Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

CADDPM: Contour-Aware Denoising Diffusion Probabilistic Model for Mural Inpainting

qi wu

bs231011008@sust.edu.cn

Shaanxi University of Science and Technology

Qinchuan Zhan

Shaanxi University of Science and Technology

Yuhua Li

Shaanxi University of Science and Technology

Guolong Du

Shaanxi University of Science and Technology

Ruichen Lu

Shaanxi University of Science and Technology

Article

Keywords: Mural inpainting, Diffusion model, Contour-aware, Multi-scale features

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11321922/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

CADDPM: Contour-Aware Denoising Diffusion Probabilistic Model for Mural Inpainting

Qi Wu[1], Qinchuan Zhan[1], Yuhua Li[1], Guolong Du[1], Ruichen Lu[1]

Abstract: Murals are vital cultural heritage, but aging, environmental erosion, and historical damage cause structural fragmentation and color degradation, while inpainting faces structural discontinuity and stylistic inconsistency. To address these, we propose a contour-aware denoising diffusion probabilistic model (CADDPM) for non-contact digital mural restoration. Sobel operator extracts contours from masks, and a lightweight convolutional encoder derives contour embeddings injected into denoising time-step embeddings, providing shape and scale priors for missing regions. Atrous spatial pyramid pooling at U-Net bottleneck enhances multi-scale context modeling. A multi-task loss integrates pixel, perceptual, style, and edge-alignment constraints to ensure structural accuracy, semantic consistency, and color-texture harmony. Inference uses a progressive sampling constraint with free generation, soft blending, and hard constraint stages to mitigate boundary artifacts. Experiments on MuralDH show CADDPM outperforms DDPM, RePaint, and LaMa in PSNR, SSIM, and LPIPS. This framework offers a controllable, interpretable path for digital mural restoration, supporting documentation, analysis, and long-term preservation.

Keywords: Mural inpainting; Diffusion model; Contour-aware; Multi-scale features

Introduction

Ancient murals, as carriers of historical memory and cultural heritage, are vital to heritage conservation. Murals at the Mogao Grottoes and other sites span millennia, conveying rich information on religious evolution, social customs, and artistic expression. However, natural erosion, environmental deterioration, and human interference have caused paint loss, network of cracks, flaking, salt efflorescence, and fading, threatening their artistic integrity and long-term preservation[1]. Surveys show over 4,000 m^2 of murals and over 250 caves at Mogao are affected[2]. Traditional physical restoration is limited by irreversibility and material aging, limiting large-scale physical restoration and

[1]School of Art & Design, Shaanxi University of Science & Technology University, Xian, China.
Corresponding author: Qi Wu, E-mail:bs231011008@sust.edu.cn

digital archiving. With improved digital acquisition, digital preservation and virtual restoration have become important directions[3]. Deep learning offers new approaches, but generic inpainting struggles with large irregular damage and complex textures.

Image inpainting technology has undergone a complete evolution from traditional methods to deep learning. Early research centered on mathematical modeling and statistical learning. Buades et al.[4] introduced the non-local means denoising algorithm, laying the foundation for self-similarity-based inpainting; Criminisi et al.[5] proposed an exemplar-based sample block filling algorithm, which became a classic representative of the traditional inpainting paradigm. With the advent of deep learning, Pathak et al.[6] developed the context encoder, pioneered the direction of combining self-supervised feature learning with inpainting; Yu et al.[7] introduced the contextual attention mechanism, and Liu et al.[8] proposed partial convolution operations, making breakthroughs in inpainting quality and adaptability to irregular holes, respectively; Zeng et al.[9] designed a pyramid context encoder network to optimize multi-scale fusion; Transformer architectures further strengthened the ability to model long-range dependencies[10]. In recent years, diffusion models have become a research hotspot. RePaint[11] and latent diffusion models[12] have achieved state-of-the-art results in image completion and high-resolution generation. For the restoration of cultural heritage, researchers have also developed dedicated solutions[13-14].

In mural restoration, diffusion models have been widely adopted due to their multi-step iterative denoising and strong generative capability. For example, Brownian bridge diffusion enhances textures via line-drawing structure extraction and dual-prior guidance [15]; dual-branch diffusion reconstructs grayscale structure and uses cross-attention to guide texture synthesis[16]; ST-DiffNet improves semantic coherence through structural and textural branches[17]; DiffuMural achieves high-fidelity restoration via multi-scale convergence and cooperative guidance[18]. Li et al.[19] improved the U-Net noise estimation network under DDPM by introducing hybrid attention and atrous spatial pyramid pooling (ASPP). Li et al.[20] further proposed DMEA-Net with enhanced multi-head contextual attention, strengthening local texture extraction and channel interaction. Zhang et al.[21] proposed edge-constrained diffusion and adversarial optimization, using TERS to reconstruct grayscale sketches as explicit structural priors, injecting edge information into

