identification, achieving 91.95% accuracy across three lithologies: Mudstone, shale, and sandstone, demonstrating superior generalization capabilities. In 3D digital rock reconstruction, traditional methods face challenges such as cost and accuracy limitations (Zhao et al., 2021). Deep learning techniques like Generative Adversarial Networks (GAN) and Variational Autoencoders (VAE) can generate realistic rock samples from extensive image datasets, improving reconstruction efficiency and accuracy (Chi et al., 2023). Feng et al. (2020) employed GAN for 3D digital rock reconstruction from 2 dimensions (2D) slices, demonstrating faster reconstruction speeds compared to traditional methods. GAN has been a common choice in digital rock image generation (Liu et al., 2024), It consists of two components, a generator and a discriminator, both engaging in competition to acquire the ability to generate high-quality images. However, despite the success of GAN in some applications, their limitations are becoming apparent in the field of digital core. As it involves the simultaneous training of two networks, problems such as mode collapse often occur, which leads to a lack of diversity in the generated rock images.

Stable diffusion (SD), as a kind of latent diffusion models, since its first release in August 2022, has rapidly emerged in the field of artificial intelligence-generated content (AIGC). Compared with GAN, SD uses an autoregressive training process that does not involve a zero-sum game between two networks, and the training process is relatively more stable. It is gradually showing the trend of replacing traditional GAN with its excellent image generation ability and wide range of application scenarios. In the generation process, SD iteratively transforms the noisy image into the target image, which is a more controllable and stable process (Croitoru et al., 2023), and is less likely to produce the pattern noise problem. In addition, since SD is based on Markov chains, the generated samples can cover the entire data distribution, the results tend to be more diverse (Gong et al., 2023). One of the most significant advantages of SD over many other generative models (e.g., Midjourney, Dall-E, etc.) lies in its open-source nature, which means that researchers can further investigate its internal working mechanism and extend it with new modules according to specific needs. It is also highly extensible, which includes facilitating the integration of super-resolution algorithms to enhance image quality further, leveraging ControlNet for finer control over the generation process, and amalgamating data from various modalities for multi-modal generation. It can also be extended to the field of video generation to produce high quality video content, which has wide application in various domains including illustration, game design, and electronic commerce.

Despite these advancements, the application of SD in digital rock remains sparse and incomplete. Only two articles have been reported for the application of SD in the field of digital rock analysis, to the best of our knowledge. Ma et al. (2023a) proposed the concept of utilizing SD as a data augmentation technique, generating new images akin to the original dataset by simulating the process of adding and eliminating random noise. Applied to image segmentation tasks in digital rock physics, this method significantly enhances the segmentation performance of neural network models such as U-Net, Attention-U-Net, and TransUNet, particularly adept at handling images with intricate pore morphology. Ma et al. (2023b) utilized SD to enhance micro-CT image resolution of rock samples, which employs progressive diffusion noise processing and cascaded diffusion models for accelerated training and high-fidelity image production. They explored SD's application in single-image super-resolution, introducing shortcut sampling for diffusion. Additionally, they discussed using diffusion models in blind super-resolution methods, emphasizing their effectiveness in capturing long-range dependencies in images. The limited number of studies indicates that there is still much unexplored potential in this field. Many functionalities of SD have yet to be fully explored and developed for digital rock analysis. Future research should focus on further exploring the capabilities of SD in enhancing rock image resolution, denoising, image expansion, and other related tasks. Additionally, there is a need to address the challenges associated with the application of SD in digital rock.

In this study, the principle of SD for image generation and its application in digital rock analysis are explored (Fig. 1). The basic principle of SD image generation is to start from pure Gaussian noise and gradually denoise to reconstruct the image. Its potential applications in the digital rock domain include enhancing image resolution, denoising, image expansion, text to image, and digital core reconstruction from 2D to 3D, etc. For super-resolution reconstruction and improving image quality through denoising, the results show that SD is more capable of restoring the details and realism of digital rock images while preserving the structure of the original image. Also, it is observed that SD shows good performance in image expansion tasks, which brings a new perspective to digital core analysis. It has the potential to rapidly extend the generation of full-field data using local information, thereby reducing scanning costs. The text to image function of SD also provides a new approach for generating digital core images by given prompts. Due to the lack of generalization ability of pre-trained models in the digital core domain, there is still a need to fine-tune or retrain the models in terms of rebuilding 3D digital rock based on 2D images. Further investigation reveals the possibilities of diffusion models for filling in missing images, lithologic discrimination, and even in designing algorithms. It is found that diffusion models themselves are undergoing continuous development and optimization. In sum, SD shows promising applications in the field of digital core analysis.

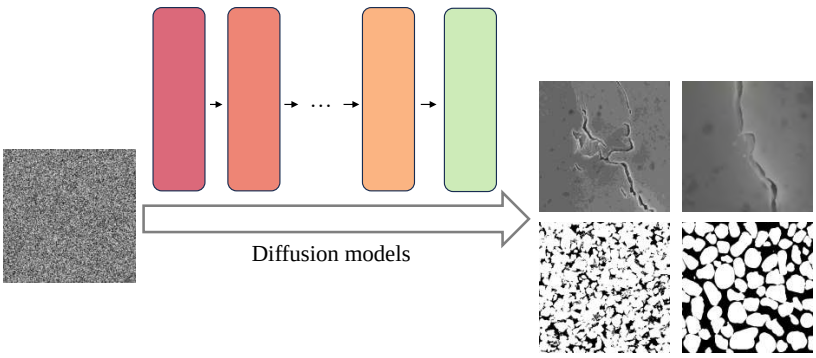

Fig. 1. Diagram of the generating digital rock images using stable diffusion.

2. Method

The wide application of SD benefits from the theoretical support of diffusion models. At present, diffusion models can be categorized into three groups (Yan et al., 2024): Denoising diffusion probabilistic models (DDPM), score-based generative models, and stochastic differential equations (SDE). DDPM achieves efficient and high-quality sample generation by introducing time-dependent noise and variational lower bound optimization objectives. Here, DDPM will be taken as the main entry point to explore the fundamentals of diffusion models.

DDPM is a generative model based on a diffusion process, which learns the essential features of the data distribution through stepwise noise addition and denoising process, to generate samples similar to the original data distribution (Ho et al., 2020). It comprises two parameterized Markov chains: A forward process and a reverse process. Utilizing variational inference, it generates samples that align with the original data distribution within a finite timeframe (Fig. 2).

In the forward process, given the initial data distribution $x_0 \sim q(x)$, for time $t \in [1,T]$, DDPM will gradually increase Gaussian noise according to the x_{t-1} obtained from the previous step as (Chang et al., 2024):

$$\begin{aligned} &x_t = \sqrt{\beta_t} \times \varepsilon_t + \sqrt{1-\beta_t} \times x_{t-1}, \quad \varepsilon \sim \mathcal{N}(0,\mathbf{I}), \\ &0 < \beta_1 < \beta_2 < \beta_3 < \beta_{t-1} < \beta_t < 1 \end{aligned} \tag{1}$$

where ε_t is the noise obtained by random sampling based on the standard normal distribution. $\beta_t \in (0,1)$ is the variance used for each step. The variances of different steps are set in advance as a variance schedule, β is gradually incremented in each step, indicating that the diffusion speed becomes faster. When the diffusion steps T is large enough, the final image x_T converges to a pure noise (Ho et al., 2020). Since the low efficiency of stepwise iteration, $\alpha_t := 1-\beta_t$ and $\overline{\alpha_t} := \prod_{s=1}^{t} \alpha_s$ are introduced. After reparameterization, the entire forward process can be described as:

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{\overline{\alpha}_t}\mathbf{x}_0, (1-\overline{\alpha}_t)\mathbf{I}\right) \tag{2}$$

where $\mathbf{I}$ is an identity matrix with the same dimension as x_0. In this way, given x_0, the denoised image x_t can be obtained at any time.

