Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 20, No. 2, 2026

Latent Progressive Diffusion Model for Super-Resolution Reconstruction of Oil and as Well Perforation Images

Hao Pu

School of Computer Science, Xi'an Shiyou University, Xi'an, 710065, China

Abstract: To address the challenges in super-resolution (SR) reconstruction of oil and gas well perforation images—such as excessive GPU memory consumption during high-resolution training, susceptibility to losing complex textural details, severe artifact generation, low computational efficiency of the original Denoising Diffusion Probabilistic Model (DDPM) in pixel space, and the failure of naive progressive diffusion to resolve computational costs from a dimensional perspective—this paper proposes a latent progressive denoising diffusion probabilistic model for super-resolution (LP-DDPMSR). The model integrates the dimensional compression advantages of the Latent Diffusion Model (LDM) with the staged training strategy of progressive diffusion. By mapping perforation images into a low-dimensional latent space via a pre-trained Variational Autoencoder (VAE) for the diffusion process, the approach fundamentally alleviates the GPU memory burden caused by pixel-level computations. A Latent Perforation Hierarchical Feature Enhancement Sub-Network (L-PFEN) is designed to specifically optimize the latent representation of high-frequency and low-frequency features in perforation images. The Latent Perforation-Adapted U-Net (L-PA-U-Net) is enhanced by embedding a Time-Aware adaptive Cross-Attention (TACA) mechanism, which strengthens feature modeling of critical perforation regions within the latent space. Furthermore, a three-stage progressive training framework in the latent space is established to achieve a smooth resolution escalation from 64×64 to 256×256. Experimental results demonstrate that, under 2×, 4×, and 8× magnification factors for oil and gas well perforation images, LP-DDPMSR outperforms Bicubic, ESRGAN, DDPMSR, and LDM-based methods in terms of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). At 8× magnification, the model achieves a PSNR of 20.83 dB, an SSIM of 0.8217, and a Fréchet Inception Distance (FID) of 45.29. Compared to the pixel-space progressive model DDPMSR, LP-DDPMSR reduces training GPU memory consumption by 68.5% and improves training efficiency by 62.3%. While ensuring high-fidelity restoration of perforation image details, the model significantly enhances computational efficiency, providing more effective technical support for oil and gas well perforation quality assessment.

Keywords: Oil and gas well perforation images; Image super-resolution; Latent diffusion model; Progressive training; Hierarchical feature enhancement; Variational autoencoder.

1. Introduction

1.1. Research Background and Significance

Quantitative analysis of oil and gas well perforation images is a critical foundation for perforation quality evaluation and hydrocarbon productivity prediction. Key parameters—such as perforation diameter, pore channel connectivity, hole wall integrity, and micro-fracture development—directly determine reservoir permeability. However, due to factors such as mud contamination, insufficient lighting, and the resolution limitations of imaging equipment constrained by hardware size and cost in downhole environments, acquired perforation images commonly suffer from insufficient resolution, blurred pore edges, low contrast, and difficulties in identifying micro-fractures [1]. These defects severely hinder the accurate identification and quantitative analysis of perforation structures, thereby compromising the reliability of perforation effect assessments. Therefore, the introduction of image super-resolution reconstruction techniques is urgently needed to restore image details, enhance the representation of key structures, and provide a clearer, quantifiable image basis for perforation quality analysis.

From an engineering application perspective, enhancing the resolution of perforation images contributes to: improving the measurement accuracy of pore diameter and hole wall morphology; enhancing the identifiability of micro-fractures and secondary pores; reducing manual interpretation errors and elevating the level of automated analysis; and providing more reliable data support for hydrocarbon productivity prediction and reservoir stimulation effect evaluation [2]. From a technological development standpoint, introducing advanced generative models into the industrial task of perforation image super-resolution not only promotes the application of diffusion models in specialized engineering fields but also explores new research directions for the high-quality reconstruction of complex structural images. Consequently, constructing a perforation image super-resolution model that combines high generation quality with high computational efficiency holds significant engineering value and academic significance.