SD-inpainting via ControlNet, and a non-local color-consistency adversarial network for global color coordination. Peng et al.[22] proposed DCADif, a decoupled conditional adaptive time-dynamic fusion framework, dynamically adjusting structural condition weights across denoising stages. APLDiff[23] balances inference efficiency and visual realism via physical degradation simulation and two-stage multi-scale sampling. MAPGR[24] reformulates inpainting as multi-stage reasoning constrained by diagnostic evidence. In addition, text-guided pretrained models[25] and gradient-guided inpainting[26] offer new ideas. These works show that structural priors and multimodal guidance can effectively improve restoration quality.

Regarding attention mechanisms and multi-scale fusion, Chu et al.[27] introduced multiple attention mechanisms for Yongle Palace murals to optimize structural information and directional attention. Lv et al.[28] proposed a multimodal controlled diffusion model for Kizil Grottoes murals, integrating text and multidimensional visual features. Liu et al.[29] developed an edge-guided and multi-scale spatial feature fusion algorithm using rotation-invariant Scharr filters and pyramid-fused Mamba blocks to aggregate multi-scale semantics, improving peak signal-to-noise ratio (PSNR) by 0.04–0.67% and reducing learned perceptual image patch similarity (LPIPS) by 1.18–14.45% on a Dunhuang mural dataset. Zhang et al.[30] proposed CAAT-GAN, combining coordinate attention aggregated transformations with a U-Net discriminator. Liu et al.[31] designed MGRT with joint gated attention and a residual dense Transformer, reducing computational complexity from quadratic to linear. For diffusion semantic stability, Peng et al.[22] proposed the DCADif diffusion bridge framework with a structure restorer to stabilize overall semantic generation. AMAN[32] achieved blind mural inpainting via an attention-modulated adversarial network, improving FID by 12% and PSNR by 2.1 dB. A text- and edge-guided diffusion method for Thangka murals[33] balances damage repair and artistic style preservation.

In terms of structural prior guidance, DiffInpaint[34] uses line drawings as conditional input and guides the reverse generation of a latent diffusion model through two-stage training; Low-rank structure guided diffusion[35] mines global low-rank priors of murals and captures non-local similarity via orthogonal Tucker decomposition; SGPR[36] proposes a structure-guided proxy restoration workflow that masters artistic style through proxy

learning and combines selective feature extraction for high-fidelity restoration; SGCM-Net[37], based on a CNN-Mamba hybrid architecture, adaptively fuses edge priors at multi-scale levels through structure-guided fusion blocks; PRGSM[38] constructs multi-level similarity references and performs progressive reference-guided training, combined with joint sparse reference matching to suppress irrelevant responses. The above works improve mural restoration quality from perspectives such as attention optimization, multi-scale fusion, structural prior guidance, and multimodal control. The core innovation of CADDPM lies in using damage mask contours as global shape statistical priors, achieving efficient contour-aware conditional modulation through a lightweight encoder, and directly injecting global contour embeddings into the time-step embeddings of the diffusion denoising process, so that global contour conditions continuously participate in modulation at every denoising step. Meanwhile, a multi-task loss function unifies perceptual, style, and edge constraints into the optimization objective, achieving continuous coupling between global contour conditions and the generation process.

However, existing methods still have shortcomings. First, most methods treat damage masks merely as binary indicators and fail to fully exploit statistical information reflected by damaged-region contours, such as overall shape, area proportion, and distribution scale, leading to insufficient global layout constraints in large-area damage restoration[39]. Second, although style constraints have been introduced, explicit joint optimization of color harmony and style consistency is lacking, making restored regions prone to hue deviation and texture disharmony[25]. Third, inference often adopts a simple hard-covering strategy, producing visible artifacts at restoration boundaries[26]. Fourth, mural damage scales vary considerably, and existing networks lack efficient multi-scale context modeling at the bottleneck layer[19]. Fifth, existing mural restoration datasets are limited in scale, with scarce annotated samples and no unified evaluation benchmark, making fair comparison difficult[18]. From the perspective of training objectives, most methods rely on pixel-wise MSE or perceptual loss, lacking joint optimization of mural artistic style and boundary transition, so restoration results easily suffer from local texture blurring and global style drift.To address these limitations, this paper proposes a contour-aware CADDPM, which improves the baseline DDPM in four aspects:

1. A damage-contour-aware conditional modulation is designed. Sobel edge detection extracts contours from the damage mask, and a lightweight convolutional encoder with global average pooling generates a global contour embedding, which is injected into the U-Net time-step embedding to provide shape and scale priors for missing regions.