The reverse process starts at $x_T \sim N(0,I)$ and denoising is performed as (Mao et al., 2023):

$$\begin{aligned} &\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left[\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}\varepsilon_\theta(\mathbf{x}_t,t)\right] + \sigma_t\mathbf{z}, \quad \sigma_t \sim N(0,1), \\ &\begin{cases} \mathbf{z} \sim \mathcal{N}(\mathbf{0},\mathbf{I}), & t > 1 \\ \mathbf{z} = \mathbf{0}, & t \le 1 \end{cases} \end{aligned} \tag{3}$$

where $\varepsilon_\theta(\mathbf{x}_t,t)$ is the model that estimates the noise from $\mathbf{x}_t$ and t, and θ is the training parameter for the model, σ_t is a hyperparameter that follows a normal distribution, and $\mathbf{z}$ is a Gaussian noise. The whole reverse process can be represented as:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}\left(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t,t), \sigma_t^2\mathbf{I}\right) \tag{4}$$

In order to make the generated image match the training data distribution as closely as possible, $\varepsilon_\theta(\mathbf{x}_t,t)$ needs to be trained so that its prediction $\hat{\varepsilon}$ is similar to the true ε. The solution of minimizing the usual variation bound on negative log likelihood is used in DDPM:

$$\begin{aligned} &\mathbb{E}_q[\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_T \mid \mathbf{x}_0)\|p(\mathbf{x}_T))}_{L_T} \\ &+ \sum_{t>1}\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_{t-1} \mid \mathbf{x}_t,\mathbf{x}_0)\|p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}} \underbrace{-\log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0}] \end{aligned} \tag{5}$$

where Kullback-Leibler (KL) divergence is used to compare $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ with forward process posteriors. As all KL divergences in Eq. (5) involve comparisons between Gaussians, they can be computed in a Rao-Blackwellised manner using closed form expressions. Since L_T is an independent constant with no reliance on θ , it can be ignored during training. Ho et al. (2020) used an independent discrete decoder derived from $\mathcal{N}\left(\mathbf{x}_0; \mu_\theta(\mathbf{x}_1,1), \sum_\theta(\mathbf{x}_1,1)\right)$ to simulate L_0. In actual training, they found the following variant of the variational bounds beneficial for sample quality:

$$L_{\text{simple}}(\theta) := \mathbb{E}_{t,\mathbf{x}_0,\varepsilon}\left[\left\|\varepsilon - \varepsilon_\theta\left(\sqrt{\overline{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\overline{\alpha}_t}\varepsilon, t\right)\right\|^2\right] \tag{6}$$

Therefore, DDPM takes Eq. (6) as loss function and uses U-Net to train the noise parameters.

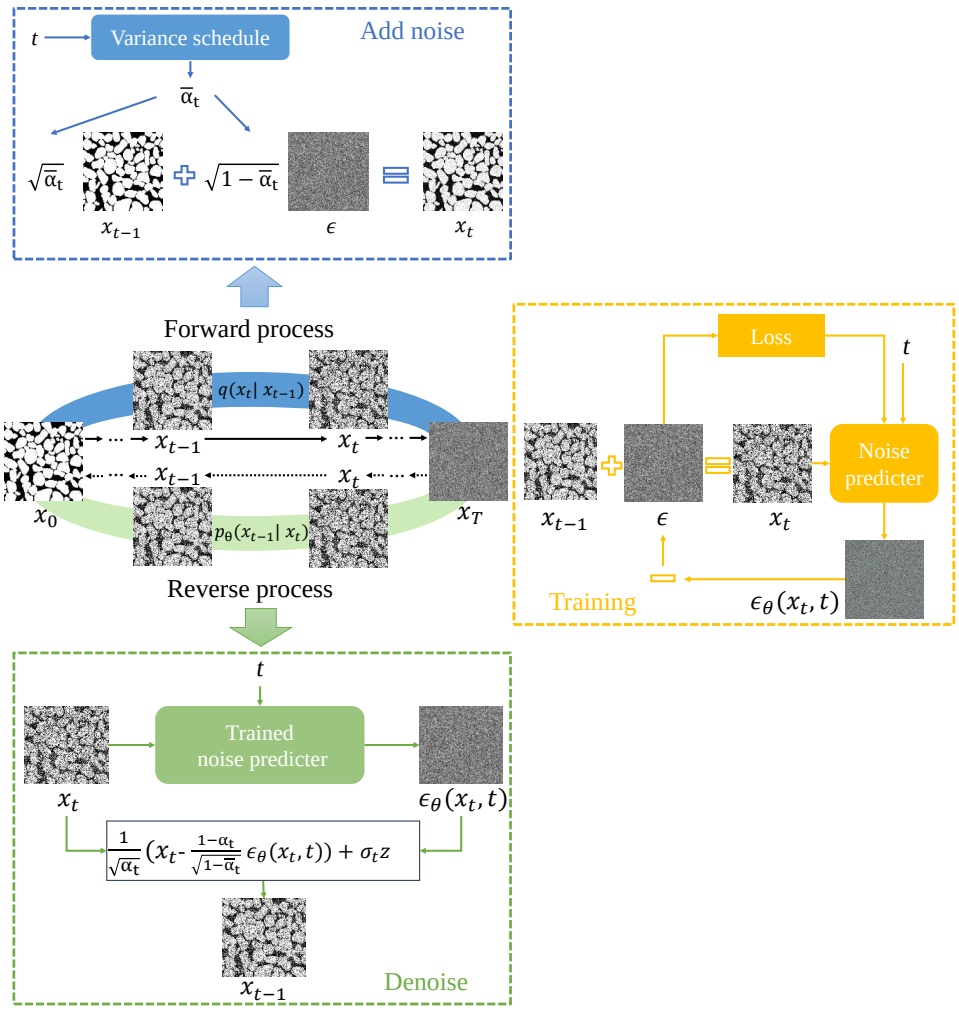

Fig. 2. Process of DDPM.

As a practical application software of diffusion models, SD introduces VAE to achieve the mapping of images to low-dimensional latent space (Rombach et al., 2022), thereby enhancing the efficiency of image generation and processing. Furthermore, SD incorporates modules such as cross-attention layers and the contrastive language-image pre-training text encoder to enhance its adaptability to textual inputs. The SD network structure orchestrates the interaction between input data, conditioning mechanisms, denoising U-Net, and latent space to enable the generation of high-quality images.

3. Applications and results

3.1 Super-resolution

In the process of digital rock imaging and data processing, resolution plays a pivotal role in capturing image details and ensuring the precision of subsequent analyses. High-resolution digital rock images are instrumental in accurately discerning minute features and variations in the core, which is crucial to the study of physical properties of rocks, fluid flow behavior and reservoir evaluation (Chen et al., 2020). By continuously optimizing imaging techniques to enhance the resolution of rock images, significant strides will be made in acquiring more precise and reliable geological information, thereby enhancing the efficacy of oil and gas exploration (Shan et al., 2022).

Diffusion models provide an innovative solution for image super-resolution reconstruction. When generating an image, the random walk process is iteratively executed in diffusion models, and at each iteration, the pixel values are updated with specific transition probability based on the current pixel value and the state of the surrounding pixels (Xie et al., 2023). Thus, they can progressively increase the details and clarity on low-resolution images, ultimately generating high-resolution images. During this process, they can learn more structural information from the input image and try to maintain these structures as much as possible while enhancing the resolution. Therefore, they can generate more natural image reconstruction results and reduce the potential introduction of artifacts and distortions (Zeng et al., 2023). Additionally, the diffusion models can better handle nonlinear transformations in images, making them more suitable for processing complex images. Furthermore, their strong scalability enables them to integrate with other techniques to achieve higher quality production of rock images.

It is a technical challenge to improve the efficiency and e-

Table 1. Stable diffusion parameters for image generation.