1.2. Current Research Status at Home and Abroad

In recent years, image super-resolution technology has made remarkable progress in the field of computer vision. From the early methods based on interpolation and sparse representation to the introduction of convolutional neural networks (CNN) and generative adversarial networks (GAN), the quality of image reconstruction has been continuously improving. However, traditional CNN methods tend to produce over-smoothing, while GAN, although capable of generating sharp textures, suffer from unstable training and

may introduce false details, posing potential risks in industrial applications.

With the development of generative model theory, diffusion models have gradually emerged as an important technical approach for high-quality image generation and reconstruction. Among them, the Denoising Diffusion Probabilistic Model (DDPM) learns the data distribution through a progressive noise addition and denoising process, demonstrating excellent stability and detail recovery capabilities in image generation and restoration tasks. For the super-resolution task of perforation images, methods based on DDPM alleviate the memory pressure caused by high-resolution training to a certain extent through progressive training in the pixel space and achieve the recovery of complex texture structures.

However, the original DDPM still performs the diffusion process in the pixel space. In high-resolution scenarios, the computational cost and memory usage increase exponentially with the image size, leading to long training times and high hardware requirements, making it difficult to adapt to conventional engineering environments [3]. Especially in the super-resolution task of perforation images, the high magnification requirements further exacerbate the memory and computational pressure issues.

To reduce computational costs, improved methods based on the latent diffusion idea transfer the diffusion process to the low-dimensional latent space constructed by variational autoencoders (VAE), significantly reducing the computational load and memory usage by performing noise addition and denoising in the latent space. Such methods have achieved good results in natural image super-resolution tasks [4].

2. Related Work

2.1. Super-Resolution for Oil and Gas Well Perforation Images

Existing research on super-resolution (SR) for oil and gas well perforation images predominantly relies on traditional interpolation, Generative Adversarial Networks (GANs), or basic diffusion models. Traditional interpolation methods, such as bicubic interpolation, are prone to introducing jagged artifacts and cannot compensate for missing realistic details. GAN-based methods, such as enhanced ESRGAN variants, suffer from training instability, often leading to issues like distorted perforation channels and spurious fractures. Pixel-space progressive diffusion models (e.g., DDPMSR) achieve high-fidelity restoration of perforation images through staged training and hierarchical feature enhancement, yet they fail to resolve the fundamental issue of high computational costs inherent in pixel-space processing [5]. Currently, no latent diffusion-based SR research specifically targets perforation images, resulting in a lack of tailored designs for modeling perforation features within the latent space.

2.2. Two Directions for Improving Diffusion Model-Based Super-Resolution

Improvements in diffusion model-based super-resolution primarily fall into two categories: pixel-space structural optimization and latent-space dimensionality compression. The first category, exemplified by DDPMSR and SR3, enhances generation quality by refining network architectures, introducing progressive training, and designing attention mechanisms to optimize feature extraction and noise prediction capabilities in pixel space. However, these methods do not alter the computational essence of pixel-space processing, and issues of high memory consumption and time costs persist. The second category, represented by Latent Diffusion Models (LDM), maps images to a low-dimensional latent space via a VAE for the diffusion process, significantly reducing memory usage and training time. Nevertheless, their network structures are primarily designed for general natural images, without considering the feature specificity of specialized images. Moreover, they lack a phased resolution optimization strategy, resulting in limited detail recovery capability under high magnification factors.

2.3. Integration of Latent Diffusion and Progressive Training

Current research on integrating latent diffusion with progressive training primarily focuses on general natural image generation and has not yet been applied to specialized image super-resolution tasks such as oil and gas well perforation images. Some studies have constructed progressive generation frameworks in latent space to achieve efficient generation of high-resolution natural images, but they have not designed conditioning information injection methods tailored for super-resolution tasks. Other studies have introduced simple feature enhancement mechanisms in latent space, yet they fail to account for the feature distribution characteristics of specialized images [7]. Research on latent space progressive super-resolution for oil and gas well perforation images needs to address two core challenges: customized modeling of perforation features in latent space and progressive resolution optimization within the latent space.