2. An atrous spatial pyramid pooling module is introduced at the U-Net bottleneck to enhance context modeling for large-area damage through multi-scale atrous convolution.

3. A multi-task joint loss function is constructed, integrating pixel-wise MSE, perceptual loss, style loss, and edge-alignment loss, to jointly constrain restoration quality from four dimensions: structure, semantics, style, and boundary.

4. A progressive sampling constraint (PSC) strategy is proposed, dividing reverse diffusion into free generation, soft blending, and hard constraint stages. During inference, known-region constraints are gradually introduced with boundary feathering to alleviate boundary artifacts.

Experimental results on the MuralDH[40] dataset show that CADDPM outperforms representative methods such as DDPM, RePaint, and LaMa in PSNR, structural similarity index measure (SSIM), and LPIPS, and ablation studies further verify the independent contributions and synergistic effects of each improvement.

Methods

Denoising Diffusion Probabilistic Model (DDPM)

As shown in Fig. 1, the DDPM[41] is a Markov chain-based generative model that generates images through two processes: a forward diffusion process that gradually adds noise to data, and a reverse denoising process that learns to recover the original data from noise.

Forward diffusion process. Gaussian noise is gradually added to the original image over T time steps. Given a clean image x_0, each step x_t is transformed according to the predefined noise schedule parameter β_t as follows:

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

where x_{t-1} is the image at step $t-1$, and β_t is the noise variance at step t, controlling the noise intensity added at each step. This process defines the single-step transition probability from x_{t-1} to x_t. Using the reparameterization trick, x_t can be sampled directly from x_0 at any time step t:

$$q(x_t|x_0)=N(x_t;\sqrt{\bar{\alpha}_t}\,x_0,(1-\bar{\alpha}_t)I) \tag{2}$$

where $\bar{\alpha}_t=\prod_{s=1}^{t}(1-\beta_t)$. As t increases, $\bar{\alpha}_t$ gradually approaches 0, and x_t gradually approaches standard Gaussian noise. The reverse denoising process aims to learn a parameterized denoising distribution $p_\theta(x_{t-1}|x_t)$, with the training objective of making the model predict the noise added at each step:

$$L_{mse}=E_{t,x_0,\varepsilon}[\|\varepsilon-\varepsilon_\theta(x_t,t)\|^2] \tag{3}$$

where ε denotes the actual noise added, ε_θ denotes the noise predicted by the model, and x_0 denotes the original clean image. This loss function trains the denoising network by minimizing the noise prediction error. During inference, the model starts from pure Gaussian noise $x_T\sim N(0,I)$ and gradually denoises to recover the original image. This paper adopts a linear noise schedule, with β increasing linearly from 1×10^{-4} to 2×10^{-2}, and a total number of steps $T=1000$.

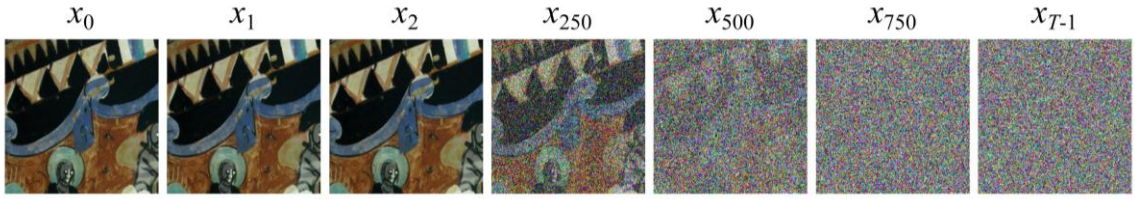


Fig.1 | Schematic illustration of the forward diffusion and reverse denoising processes of DDPM

Improved CADDPM Framework

Although standard DDPM and its inpainting variant RePaint perform well in natural image inpainting, directly applying them to mural restoration still has certain limitations.

First, such methods treat the damage mask merely as a binary indicator distinguishing known and unknown pixels, failing to exploit the geometric structure of the damaged regions themselves. Second, the training objective is only the mean squared error (MSE) of noise prediction, lacking explicit constraints on perceptual quality, artistic style, and boundary transition. Third, the inference stage adopts a simple hard-covering strategy for known regions, which easily produces visible artifacts at restoration boundaries. In addition, the scale distribution of mural damage regions varies significantly, and existing networks lack efficient multi-scale context modeling capability.

To address these limitations, this paper proposes CADDPM (Contour-Aware Denoising Diffusion Probabilistic Model), whose overall architecture is shown in Fig. 2. The method improves the baseline model in four aspects: designing a contour condition encoder to extract geometric features from the damage mask and inject them into the denoising network; proposing a progressive sampling constraint strategy that gradually introduces known-region constraints in three stages; introducing an atrous spatial pyramid pooling module at the U-Net bottleneck to enhance multi-scale feature extraction; and constructing a multi-task joint loss function to jointly constrain restoration quality from four dimensions: structure, semantics, style, and boundary.