Method	Parameters	Value
SD	Sampler	DPM++2M SDE Karras
	Image size	960×800
	Steps	49
	CFG scale	1.5
	Denoising strength	0.03
Tiled diffusion	Method	MultiDiffusion
	Latent space block size	96×96
	Latent space block overlap	78
	Latent space block batch size	4
Tiled VAE	Encoder block size	3,072
	Decoder block size	192

nsure the quality of large rock image processing. Tiled diffusion and tiled VAE, as an extended application of SD, can ensure the overall picture quality while achieving the efficient use of computing resources. Tiled diffusion (Bar-Tal et al., 2023) is an image generation method based on diffusion models. It adopts a block-based processing approach, and utilizes two state-of-the-art diffusion tiling algorithms (i.e., mixture of diffusers and multi-diffusion), as well as the tiled VAE algorithm and tiled noise inversion. The method divides the image into multiple small blocks/tiles and applies independent diffusion processes to each small block. Then, it uses information from neighboring small blocks as conditions to aid in generation. This local contextual condition significantly enhances the generated results, maintaining the coherence of the image while efficiently optimizing graphics memory usage.

Tiled diffusion and tiled VAE are applied to the restore task of SD combined with the Swin-Conv-UNet (SCUNet) super-resolution algorithm. Realistic Vision V2.0 is chosen as the base model, and the DPM++ 2M SDE Karras sampler is adopted. In the image restoration process, the denoising strength parameter determines how much the model depends on the original image when generating. By adjusting it, a balance can be struck between preserving image details and avoiding excessive smoothing. The denoising strength is finally set to 0.03 for performing high-definition restoration on 10 sets of crack scanning electron microscope (SEM) images. SCUNet is a deep learning model designed for image super-resolution reconstruction (Zhang et al., 2023a). By learning the mapping relationship between numerous low-resolution images and high-resolution images, it can transform low-resolution images into high-resolution ones. SD and SCUNet are integrated to achieve a comprehensive image enhancement process. Initially, SD is employed to denoise and enhance the input image. Subsequently, the image processed by SD is fed into SCUNet. By leveraging SCUNet's powerful super-resolution reconstruction capability, the image is further enlarged and refined. The generation parameters of stable diffusion are listed in Table 1.

A total of six algorithms, i.e., nearest interpolation, Lanczos interpolation, enhanced super-resolution GAN (ESRGAN), blind super-resolution GAN (BSRGAN), SD and SD combined with SCUNet, were used to restore the SEM images (Fig. 3), and the key parameters are listed in Table 2.

To quantitatively evaluate the image quality, the structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) are compared in Table 3. SSIM is a quality assessment metric based on structural similarity, which measures the similarity between the original image and the distorted image by comparing the luminance, contrast, and structural information of the images. PSNR is an objective quality assessment metric based on pixel differences, which quantifies the loss of image quality by calculating the mean square error between the original image and the distorted image. In general, a larger value of SSIM or PSNR indicates better image quality. The mean and standard deviation of SSIM and PSNR for 10 sets of images are calculated, utilizing scaling factors of 2, 4, and 8 times. The bold value means the best result in all algorithms.

Permeability values are then investigated to further compare different super-resolution approaches from the perspective of pore structure. Specifically, the images in Fig. 4 are translated to binary images using certain threshold such that the porosity values are the same for all images after binarization. Then, the permeability is computed by solving Stokes flow equation, assuming a constant pressure difference, and hence the permeability is proportional to the flow rate. The permeability values for different images are: High-resolution 2.51 D, low-resolution 5.12 D, nearest 2.39 D, Lanczos 2.96 D, BSRGAN 2.26 D, SD 2.62 D, where D means Darcy. It can be seen that compared to the true reference (high-resolution value) the SD provides the closest estimation, followed by the nearest interpolation. The low-resolution yields the largest error, as anticipated. More parameters for pore structure representation, e.g., pore throat size distribution, will be further analyzed in the future.

These results reveal that compared to traditional interpolation-based super-resolution algorithms (e.g., nearest interpolation), SD does not simply smooth the image by interpolation when upscaling images. Instead, it is capable of learning complex structures and information within the image and retains these details throughout the upscaling process. In contrast to other super-resolution algorithms like BSRGAN, the employment of SD for image enlargement typically mitigates the occurrence of excessive smoothing artifacts. This ability facilitates the generation of more natural image content while effectively enhancing resolution. Through comprehensive analysis, it can be concluded that SD not only maintains the good effect of SSIM and PSNR metrics but also better retains the structure of the original image, enhances details, and exhibits greater authenticity.

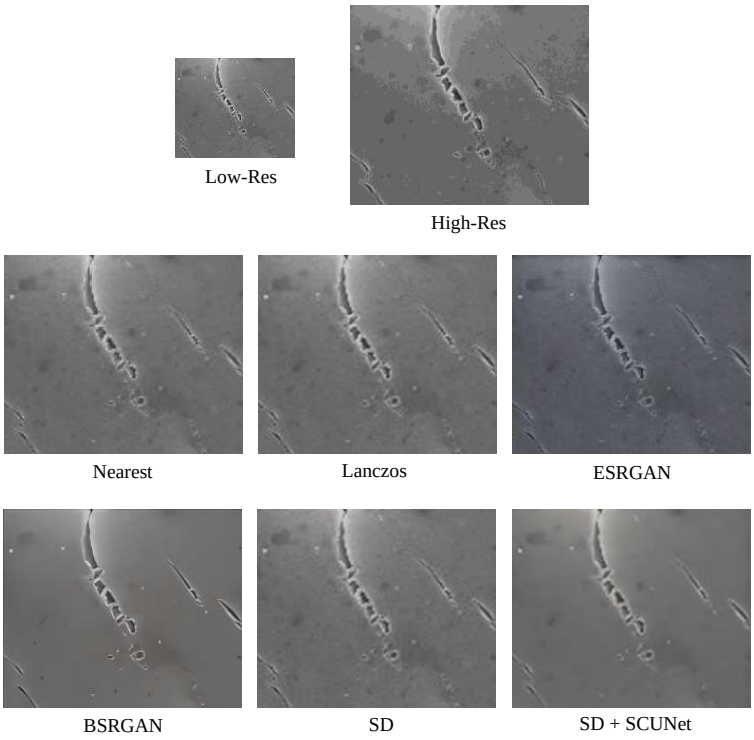

Fig. 3. Super-resolution reconstruction results of SEM images.

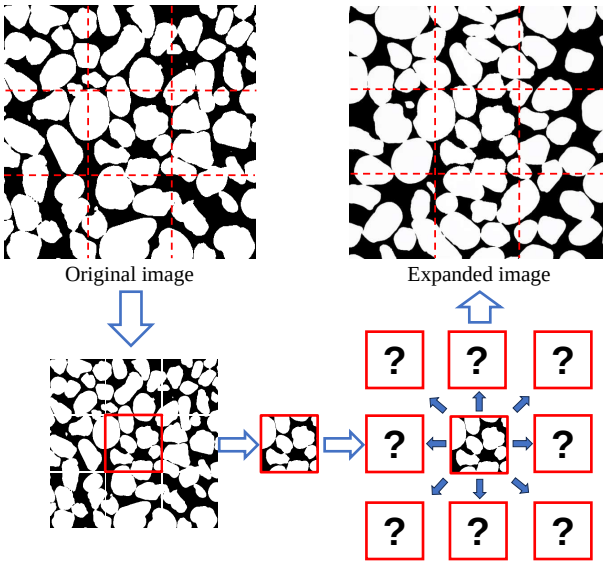

Fig. 4. Expansion of the sandpack image: The original image is cut into 9 pieces and the center part is expanded to generate the whole image.

3.2 Image expansion

Digital core analysis allows researchers to obtain detailed structural and property information of cores in a non-destructive way (Zhao et al., 2020). However, in practice, due to physical limitations or the high cost of scanning equipment, partial images of the core are typically obtained, and the unknown areas exhibit some spatial randomness (Li et al., 2023). In stable diffusion, the image expansion task can be performed to extend the original image and generate missing regions caused by equipment limitations. The model first learns the statistical regularities of structures, colors, and other features in images by training on a large dataset. Then, during the completion process, it utilizes this prior knowledge to predict and fill in missing areas, thereby generating visually complete and consistent images. Currently, Image Expansion shows broad application prospects in various fields such as AIGC, film and television production and medical image processing.