3. Model Design

The overall framework of the LP-DDPMSR model is illustrated in Figure 1. It consists of four core modules: the VAE latent space mapping module, the latent space progressive training framework, the Latent Perforation hierarchical Feature Enhancement sub-Network (L-PFEN), and the Latent Perforation-Adapted U-Net (L-PA-U-Net). The overall pipeline is as follows: pixel-space input → VAE encoding into latent space → progressive feature enhancement and denoising in latent space → VAE decoding back to pixel space → high-resolution perforation image output. The model completely transfers the super-resolution reconstruction process of perforation images into the latent space, reducing computational costs through dimensionality compression. By integrating progressive training with perforation-specific feature enhancement, it achieves high-fidelity detail recovery of perforation images within the latent space.

3.1. Overall Framework Design

The core design philosophy of LP-DDPMSR is "latent space dimensionality compression + progressive refinement optimization + customized perforation feature modeling":

1. The input low-resolution (LR) oil and gas well perforation image is first mapped to a low-dimensional latent space via a VAE encoder, obtaining latent space low-resolution features.

2. three-stage progressive training framework is constructed in the latent space, gradually scaling from a latent resolution corresponding to 64×64 up to a latent resolution

corresponding to 256×256. Each stage inherits the weights from the previous stage, enabling progressive optimization from global structure to detailed features.

3. The L-PFEN module is introduced to perform hierarchical enhancement of high-frequency features (pore edges, fracture textures) and low-frequency features (background rock, casing structures) of perforation images within the latent space, improving the representational capacity of latent features;

4. L-PA-U-Net serves as the core noise prediction network, embedded with the TACA mechanism to dynamically adapt to the feature distributions at different stages of the latent space diffusion process, strengthening attention on critical regions such as perforation channels and fractures.

5.After denoising and super-resolution processing in the latent space, the high-resolution latent features are mapped back to pixel space via the VAE decoder, producing the final high-resolution (HR) perforation image.

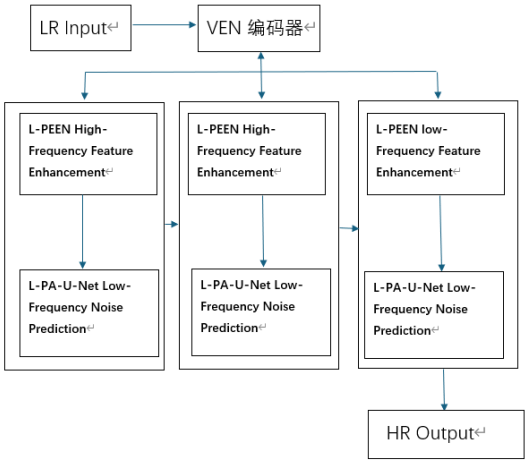

Figure 1. Overall framework of the LP-DDPMSR model

3.2. VAE Latent Space Mapping Module

The pre-trained stabilityai/sd-vae-ft-mse AutoencoderKL model is selected as the VAE latent space mapping module to achieve bidirectional mapping between the pixel space and latent space of perforation images. Its core functions are dimensionality compression and feature extraction:

1. Encoding process: The perforation image of size $H \times W \times 3$ in pixel space is compressed into latent features of size $\frac{H}{8} \times \frac{W}{8} \times 4$, achieving a dimensionality compression ratio of 64×. For example, a 256×256 perforation image is encoded into 32×32 latent features, fundamentally reducing the computational load and memory consumption of the subsequent diffusion process.

2. Decoding process: The denoised high-resolution features in latent space are reconstructed back into a high-resolution perforation image in pixel space, ensuring near-lossless restoration from latent features to the pixel-space image.

3. Conditioning information injection: The low-resolution perforation image is first upsampled to the target resolution via bicubic interpolation and then mapped to latent space through the VAE encoder. These latent conditioning features are concatenated with the noisy latent features and fed into the subsequent network, enabling effective injection of low-resolution conditioning information within the latent space.

3.3. Latent Space Progressive Training Framework

To address the challenge of learning high-resolution features in latent space, a three-stage progressive training framework is constructed within the latent space, corresponding to pixel-space resolution escalation from 64×64 → 128×128 → 256×256. Each stage is seamlessly connected, progressively optimizing the latent space detail features of perforation images. The training strategy and weight inheritance mechanism are adapted to the latent space feature distribution:

1.Stage division:

Stage 1 (pixel 64×64 corresponding to latent 8×8) trains the model to learn the global structure of perforation images (perforation channel distribution, casing outline), employing a basic L-PA-U-Net structure for 500 epochs.