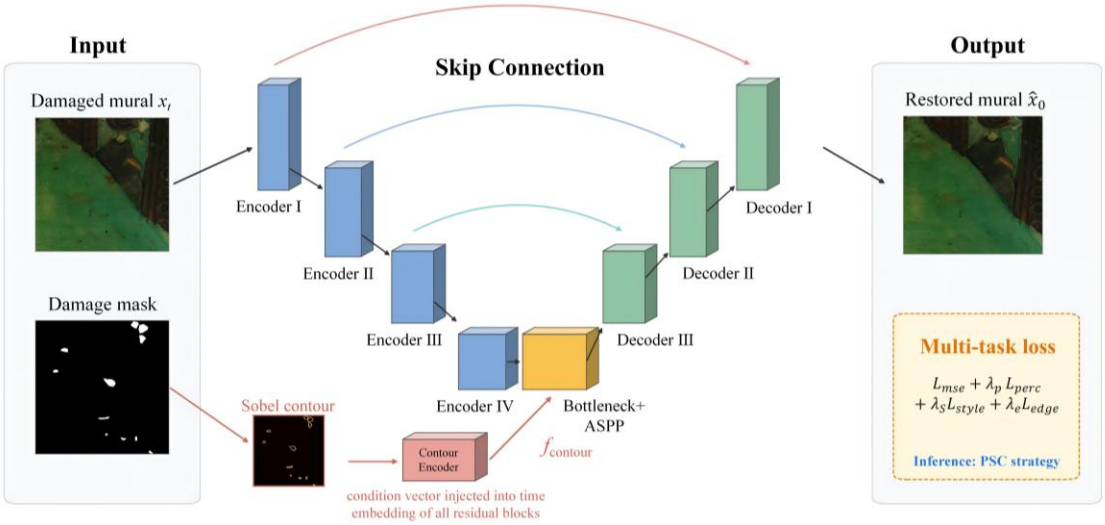


Fig. 2 | Overall architecture of CADDPM. Left: damaged mural and damage mask inputs. Middle: U-Net denoising network with an embedded ASPP module. Bottom red

path: extraction and injection of contour features. Right: restoration output and the multi-task joint loss used for training.

Damage-Contour-Aware Conditional Modulation

The boundary contour of the damage mask reflects the overall shape, spatial extent, and distribution scale of the damaged regions, and can serve as a global condition to assist the network in perceiving the location and scale of missing regions. This condition does not directly specify the internal content structure of missing regions, but rather provides a global shape statistical condition.

Given a binary damage mask $m \in \{0,1\}^{H \times W}$ (1 denotes damaged pixels), the Sobel operator is used to extract contours, yielding a two-channel contour map $C=[E_x,E_y] \in \mathbb{R}^{2 \times H \times W}$. The Sobel kernel is a fixed operator and does not participate in training. The contour map is then fed into a lightweight convolutional encoder, which consists of four cascaded convolutional layers with stride 2. The spatial resolution is halved layer by layer, while the number of feature channels is progressively deepened to 256. Each convolutional layer is equipped with Group Normalization and sigmoid linear unit (SiLU) activation, and its detailed structure is shown in Fig. 3. The spatial feature map output by the encoder is processed by global average pooling and linear projection to obtain a d-dimensional contour feature vector $f_{\text{contour}} \in \mathbb{R}^d$ (d=256).

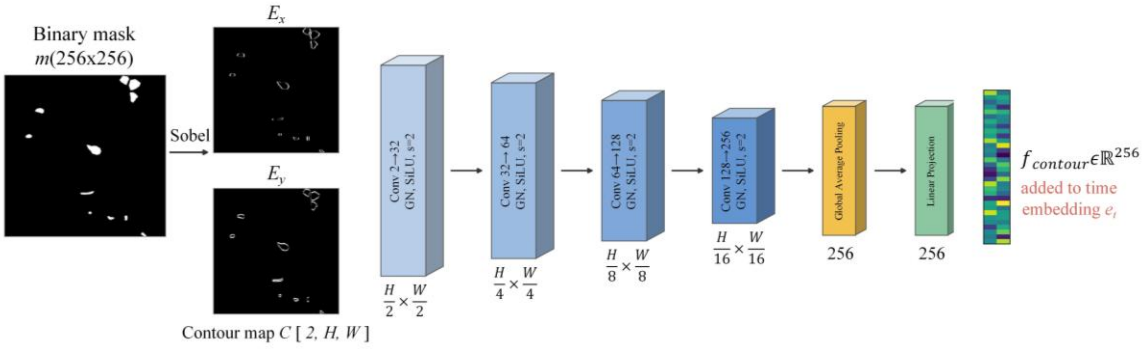


Fig. 3 | Architecture of the contour condition encoder.