Sandpack and sandstone CT images are extended using the inpaint model of stable diffusion ControlNet (Zhang et al., 2023b). The inpaint model takes inputs including the image to be extended and conditional inputs (such as textual descriptions of the areas to be filled). Text is converted into embedding vectors and then mapped to the internal representation of the inpaint model to add constraints, guiding the generation of images consistent with the text descriptions. Considering that the training image set of Realistic Vision may not include core images, our own low-rank adaptation (LoRA) was retrained. The parameters employed in the experiments are listed in Table 4.

Subsequently, the trained LoRA is embedded into Realistic Vision for fine-tuning, and ControlNet is incorporated for further control. Compared to traditional methods of image generation, ControlNet's local inpaint model not only considers the information of the repaint area itself but also utilizes contextual information from surrounding pixels to blend the repaint area with the overall picture, helping to avoid obvious seams and ensuring smooth transitions at the boundaries.

Realistic Vision serves as the base model, embedding the

Table 2. Network structure parameters of super resolution algorithms.

Method	Parameters	Value
ESRGAN-Discriminator	Number of input channels	3
	Number of base intermediate feature channels	64
	Skip connection	True
ESRGAN-Generator	Number of input channels	3
	Number of output channels	3
	Number of base intermediate feature channels	64
	Number of convolutional layers	16
	Scale factor	4
	Activation function	PReLU
BSRGAN-Discriminator	Number of input channels	3
	Number of discriminator filters	64
	Number of discriminator layers	3
	Normalization type	Batch normalization
	Activation function	LeakyReLU
BSRGAN-Generator	Number of input channels	3
	Number of output channels	3
	Number of base intermediate feature channels	64
	Growth channels in ResidualDenseBlock_5C	32
	Number of RRDB blocks	23
	Scale factor	4
SCUNet	Number of input channels	3
	Residual block configuration	2, 2, 2, 2, 2, 2, 2
	Dimension of features	64
	Activation function	ReLU

trained LoRA and the control_v11p_sd15_inpaint model, and expanding the images by 9 times (Figs. 4 and 5).

It can be observed that natural transitions in the boundary regions of the expanded images are allowed by SD, enabling fine adjustments and blending.

During the training process, it learns features from a large amount of data, enabling it to recognize and understand complex structures within images. Actually, the SD is not able to reproduce the original image exactly (actually it would be weird if it is able to), but the properties statistically estimated from the expanded image are close to those from the original image, e.g., porosity, permeability, and spatial relation. In the task of image expansion, the model can utilize this learned knowledge to generate content similar to the regularities of the original image, inferring the whole from the parts, thereby achieving high-quality image expansion.

3.3 Image denoising

The denoising of digital core images forms the basis for subsequent image analysis. Denoising algorithms need to remove noise while preserving as much detail as possible, thus providing a clearer and more accurate dataset for further research. The principle of denoising in the diffusion models is mainly based on forward noise addition and reverse denoising processes. In the forward process, the model gradually adds noise to the image until it is nearly destroyed and becomes almost pure noise, hiding useful information within the noise (Pan et al., 2023). Then, in the reverse process, it learns the probability distribution of pixel values in the image and uses this distribution for denoising, continuously adjusting the pixel values in the image to recover the original image from a state of pure noise. This probabilistic approach ensures that the denoising results are more natural and realistic.

Here, the noise ratio is set to 80%, and the white noise process is applied to the sandstone images. Then, the noisy

Table 3. Image quality assessment using different algorithms.

Algorithm	SSIM			PSNR		
	×2	×4	×8	×2	×4	×8
Nearest	0.755 ± 0.136	0.731 ± 0.126	0.742 ± 0.120	29.638 ± 2.913	28.629 ± 2.608	27.828 ± 1.922
Lanczos	0.778 ± 0.127	0.759 ± 0.120	0.776 ± 0.111	30.643 ± 3.028	29.631 ± 2.751	28.675 ± 1.911
ESRGAN	0.761 ± 0.129	0.710 ± 0.143	0.703 ± 0.146	21.525 ± 0.718	21.294 ± 0.736	21.152 ± 0.721
BSRGAN	0.787 ± 0.115	0.769 ± 0.113	0.775 ± 0.107	29.607 ± 2.286	28.385 ± 2.130	26.829 ± 1.380
SD	0.746 ± 0.129	0.761 ± 0.119	0.778 ± 0.109	29.670 ± 2.947	29.428 ± 2.698	28.651 ± 1.856
SD+SCUNet	0.791 ± 0.113	0.774 ± 0.122	0.790 ± 0.104	29.679 ± 2.142	28.908 ± 2.405	28.700 ± 1.538

Table 4. Parameters of LoRA network for image expansion.

Parameters	Value
Image size	256×256
Data amount	400
Max epochs	20
Batch size	1
Learning rate	1.00×10^{-4}
Learning rate scheduler	Cosine with restarts
Learning rate warmup steps	0
Learning rate scheduler num cycles	1
Optimizer type	AdamW8bit
Clip skip	2
Mixed precision	Fp16
Network dim	32
Network alpha	32

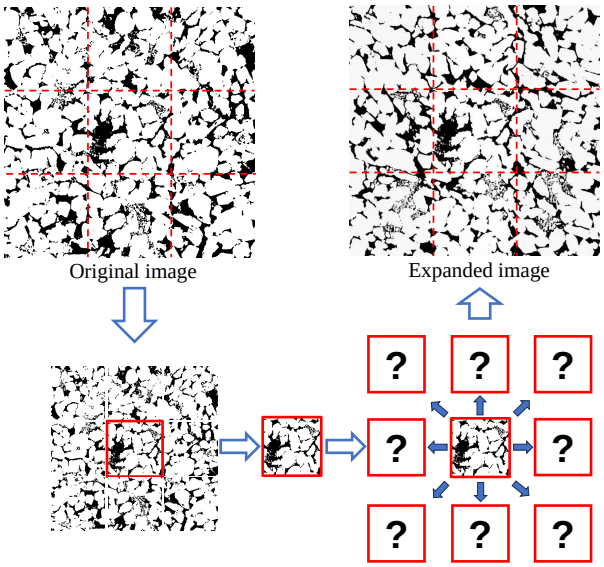

Fig. 5. Expansion of the sandstone image: The original image is cut into 9 pieces and the center part is expanded to generate the whole image.

image is denoised using SD, as well as mean filtering, Wiener filtering, median filtering, and autoencoder (Fig. 6). For autoencoder, 700 sandstone images of size 240×240 are used for training. In the case of mean filtering, Wiener filtering, and median filtering, the appearance of black borders around the generated images is primarily due to the convolution operation typically involved in traditional denoising methods. During convolution operations, to ensure the correctness of the algorithm and the handling of boundary pixels, the edges of the image are often padded with zero values. Since zero values typically correspond to black pixels in color images, the edge areas of the image will exhibit black borders after convolution processing. Therefore, when using such denoising methods in practice, appropriate edge handling methods need to be combined.

The SSIM and PSNR are calculated for SD (SSIM: 0.632, PSNR: 11.211), mean filtering (SSIM: 0.392, PSNR: 10.785), Wiener filtering (SSIM: 0.648, PSNR: 11.765), median filtering (SSIM: 0.662, PSNR: 10.885), and autoencoder (SSIM: 0.638, PSNR: 12.486). And the results indicate that mean filtering performs poorly, while the effect of SD is comparable to Wiener filtering, median filtering and autoencoder. Through the analysis of denoising results, it is evident that after mean filtering, a large number of noise points remain uncleared from the image. After wiener filtering, burr and roughness appear at the edge of the pore, which may affect the accurate identification and analysis of pore structures. After the application of median filtering, although most noise points in the image are effectively removed, an over-smoothing phenomenon occurs along the pore edges, resulting in the loss of significant amounts of detailed information. Additionally, some pores are mistakenly connected together, significantly impacting the authenticity and accuracy of the pore network. For autoencoder, the repaired images of the pore structure have become distorted and blurred. However, while SD fails to completely remove all noise points, it excels in preserving pore details and authenticity. The image processed by SD exhibits clearer pore edges, closely resembling the real microscopic structure of sandstone, and avoids excessive smoothing along the pore edges, thus retaining critical reservoir information.