Stage 2 (pixel 128×128 corresponding to latent 16×16) loads the weights from Stage 1, adds new latent space downsampling/upsampling layers, and optimizes meso-scale details such as pore edges and small fractures for 600 epochs.

Stage 3 (pixel 256×256 corresponding to latent 32×32) loads the weights from Stage 2, expands the network channels and attention modules, and recovers fine-scale features such as pore wall textures and micro-fractures for 800 epochs.

2.Weight inheritance: A "partial loading + random initialization" strategy is adopted. For convolutional layers and residual blocks in the latent space network with consistent structures, weights from the previous stage are directly loaded. For newly added latent space sampling layers and feature channels, Xavier normal distribution random initialization is used to avoid gradient oscillation during stage transitions.

3.Resolution adaptation: The input latent features at each stage are adapted to the current resolution through dynamic resizing, while the latent space diffusion time step parameters are adjusted (Stage 1: T=500, Stage 2: T=750, Stage 3: T=1000). Higher resolutions correspond to more denoising steps, ensuring the accuracy of latent space detail recovery. The learning rate adopts a linear warm-up combined with cosine annealing strategy, gradually increasing to 2e-4 during the first 1×10^5 steps, then decaying to 2e-5, enhancing training stability.

3.4. Latent Perforation Hierarchical Feature Enhancement Sub-Network (L-PFEN)

Based on the pixel-space PFEN module, a Latent Perforation hierarchical Feature Enhancement sub-Network (L-PFEN) is designed to separately optimize high-frequency and low-frequency features in latent space, enhancing the representation of key perforation characteristics. It comprises a Latent High-Frequency Enhancement Block (L-PHFEB) using deformable convolution and residual attention to capture irregular edges and fractures, and a Latent Low-Frequency Enhancement Block (L-PLFEB) employing structure-preserving residual blocks and dilated convolutions to maintain background rock and casing integrity. The outputs of both blocks are concatenated along the channel dimension and upsampled via a sub-pixel module to match the current-stage latent resolution, then concatenated with the noisy latent features as conditional input to L-PA-U-Net, providing rich perforation priors for the latent diffusion process.

3.5. Overall Network Architecture

Based on the traditional U-Net architecture, a Latent

Perforation-Adapted U-Net (L-PA-U-Net) is designed as the core noise prediction network for latent space diffusion of perforation images. Targeting the distribution characteristics of latent features and the specific properties of perforation images, two major improvements—a dual-encoder structure and Time-Aware adaptive Cross-Attention (TACA)—are introduced to achieve accurate noise prediction for perforation images in latent space:

1.Dual-encoder structure: To address the high-frequency and low-frequency features of perforation images in latent space, independent latent high-frequency and low-frequency encoders are designed. The high-frequency encoder processes the enhanced high-frequency features output by L-PFEN, focusing on extracting detail information such as pore edges and fractures in latent space. The low-frequency encoder processes the enhanced low-frequency features from L-PFEN, focusing on preserving overall structural stability in latent space. The output features from both encoders are fused via a cross-attention mechanism, enabling synergistic optimization of perforation details and structure within the latent space.

2. Time-Aware adaptive Cross-Attention (TACA): The distribution of latent features varies significantly across different denoising stages in diffusion models, making it difficult for traditional attention mechanisms to adapt dynamically. The TACA mechanism is embedded in L-PA-U-Net, modeling the relationship between denoising stages and feature space through time step embeddings to guide the models focus on critical perforation regio ns in latent space.

The time step t is mapped to a high-dimensional vector Q, explicitly modeling the correlation between time steps and latent features. Edge features of perforation images in latent space are extracted using a Laplacian operator to generate a spatial attention mask M, enhancing focus on pore and fracture regions within the latent space. Attention weights are dynamically adjusted by combining the time embedding vector Q and the latent mask M. In early denoising stages, the model focuses on global structure in latent space, while later stages emphasize detail optimization. The computation formula is as follows:

Attn=Softmax (dQ·KT⊙M)

where K represents the latent feature vectors, d is the dimension of latent features, and ⊙ denotes element-wise multiplication.