The damage mask is processed by the Sobel operator to extract horizontal and vertical edge responses, forming a two-channel contour map. Four cascaded convolutional blocks progressively downsample and deepen the feature channels, and

finally output a 256-dimensional global contour embedding through global average pooling and linear projection.

The global contour embedding is used for conditional modulation by adding it to the time-step embedding. The original time-step embedding e_t is generated from sinusoidal positional encoding through a two-layer MLP. The modified embedding is:

$$e'_t = e_t + W_c \cdot f_{contour} + b_c \tag{4}$$

where W_c and b_c are learnable parameters. The modified embedding e'_t is used in all residual blocks of the U-Net, allowing the global contour statistics of damaged regions to participate in conditional modulation at every resolution level, thereby continuously providing shape and scale priors of the damaged regions during denoising. The only newly added trainable modules are the convolutional encoder and the projection layer, resulting in a negligible increase in the number of parameters.

Multi-Scale Feature Enhancement Module

To enhance the structural reconstruction capability for large-area damaged regions, an Atrous Spatial Pyramid Pooling module is introduced at the U-Net bottleneck, as shown in Fig. 4. ASPP captures multi-scale context through five parallel branches: a 1×1 convolution branch for local features, three 3×3 atrous convolution branches with dilation rates 6, 12, and 18 to progressively enlarge the receptive field, and a global average pooling branch for global semantic information. The outputs of all branches are concatenated, fused via a 1×1 convolution, and residually connected to the input feature map.

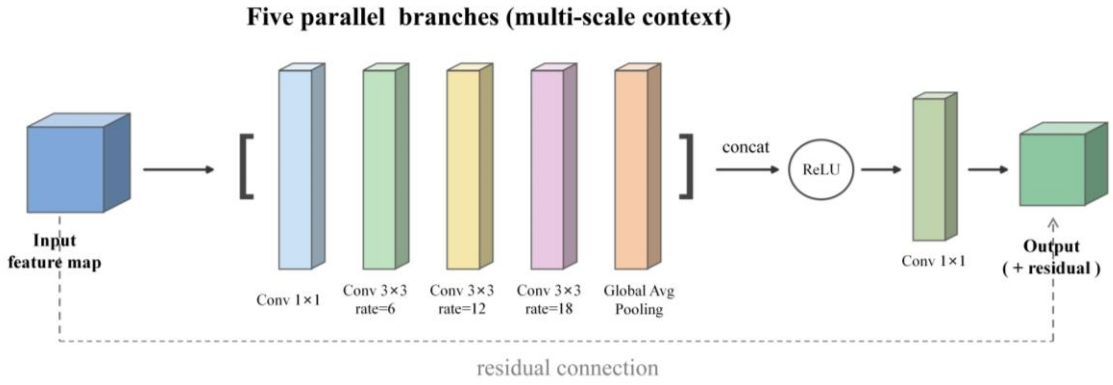


Fig. 4 | Architecture of the ASPP module.

The input feature map is processed by five parallel branches—1×1 convolution, atrous convolutions with dilation rates 6, 12, and 18, and global average pooling—to extract multi-scale context. The branch outputs are concatenated, fused via a 1×1 convolution, and residually connected to the input features.

ASPP is placed only at the bottleneck layer because the multi-scale requirements of mural restoration are concentrated at the global semantic level. The bottleneck layer controls the parameter count while enabling the model to acquire multi-scale perception at the the highest semantic level.

Multi-Task Joint Loss Function

Standard DDPM only optimizes the MSE loss for noise prediction. For mural restoration, the restored content must simultaneously satisfy structural accuracy, semantic consistency, stylistic harmony, and natural boundary transition. We recover x_0 from the noise prediction:

$$\hat{x}_0=\frac{x_t-\sqrt{1-\bar{\alpha}_t}\cdot\varepsilon_\theta(x_t,t)}{\sqrt{\bar{\alpha}_t}} \tag{5}$$

Three complementary losses are introduced on $\hat{x}_0$. The perceptual loss[42] measures the weighted L_1 distance between $\hat{x}_0$ and the ground-truth image x_0 across multiple feature layers of a pretrained VGG-19 network:

$$L_{perc}=\sum_k w_k\cdot|\phi_k(\hat{x}_0)-\phi_k(x_0)|_1 \tag{6}$$

Where ϕ_k denotes the output of the k-th feature block of VGG-19. Four feature blocks from shallow to deep are selected, with corresponding weights w_k set to 1.0, 0.5, 0.25, and 0.125, respectively. The VGG parameters are frozen during training. Murals have distinctive color distributions and texture patterns. To preserve these stylistic attributes in the restored regions, a Gram matrix-based style loss[43] is introduced:

$$L_{style}=\sum_k|G_k(\hat{x}_0)-G_k(x_0)|_1 \tag{7}$$

where G_k denotes the Gram matrix of the k-th VGG feature block. Visible seams at restoration boundaries are a common artifact, so an edge alignment loss is designed.