In addition to binary images, the denoising capability of SD is also tested using grayscale images (Fig. 7). In terms

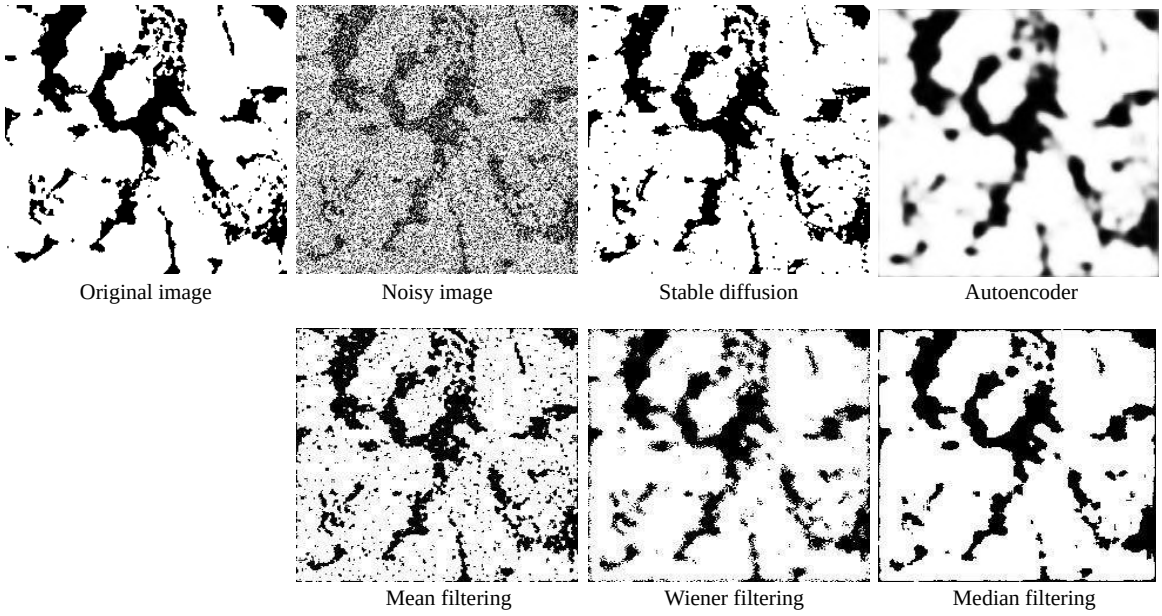

Fig. 6. Restoration of sandstone images by different denoising algorithms.

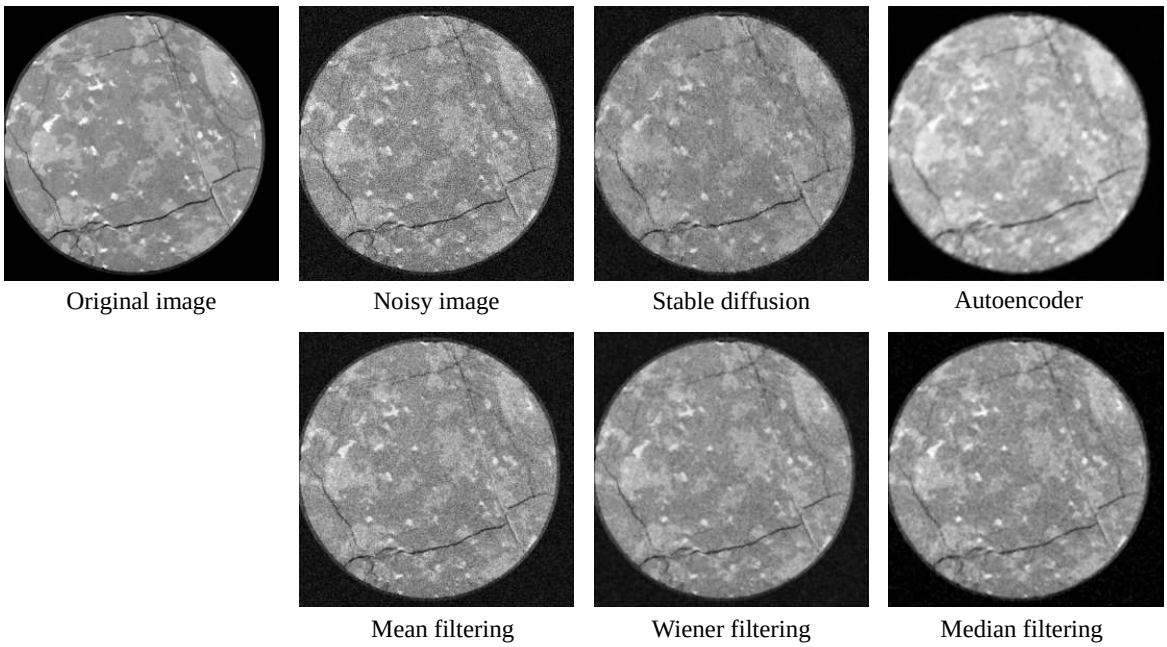

Fig. 7. Restoration of grayscale images by different denoising algorithms.

of quantitative metrics, the autoencoder performs well, with restoration results numerically surpassing SD and conventional filtering methods. However, its restored images do not well preserve the morphological characteristics of cracks. Combining SSIM, PSNR and actual denoising results analysis, the SD shows good performance in maintaining overall clarity and image details.

3.4 Reconstruction from 2D to 3D

After obtaining the 3D digital rock model, researchers can gain a better understanding of key parameters of reservoir such as porosity and permeability, and thus predict and evaluate oil and gas reserves and production capacity more accurately. However, due to technological limitations or sampling costs, the acquisition of 2D slice data in practical applications is often limited. Therefore, the random reconstruction from 2D slices to 3D digital rock becomes particularly important (Zheng and Zhang, 2022).

SD provides a new approach to the reconstruction of 2D to 3D digital rock. Deforum, an SD-based video generation plugin, can generate a consecutive sequence of images and compile them into a video, solely relying on textual descriptions or reference images as its foundation (Hu, 2023). In Deforum, users can describe the visuals of key frames

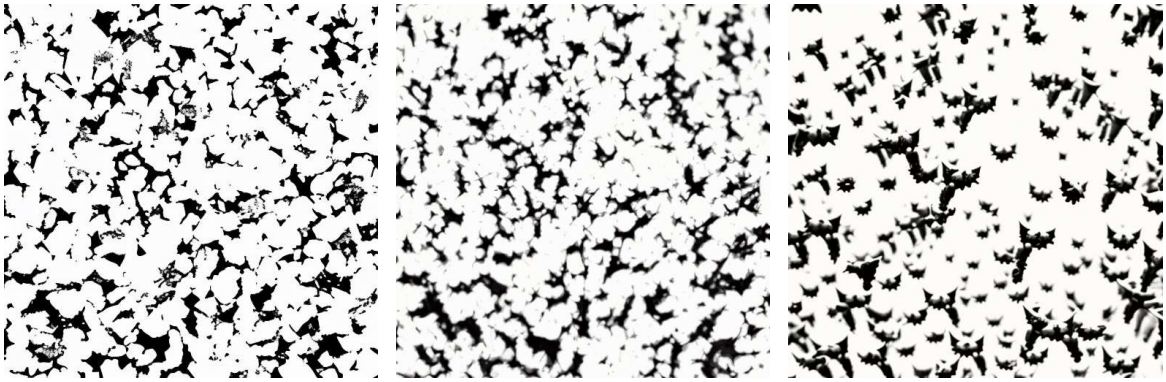

Fig. 8. Generated images of different frames.

by positive and negative prompts. After receiving the input, Deforum utilizes the image-to-image function of SD models to generate a series of image frames that match the input description through iterative and optimization processes. During the generation of each frame, Deforum uses information from previous frames and subtle changes to predict the content of the next frame, which may involve variations in object positions and adjustments to shapes, etc. By continuously generating these slight changes in frames, Deforum can create smooth animation effects.