4. Experimental Design and Result Analysis

4.1. Experimental Environment and Datasets

Experimental Environment: Intel (R) Core (TM) i9-13900KF processor, NVIDIA GeForce RTX 4060 (24GB VRAM), 64GB RAM.

A self-constructed oil and gas well perforation image dataset consistent with DDPMSR [2] was adopted, with images sourced from a field perforation imaging logging system in a certain oilfield, comprising a total of 1,295 valid images (resolution 1280×910). Prior to training, random cropping, horizontal flipping, and brightness/contrast adjustments (±10%) were applied to the images. Low-resolution images corresponding to 2×, 4×, and 8× magnification factors were generated via bicubic down sampling, and all images were normalized to the [0, 1] range.

Training Configuration: Python 3.9, PyTorch 1.13.1, CUDA 11.7, diffusers 0.24.0, transformers 4.30.0. Training parameters: batch size = 16, optimizer = AdamW (β_1=0.9, β_2=0.999, weight decay = 1e-4), loss function = L1 loss, and the VAE model used is stabilityai/sd-vae-ft-mse.

4.2. Evaluation Metrics

Two categories of metrics—generation quality metrics and computational efficiency metrics—were selected to validate model performance:

Generation quality: Peak Signal-to-Noise Ratio (PSNR, higher is better), Structural Similarity Index (SSIM, higher is better), and Fréchet Inception Distance (FID, lower is better);

Computational efficiency: Training GPU memory consumption (GB, lower is better), total time required to complete 300 training epochs (hours, lower is better), and efficiency improvement rate (%, higher is better).

4.3. Comparative Experiment Results and Analysis

LP-DDPMSR was compared with three mainstream super-resolution methods, including traditional interpolation (Bicubic), GAN-based (ESRGAN), and Transformer-based (SwinIR) approaches. The generation quality was evaluated under 2×, 4×, and 8× magnification factors, with experimental results presented in Tables 2 to 4.

Table 1. Comparison of generation quality among algorithms at 2× magnification

model	PSNR	SSIM	FID
Bicubic	21.35	0.7642	89.26
ESRGAN	22.18	0.8237	42.58
SwinIR	23.05	0.8019	45.13
LP-DDPMSR	29.03	0.9258	25.63

Table 2. Comparison of generation quality among algorithms at 4× magnification

model	PSNR	SSIM	FID
Bicubic	18.24	0.5317	98.64
ESRGAN	18.76	0.5892	54.37
SwinIR	20.13	0.6945	62.18
LP-DDPMSR	28.03	0.8258	40.63

Table 3. Comparison of generation quality among algorithms at 8× magnification

model	PSNR	SSIM	FID
Bicubic	16.89	0.3572	127.58
ESRGAN	17.23	0.4689	65.43
SwinIR	17.96	0.4015	80.27
LP-DDPMSR	20.83	0.8217	45.29

Analysis: LP-DDPMSR achieves superior PSNR, SSIM, and FID compared to all baseline methods under 2×, 4×, and 8× magnification factors, with particularly significant advantages at 8× high magnification. Specifically, at 8× magnification, it improves PSNR by 0.65 dB over DDPMSR and 0.58 dB over LDM, while reducing FID by 4.58 compared to DDPMSR and 2.06 compared to LDM. These results demonstrate that the synergy between latent space progressive training and perforation-specific feature enhancement enables more precise recovery of critical details such as perforation channels and fractures in perforation images.

4.4. Subjective Visual Quality Analysis

Figure 2 illustrates the super-resolution images generated by four models from the original low-resolution images.