$$L_{\text{edge}}=\frac{1}{|B|}\sum_{p\in B}|\nabla\hat{x}_0(p)-\nabla x_0(p)| \tag{8}$$

where B denotes the boundary region of the damage mask defined by morphological dilation and erosion (kernel radius r=3). The final training objective is a weighted combination of the four losses:

$$L=\lambda_{\text{mse}}L_{\text{mse}}+\lambda_{\text{perc}}L_{\text{perc}}+\lambda_{\text{style}}L_{\text{style}}+\lambda_{\text{edge}}L_{\text{edge}} \tag{9}$$

In the experiments, we set λ_{mse}=1.0, λ_{perc}=0.1, λ_{style}=0.5, and λ_{edge}=0.2. The style loss is assigned a relatively large weight because mural restoration imposes extremely high requirements on color and texture harmony, and color deviation or texture disharmony in restored regions can severely degrade visual quality.

Progressive Sampling Constraint Strategy

During inference, undamaged regions should remain unchanged. RePaint performs inpainting by replacing known regions at each reverse diffusion step. However, direct hard replacement in the early denoising stage introduces discontinuities at boundaries. As shown in Fig. 5, this paper proposes a Progressive Sampling Constraint (PSC) strategy, which divides the reverse diffusion process into three stages controlled by two ratio parameters, r_{soft} and r_{hard}:

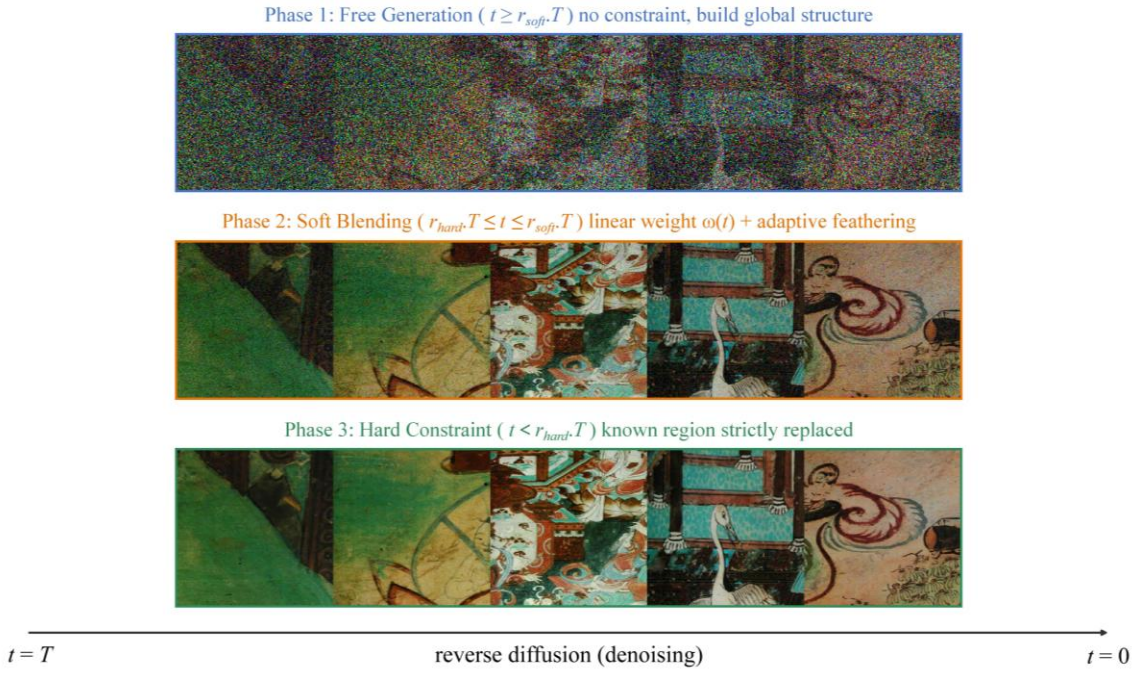


Fig. 5 | Schematic illustration of the progressive sampling constraint strategy.

The reverse diffusion process is divided by time steps into three stages: free generation, soft blending, and hard constraint. From top to bottom, these correspond to high-noise, low-noise, and final denoised results, respectively, and the constraint strength on known regions gradually increases as denoising proceeds.