From the perspective of video generation, 2D images are inputted to produce a continuous sequence of images, enabling the creation of 3D digital core samples. Using sandstone CT images as the initial input, the Realistic Vision model is employed to generate a 90-frame video, mimicking 90 layers in 3D space (Fig. 8).

It is observed that within the first 20 frames of variation, the images generated by Deforum are relatively reasonable. However, beyond 20 frames, the differences between the generated images and the original ones significantly increase, and the generated shapes often become illogical. This phenomenon is related to the limitations of using existing pre-trained models in the digital rock domain. Most of the SD models are typically trained on large-scale, diverse, and everyday image datasets, which may not include or adequately represent the specificity of rock images. Therefore, when attempting to generate content related to digital core, they may fail to accurately capture these subtle but critical features, resulting in significant deviations between the generated images and the expected results. Furthermore, prompts play a crucial role in guiding the models to generate images. However, due to the inherent differences between SD models' training set and rock images, the models may struggle to understand or accurately interpret professional terms and descriptions related to digital rock, further exacerbating the mismatch between the generated results and the expectations. Therefore, to address these issues, it is critical to retrain the SD models for rock images. By leveraging specific datasets to train the models, they are able to acquire a profound understanding of the distinctive features and patterns inherent in rock images, thereby enhancing the precision of the generated outcomes.

In summary, despite the limitations of SD in this area, it is anticipated that retraining the models and optimizing the design of prompts will facilitate the overcoming of these challenges and enhance the quality of the generated images. This will offer enhanced support for the 2D to 3D reconstruction of digital rock.

3.5 Text to image

Text to image is an important application direction of Stable diffusion. In the text to image task, the diffusion models initially take a descriptive text as input and encode it into a fixed-length vector using a pre-trained text encoder to capture the main information. Then, the text vector is combined with random noise to generate conditional embeddings. These embeddings serve as guidance to generate images that match the input text description (Rombach et al., 2022). Throughout the iteration process, diffusion models gradually decode from the conditional embeddings to eventually generate the expected images.

Consideration is given to applying text to image in the field of digital rock, aiming to rapidly generate rock images that meet specified requirements by leveraging given instructions or descriptions. SEM-XRD images containing 12 minerals are annotated. The text includes the proportion of each mineral in the entire image (for example, X takes up 0.00178000 of the entire picture, where X represents the mineral identifier). Afterwards, LoRA is retrained with 2,000 sets of images and labels, with some important parameters listed in Table 5.

The SD model is fine-tuned using LoRA. Subsequently, five SEM-XRD images (not included in the training set) are randomly selected, and their mineral proportions are used as prompts to generate images in SD. The text content includes both positive and negative prompts. Positive prompts are used to guide the depiction of specific minerals and their proportions in the image, such as "A takes up 0 of the entire picture, B takes up 0.00072800 of the entire picture", and so on, until "L takes up 0 of the entire picture". In addition, to avoid generating low-quality or unexpected images, a series of negative prompts are used, including "poorly drawn", "low resolution", "low quality", and "worst quality".

Considering the stochastic nature of text to image conver-

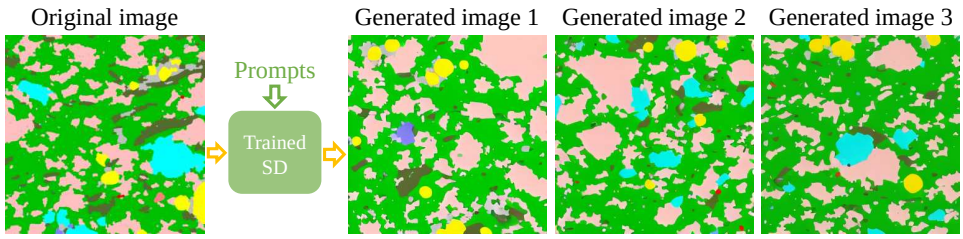

Fig. 9. Generated images based on given prompts.

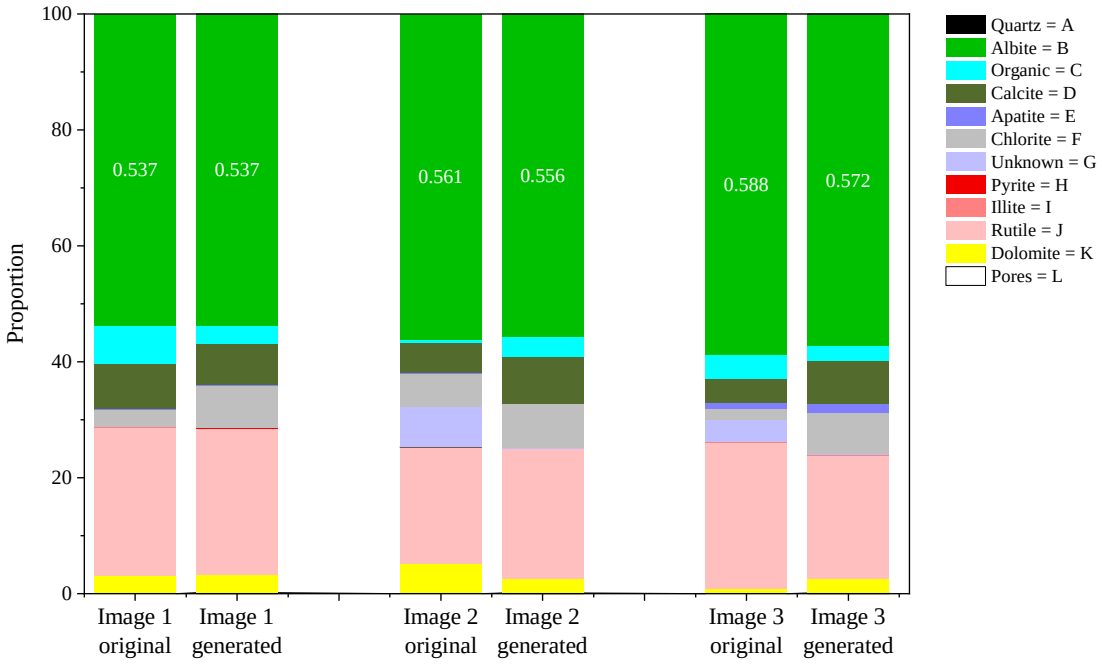

Fig. 10. Proportion of minerals in color for the original and generated images.

Table 5. Parameters of LoRA network for text to image.

Parameters	Value
Image size	320×320
Data amount	2,000
Max epochs	20
Batch size	1
Unet learning rate	1.00×10^{-4}
Text encoder learning rate	1.00×10^{-5}
Prior loss weight	1
Shuffle caption	True
Keep tokens	0
Optimizer type	AdamW8bit
Network dim	32
Network alpha	32
Xformers	True

sion, for each set of prompts, 8 images are generated, and their color proportions are calculated, followed by taking the mean. Some generated results are presented (Fig. 9), and the color proportions of the original and the generated images are displayed in Fig. 10.

The results show that when the proportion of the most dominant color (Albite, which has a mineral number of C) is set as prompts within the range of 0.5-0.6, a high consistency in color proportions with the original images is exhibited by the generated images. However, when adjusting the prompts to around 0.4 or 0.7, differences in color proportions between the generated and original images emerge. This phenomenon may be closely related to the distribution of such colors in the training dataset. Consistency in color proportions between the generated and original images can only be ensured when the color proportion distribution in the testing dataset is adequately covered by the training dataset.