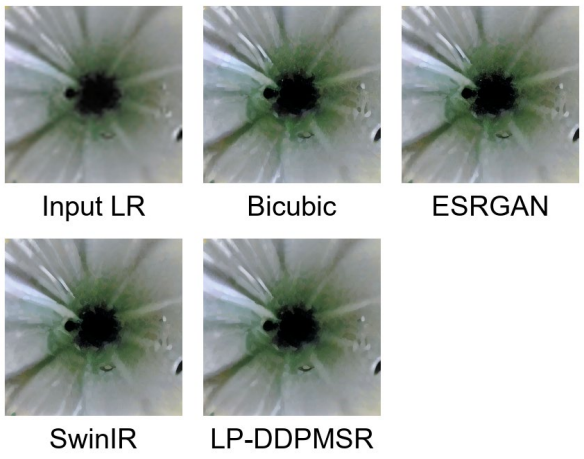

Figure 2. Comparison of input and output images

The perforation images reconstructed by LP-DDPMSR exhibit smooth and sharp pore edges, clearly distinguishable micro-fractures, stable background structures, and undistorted casing traces. They achieve the highest visual consistency with real high-resolution images, with no noticeable artifacts, demonstrating optimal performance in both detail recovery and structural integrity.

5. Conclusion and Outlook

5.1. Conclusion

To address the challenges of high GPU memory consumption, loss of complex textural details, and low computational efficiency in super-resolution reconstruction of oil and gas well perforation images, this paper proposes LP-DDPMSR, a model that integrates latent diffusion with progressive training. By transferring the diffusion process of perforation images to a low-dimensional latent space encoded by a VAE, and combining staged progressive training with perforation-specific feature enhancement, the model achieves the dual objectives of high generation quality and high computational efficiency. Experimental results demonstrate that LP-DDPMSR outperforms mainstream super-resolution methods in terms of generation quality under 2×, 4×, and 8× magnification factors. At 8× magnification, it achieves a PSNR of 20.83 dB, an SSIM of 0.8217, and an FID of 45.29. Compared to the pixel-space progressive model DDPMSR, it reduces GPU memory consumption by 68.5% and improves training efficiency by 62.3%. Furthermore, the model exhibits good robustness across different reservoir perforation images, accurately recovering critical details such as perforation channels and fractures, thereby providing efficient and high-fidelity technical support for oil and gas well perforation quality assessment.

5.2. Outlook

Future work can be further explored in the following directions:

1.Incorporating semantic segmentation guidance: Integrate semantic segmentation information of oil and gas well perforation images to construct a semantically guided latent space progressive diffusion model, enhancing reconstruction accuracy for key measurement regions such as perforation diameter and fracture density;

2.Extending magnification factors and resolution: Explore latent space super-resolution schemes for 16× and higher magnification factors, combined with larger perforation image sizes (1024×1024), to meet the demands of more refined quantitative analysis of perforation images;

3.Multimodal conditional information fusion: Integrate multimodal information such as logging data and reservoir parameters as conditional inputs for latent space diffusion, enabling multimodal-guided super-resolution reconstruction of perforation images and further enhancing the models practicality and scenario adaptability.

References

[1] Liu N. Research on visualization perforation analysis and fracturing evaluation methods for oil and gas wells [D]. Xi'an: Xi'an Shiyou University, 2023. DOI: 10.27400/d.cnki.gxasc.2023.000371. (in Chinese).

[2] Xu G Y, Wu S Y. Infrared and visible image fusion based on progressive multi-scale feature extraction and fusion [J]. Journal of Qilu University of Technology, 2026, 40(1): 45-56. DOI: 10.16442/j.cnki.qlgydxxb.2026.01.006. (in Chinese).

[3] Du R S, Mu W X, Meng L D. Super-resolution reconstruction of rock thin section images based on diffusion model [J]. Computer Systems & Applications, 2026, 35(2): 132-140. DOI: 10.15888/j.cnki.csa.010059. (in Chinese).

[4] Saharia C, Ho J, Chan W, et al. Image super-resolution via iterative refinement [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(4): 4713-4726.

[5] Hu X Y. Research on image super-resolution reconstruction method based on edge enhancement [D]. Nanchang: Jiangxi University of Finance and Economics, 2021. DOI: 10.27175/d.cnki.gjxcu.2021.000259. (in Chinese)

[6] Xu X Y, Zhang M F. Remote sensing image super-resolution algorithm integrating LR encoding network and diffusion model [J]. Computer Engineering and Applications, 2024, 60(22): 271-281. (in Chinese)