Stage 1 — Free generation ($t \geq r_{soft} \cdot T$): no constraint is applied, and the model freely generates content over the entire image to establish a globally coherent structure.

Stage 2 — Soft blending ($r_{hard} \cdot T \leq t < r_{soft} \cdot T$): known regions are gradually blended with the model output through a linearly increasing weight:

$$\omega(t) = \frac{r_{soft} \cdot T - t}{(r_{soft} - r_{hard}) \cdot T} \tag{10}$$

The blended result for each pixel is:

$$x_{t-1} = \omega(t) \cdot m x_{t-1}^{orig} + (1 - \omega(t) m) \cdot x_{t-1}^{model} \tag{11}$$

where x_{t-1}^{orig} is the result of adding noise to the original image up to step $t-1$, and x_{t-1}^{model} is the denoised output of the model. The mask boundary is feathered by mean filtering, and the adaptive kernel radius decreases as $\omega(t)$ increases.

Stage 3 — Hard constraint ($t < r_{hard} \cdot T$): known regions are strictly covered by the original image:

$$x_{t-1} = m \cdot x_{t-1}^{orig} + (1 - m) \cdot x_{t-1}^{model} \tag{12}$$

In the experiments, $r_{soft}=0.6$, $r_{hard}=0.3$, and the maximum feathering radius $r_{max}=5$ pixels. These thresholds were determined by grid search on the validation set.

Experimental Results and Analysis

To verify the effectiveness of the proposed method, this paper systematically evaluates CADDPM from three dimensions: quantitative evaluation, qualitative comparison, and ablation experiments, and compares it with mainstream inpainting methods such as DDPM, RePaint, and LaMa.

Experimental Setup

Dataset

This paper uses the MuralDH dataset and the public subset of DiffuMural[44]. MuralDH was constructed by Xu et al. and contains 961 damaged mural images with pixel-level damage mask annotations, covering typical deterioration types such as paint loss, cracks, and fading. As shown in Table 1, we split the dataset into 700/110/201 for training, validation, and test sets; the remaining 4,129 unannotated images are used to augment the training samples. The public subset of DiffuMural contains 263 Dunhuang mural images without mask annotations, all of which are included in the training set. Together, they comprise a total of 1,224 annotated images.

During training, images with paired masks directly use the real masks; unannotated images, including Mural512 and the DiffuMural subset, adopt block, crack, and mixed random mask generation strategies to synthesize training data. The training set contains 5,092 images in total. All images are uniformly resized to 256×256, and random horizontal flipping is used for data augmentation. The supplementary images are only used to augment the training set and do not participate in test evaluation, so as to ensure fair comparison with the MuralDH benchmark.

Table 1 | Statistics of the MuralDH dataset. Includes synthetic data with random masks from Mural512 and the DiffuMural subset

Subset	Total	Training	Validation	Test
Mural_seg	961	700	110	201
DiffuMural subset	263	263		
Other mural images	4,129	—	—	—
Total	5353	5092*	110	201

Experimental Environment

Hardware environment: GPU: NVIDIA GeForce RTX 4090 (24 GB VRAM); CPU: 16-core processor; RAM: 64 GB.

Software environment: Operating system: Ubuntu 22.04 LTS; deep learning framework: PyTorch 2.x; CUDA version: 12.x; Python version: 3.10.

Training configuration: Adam optimizer with an initial learning rate of 2×10^{-5}, batch size 8, and 100 training epochs. A linear noise schedule is adopted, with diffusion steps T=1000 and a gradient clipping threshold of 1.0. The inference time is approximately 4.8 s per image.

Evaluation Protocol

Since the MuralDH test set consists entirely of real damaged murals without corresponding clean ground-truth images, this paper follows the evaluation protocol of the original MuralDH paper. Specifically, 200 mural images are randomly selected from Mural512, and irregular masks covering 10%–20% of the image area are generated on each image, including block, crack, and mixed types. The masked regions are blackened and fed into each inpainting model. The original image before mask coverage is used as ground truth, and three reference-based metrics—PSNR, SSIM, and LPIPS—are computed. This protocol is consistent with the original MuralDH paper to ensure experimental comparability. All metrics are computed uniformly over the entire image.

Evaluation Metrics

This paper adopts three widely used image inpainting evaluation metrics, covering pixel-level, structure-level, and perceptual-level aspects.

(1) PSNR, in dB. PSNR measures the pixel-level reconstruction quality between the restored image and the reference image based on MSE, calculated as:

$$PSNR=10\cdot\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{13}$$

where MAX_I is the maximum pixel value (255 for 8-bit images), and MSE is the mean squared error between the restored and reference images. A higher PSNR indicates better pixel-level reconstruction quality.