3.6 Data augumentation

In the field of digital rock, diffusion models can serve as a powerful data augmentation technique to improve the generalization capability of deep learning models. By simulating the process of adding and eliminating random noise, they can be used to generate new images that are similar to the original dataset. Ma et al. (2023a) utilized diffusion models to generate a large CT/SEM and binary segmentation

dataset from a small initial dataset. They further test the threshold segmentation performance of U-Net, Attention-U-net, and TransUNet on augmented images, and find that the dataset generated by diffusion models significantly improved the performance, especially in handling images with complex pore morphology. This approach not only expands the training dataset, increasing the diversity and complexity of the data, but also enhances the segmentation accuracy and robustness of the models.

4. Future opportunities in application

4.1 Image completion

When constructing digital rock models, issues such as sample preparation quality and scanning device limitations may lead to missing images. Image completion addresses this by filling missing regions within the image. Repaint, an image completion algorithm based on DDPM, employs a pre-trained unconditional DDPM as a generative prior, which has learned to generate clear images from noise. During the restoration process, it integrates known information into the reverse DDPM process, gradually filling in missing content while maintaining existing areas' integrity (Lugmayr et al., 2022). Compared to other traditional image restoration methods, Repaint doesn't rely on specific mask types or training data distributions, showing superior generalization. It can also generate meaningful content, not just simple texture extrapolations, thanks to DDPM. Nevertheless, the application of image completion algorithm in the restoration of rock images deserves further attention.

4.2 Lithologic discrimination

Lithologic discrimination is a core task in geological research, which involves in-depth analysis and interpretation of physical properties, chemical composition, and microstructure of rocks. In recent years, deep learning algorithms have gained prominence in this field. For instance, CNN-based algorithms efficiently classify rocks by automatically extracting features like texture and color from images (Imamverdiyev and Sukhostat, 2019). In this context, diffusion models bring new possibilities to lithologic discrimination. The diffusion process captures rich contextual information in images and generate more discriminative feature representations, which may help to identify different types of rocks more accurately. Moreover, it exhibits good robustness in handling adversarial attacks, meaning it can better cope with noise and interference in images, thus maintaining accurate identification of complex images. Thus, training the diffusion models holds promise for achieving high-precision identification of complex rock images.

4.3 Fine-tuning and retraining

In this paper, training with LoRA to fine-tune tasks for SD is conducted, while further retraining diffusion models for digital rock also deserves attention. In the training process, possessing a comprehensive digital rock database is crucial. This database should contain a sufficient number of samples and diverse data variations so that the models can learn enough features and patterns. Hence, building and maintaining a high-quality digital rock database is essential for driving the integration of diffusion models with digital rock analysis.

4.4 Similar ideas from medical image analysis

Diffusion models, like those used in digital rock analysis, find application in medicine where high-precision CT scans aid in anatomical observation and parameter analysis. They have demonstrated effectiveness in medical image generation and processing. Similar to digital rock images, the human body also has complex structures. By training diffusion models, intricate human tissues, such as bone microstructures, can be effectively identified, enabling high-precision classification and recognition of complex CT images. Moreover, the availability of high-quality medical image databases facilitates the application of diffusion models in medical imaging. Kazerouni et al. (2023) provide an extensive overview of diffusion models in medical imaging, discussing various applications including reconstruction, image translation, classification, segmentation, denoising, and addressing medical challenges. In other fields utilizing high-precision CT scanning technology, diffusion models can also be used to enhance high-precision image recognition. Further research is warranted in the medical and related fields to explore the full potential of diffusion models for image generation and processing.

5. Potential research directions in methodology

5.1 Samplers

SD provides diverse sampling methods that significantly impact image quality across different tasks. While the DDPM algorithm theoretically lays the foundation for subsequent diffusion models, its practical implementation often lags due to its reliance on the first-order Markov assumption, resulting in slower denoising. In contrast, DDIM, an early sampler for diffusion models, enhances both sampling efficiency and process optimization by utilizing non-Markov diffusion processes. It exhibited enhanced generation performance over DDPM, particularly evident with smaller sampling steps, resulting in notably accelerated generation speeds.

In addition to first-generation samplers like DDIM, SD also supports DPM and DPM++ series samplers (Lu et al., 2022). The DPM++ 2M SDE Karras used in this paper was released in 2022 as a diffusion model sampler. It is currently one of the algorithms that balance generation speed and image quality well, and demonstrating excellent performance in the image processing tasks of this paper. When selecting sampling methods in practice, it is necessary to weigh various sampling methods based on the characteristics of the specific task and dataset, and test multiple sampling methods to find the suitable algorithm for the current task.

5.2 Text to video

OpenAI's video generation model, Sora, released in February 2024, revolutionizes AI "Text-to-Video" generation. Unlike previous tools like Stable Video Diffusion and Runway,

which are slow and produce short videos (3-4 seconds), Sora, utilizing Diffusion Transformer (DiT) technology, can generate high-definition videos up to one minute long. It not only interprets user input but also understands real-world entity interactions, enabling it to create detailed scenes with multiple characters and precise movements (Brooks et al., 2024). However, there's ongoing debate about Sora's understanding of physics principles. OpenAI's technical report highlights instances where Sora fails to accurately simulate physical properties. François Chollet, the founder of Keras, points out limitations in predicting the physical world by fitting high-dimensional curves to a large number of data points using machine learning models. While data-driven models can effectively simulate complex dynamics of the real world under specific conditions, such as weather prediction or wind tunnel experiments, they struggle to generalize to new situations.

Similar to Sora, SD also shares these limitations. Data-driven generative models often struggle to consider mechanisms, and although SD performs well in some tasks, further research is still needed when it comes to precise physical mechanisms.

5.3 Physics-informed approach

For digital rock images, the assessment standards predominantly rely on image quality metrics such as PSNR and SSIM, which may not be ideally suited for applications within the petroleum engineering field. For instance, in the image super-resolution techniques, the PSNR or SSIM is commonly employed as loss functions to evaluate the quality and accuracy of reconstructed images. However, such an index does not account for the underlying physical mechanisms.

In future research, diffusion models could be trained to "learn" the physical mechanisms between different images and "understand" the underlying principles, by embedding the governing equations into the loss of the neural network, such as physics-informed neural networks. This idea could be applied to generating rock images before and after hydraulic fracturing. Specifically, by training the diffusion models using images of intact rocks prior to hydraulic fracturing and images of fractured rocks post-fracture, the model could "simulate" the physical phenomena of hydraulic fracturing and predict the post-fracture digital rock images.

5.4 Light-weighting models

The architecture of diffusion models shows potential for further simplification and optimization. A recent development introduced a highly simplified architecture termed l-DAE through a process of deconstructing diffusion models. This deconstruction revealed that only a few modern components are essential for effective representation learning, while many others are superfluous. By eliminating unnecessary components, diffusion models can become more streamlined and efficient. Throughout this deconstruction process, adjustments were made to the DiT architecture to prioritize self-supervised learning. Subsequently, the tokenizer was gradually simplified, leading the model towards a more traditional DAE approach. The resulting transformed model, l-DAE, has demonstrated competitiveness in self-supervised learning tasks, surpassing benchmarks such as MoCo v3 and ViT-B (86M parameters), albeit still falling short compared to larger-scale models like ViT-L (304M parameters).

5.5 Generation of network parameters

In addition to image generation, diffusion models also have broad prospects for application in other fields. Wang et al. (2024) combined the auto-encoder framework with standard latent diffusion model to design a novel method called parameter diffusion, which synthesizes effective neural network parameters from random noise by training the latent diffusion model. Experiments showed that the method consistently generates models with comparable or higher performance than the trained networks with minimal additional cost on a variety of architectures and datasets. To confirm the generation of new parameters, they conducted further verification revealing significant differences between generated and original models. This suggests parameter diffusion's ability to produce new parameters distinct from its training data. The ability of diffusion models to generate high-performance parameters suggests a breakthrough in algorithm design and optimization, indicating their enormous potential for application.

5.6 Combination with generative foundation models

Combining diffusion models with generative foundation models may explore more interesting applications in AIGC. For example, providing textual inputs to ChatGPT to generate accurate and reasonable prompts, which are then passed to the diffusion models to generate images that match the descriptions. Alternatively, batches of rock images can be fed into diffusion models for feature extraction, followed by combining the extracted features with generative foundation models to analyze and produce detailed textual descriptions of reservoir conditions.

6. Conclusions

In this study, the applications of SD are explored in image super-resolution reconstruction, image extension, denoising and restoration, random reconstruction of 3D digital rocks based on 2D images, as well as generating rock images given specific instructions. SD is characterized by its ease of training and adjustment, high stability and strong scalability. Although SD performs well in some tasks, challenges remain in areas such as 2D to 3D reconstruction and text to image generation due to the limited generalization ability of universal models in the field of digital rock. Therefore, it is necessary to establish a comprehensive and reliable rock database to support fine-tuning or retraining. The powerful image generation capability of diffusion models provides new ideas for digital rock analysis, showing potential in filling missing parts of rock images, lithologic discrimination, and even generating network parameters, further developing algorithms. Many researchers have been inspired by the principles of diffusion models and are committed to continuously light-weighting and accelerating SD algorithms. The application of SD in the field of digital

core has shown vast research prospects, and it is expected to bring revolutionary progress to digital core analysis.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 52374017), and Science Foundation of China University of Petroleum, Beijing (No. 2462022QNXZ002).

Conflict of interest

The authors declare no competing interest.

References

Ahuja, V. R., Gupta, U., Rapole, S. R., et al. Siamese-SR: A siamese super-resolution model for boosting resolution of digital rock images for improved petrophysical property estimation. IEEE Transactions on Image Processing, 2022, 31: 3479-3493.

Bar-Tal, O., Yariv, L., Lipman, Y., et al. Multidiffusion: Fusing diffusion paths for controlled image generation. Paper Presented at Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA, 23-29 July, 2023.

Brooks, T., Peebles, B., Homes, C., et al. Video generation models as world simulators. OpenAI, 2024.

Chang, C., Peng, J., Safari, M., et al. High-resolution MRI synthesis using a data-driven framework with denoising diffusion probabilistic modeling. Physics in Medicine & Biology, 2024, 69(4): 045001.

Chen, H., He, X., Teng, Q., et al. Super-resolution of real-world rock microcomputed tomography images using cycle-consistent generative adversarial networks. Physical Review E, 2020, 101(2): 023305.

Chi, P., Sun, J., Luo, X., et al. Reconstruction of 3D digital rocks with controllable porosity using CVAE-GAN. Geoenergy Science and Engineering, 2023, 230: 212264.

Croitoru, F., Hondru, V., Ionescu, R., et al. Diffusion models in vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(9): 10850-10869.

Feng, J., Teng, Q., Li, B., et al. An end-to-end three-dimensional reconstruction framework of porous media from a single two-dimensional image based on deep learning. Computer Methods in Applied Mechanics and Engineering, 2020, 368: 113043.

Gong, K., Johnson, K., El Fakhri, G., et al. PET image denoising based on denoising diffusion probabilistic model. European Journal of Nuclear Medicine and Molecular Imaging, 2023, 51: 358-368.

Goral, J., Panja, P., Deo, M., et al. Confinement effect on porosity and permeability of shales. Scientific Reports, 2020, 10(1): 49.

Ho, J., Jain, A., Abbeel, P. Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 2020, 33: 6840-6851.

Hu, E. Deforum stable diffusion local version, 2023.

Imamverdiyev, Y., Sukhostat, L. Lithological facies classification using deep convolutional neural network. Journal of Petroleum Science and Engineering, 2019, 174: 216-228.

Kazerouni, A., Aghdam, E. K., Heidari, M., et al. Diffusion models in medical imaging: A comprehensive survey. Medical Image Analysis, 2023, 88: 102846.

Li, X., Li, B., Liu, F., et al. Advances in the application of deep learning methods to digital rock technology. Advances in Geo-Energy Research, 2023, 8(1): 5-18.

Li, Z., Deng, S., Hong, Y., et al. A novel hybrid CNN-SVM method for lithology identification in shale reservoirs based on logging measurements. Journal of Applied Geophysics, 2024, 223: 105346.

Liao, Q., Xue, L., Wang, B., Lei, G. A new upscaling method for microscopic fluid flow based on digital rocks. Advances in Geo-Energy Research, 2022, 6(4): 357-358.

Liu, X., Su, S., Gu, W., et al. Super-resolution reconstruction of ct images based on multi-scale information fused generative adversarial networks. Annals of Biomedical Engineering, 2024, 52(1): 57-70.

Lu, C., Zhou, Y., Bao, F., et al. DPM-solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. Advances in Neural Information Processing Systems, 2022, 35: 5775-5787.

Lugmayr, A., Danelljan, M., Romero, A., et al. Repaint: Inpainting using denoising diffusion probabilistic models. Paper Presented at Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, Los Angeles, 18-24 June, 2022.

Ma, Z., He, X., Kwak, H., et al. Enhancing rock image segmentation in digital rock physics: A fusion of generative ai and state-of-the-art neural networks. ArXiv Preprint ArXiv: 2311.06079, 2023a.

Ma, Z., Sun, S., Yan, B., et al. Enhancing the resolution of micro-CT images of rock samples via unsupervised machine learning based on a diffusion model. Paper SPE 214883 Presented at SPE Annual Technical Conference and Exhibition, San Antonio, Texas, 16-18 October, 2023b.

Mao, S., He, Y., Chen, H., et al. High-quality and high-diversity conditionally generative ghost imaging based on denoising diffusion probabilistic model. Optics Express, 2023, 31(15): 25104-25116.

Pan, S., Wang, T., Qiu, R. L., et al. 2D medical image synthesis using transformer-based denoising diffusion probabilistic model. Physics in Medicine & Biology, 2023, 68(10): 105004.

Rombach, R., Blattmann, A., Lorenz, D., et al. High-resolution image synthesis with latent diffusion models. Paper Presented at Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, Los Angeles, 18-24 June, 2022.

Shan, L., Bai, X., Liu, C., et al. Super-resolution reconstruction of digital rock CT images based on residual attention mechanism. Advances in Geo-Energy Research, 2022, 6(2): 157-168.

Wang, K., Xu, Z., Zhou, Y., et al. Neural network diffusion.

Wang, Y. D., Armstrong, R. T., Mostaghimi, P. Enhancing resolution of digital rock images with super resolution convolutional neural networks. Journal of Petroleum Science and Engineering, 2019, 182: 106261.

Xie, S., Zhang, Z., Lin, Z., et al. Smartbrush: Text and shape guided object inpainting with diffusion model. Paper Presented at Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, 18-22 June, 2023.

Yan, Z., Zhou, C., Li, X. A survey on generative diffusion model. Computer Science, 2024, 51(1): 273-283. (in Chinese)

Zeng, B., Liu, X., Gao, S., et al. Face animation with an attribute-guided diffusion model. Paper Presented at Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, 18-22 June, 2023.

Zha, W., Li, X., Li, D., et al. Shale digital core image generation based on generative adversarial networks. Journal of Energy Resources Technology, 2021, 143(3): 033003.

Zhang, K., Li, Y., Liang, J., et al. Practical blind image denoising via Swin-Conv-UNet and data synthesis. Machine Intelligence Research, 2023a, 20(6): 822-836.

Zhang, L., Rao, A., Agrawala, M. Adding conditional control to text-to-image diffusion models. Paper Presented at Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1-6 October, 2023b.

Zhao, J., Chen, H., Li, N., et al. Research advance of petrophysical application based on digital core technology. Progress in Geophysics, 2020, 35(3): 1099-1108.

Zhao, J., Wang, F., Cai, J. 3D tight sandstone digital rock reconstruction with deep learning. Journal of Petroleum Science and Engineering, 2021, 207: 109020.

Zheng, Q., Zhang, D. RockGPT: Reconstructing three-dimensional digital rocks from single two-dimensional slice with deep learning. Computational Geosciences, 2022, 26(3): 677-696.