(2) SSIM measures the similarity between two images from three dimensions: luminance, contrast, and structure, calculated as:

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \tag{14}$$

where μ_x and μ_y are local means, σ_x^2 and σ_y^2 are variances, σ_{xy} is covariance, and c_1 and c_2 are stabilizing constants. SSIM ranges from [0, 1], with values closer to 1 indicating better structural consistency.

(3) LPIPS uses a pretrained VGG network to extract deep features and computes the weighted distance between the restored and reference images in feature space, calculated as:

$$LPIPS(x,y)=\sum_l \frac{1}{H_l W_l}\sum_{h,w} \left| w_l \odot (\hat{x}^l_{hw}-\hat{y}^l_{hw}) \right|_2^2 \tag{15}$$

where $\hat{x}^l_{hw}$ and $\hat{y}^l_{hw}$ are channel-normalized features at layer l, H_l and W_l are the spatial dimensions of the feature map at that layer, W_l is the channel weighting vector, and $\odot$ denotes element-wise multiplication. A lower LPIPS indicates higher perceptual similarity and better consistency with human subjective evaluation.

Comparison Methods

To comprehensively evaluate the performance of CADDPM, this paper selects three representative image inpainting methods as comparison baselines:

(1) DDPM (Denoising Diffusion Probabilistic Models): a basic inpainting method based on the denoising diffusion probabilistic model. It adopts a standard hard inpainting sampling strategy, replacing known regions with the noised version of the original image after each denoising step, and uses MSE as the loss function. As a baseline for diffusion-based inpainting, DDPM provides a reference for the improvements in this paper.

(2) RePaint: a mask-conditioned sampling method based on DDPM (CVPR 2022). RePaint performs multiple resampling operations (denoise→renoise→merge) within each denoising time step, allowing information from known regions to propagate more fully into the masked regions and improving the quality of restoration boundaries. In this paper, the number of resampling operations per step is set to 5.

(3) LaMa (Large Mask Inpainting)[45]: a GAN-based image inpainting method (WACV 2022). LaMa employs Fast Fourier Convolution (FFC) to obtain a global receptive field and combines a high-receptive-field perceptual loss with a large-mask training strategy. It performs well in large-mask inpainting tasks and is a representative work among GAN-based inpainting methods.

(4) CADDPM: introduces four improvements upon DDPM: contour condition injection, ASPP-based multi-scale feature enhancement, a multi-task joint loss function, and a progressive sampling constraint (PSC) strategy.

Except for LaMa, which directly adopts the official big-lama pretrained weights, DDPM, RePaint, and CADDPM are all trained on the same MuralDH training set under the same training configuration, and tested under the same synthetic evaluation protocol to ensure fair comparison.

Quantitative Comparison Experiments

Table 2 presents the quantitative comparison results of DDPM, RePaint, LaMa, and CADDPM on the synthetic evaluation data. All metrics are computed uniformly on the 200 synthetic evaluation images.

Table 2 | Quantitative comparison results of different inpainting methods on the MuralDH synthetic evaluation data.

Method	PSNR (dB)↑	SSIM↑	LPIPS↓
DDPM	32.95	0.946	0.050
RePaint	33.07	0.950	0.045
LaMa	31.34	0.939	0.042
CADDPM（Ours）	**34.49**	**0.956**	**0.041**

As shown in Table 2, CADDPM achieves the best results on all three metrics: PSNR, SSIM, and LPIPS. Compared with DDPM, PSNR improves by 1.54 dB, SSIM by 0.010, and LPIPS decreases by 0.009. Compared with RePaint, PSNR improves by 1.42 dB, SSIM by 0.006, and LPIPS decreases by 0.004. Compared with LaMa, PSNR improves by 3.15 dB, SSIM by 0.017, and LPIPS decreases by 0.001. LaMa obtains an LPIPS of

0.042, close to CADDPM's 0.041, indicating that GAN-based methods have certain advantages in perceptual quality; however, CADDPM's advantages in PSNR and SSIM are more pronounced, demonstrating its superior overall performance in pixel-level and structure-level reconstruction. DDPM and CADDPM correspond to the Base and Full configurations in the ablation study Table 3, respectively; values are obtained from the same 200-image synthetic evaluation protocol. RePaint uses 5 resampling operations per step; LaMa adopts the official big-lama pretrained weights without fine-tuning on MuralDH.

From the perspective of method mechanisms, DDPM adopts standard covering sampling, and known regions and restored regions lack smooth transition at boundaries. RePaint improves information propagation through multiple resampling operations but lacks geometric guidance. LaMa employs single-step forward generation, which has limited control over complex color gradients and fine lines in murals. CADDPM achieves consistent superiority across the three metrics through the synergistic effect of four mechanisms: contour conditions, multi-scale features, multi-task loss, and progressive sampling.

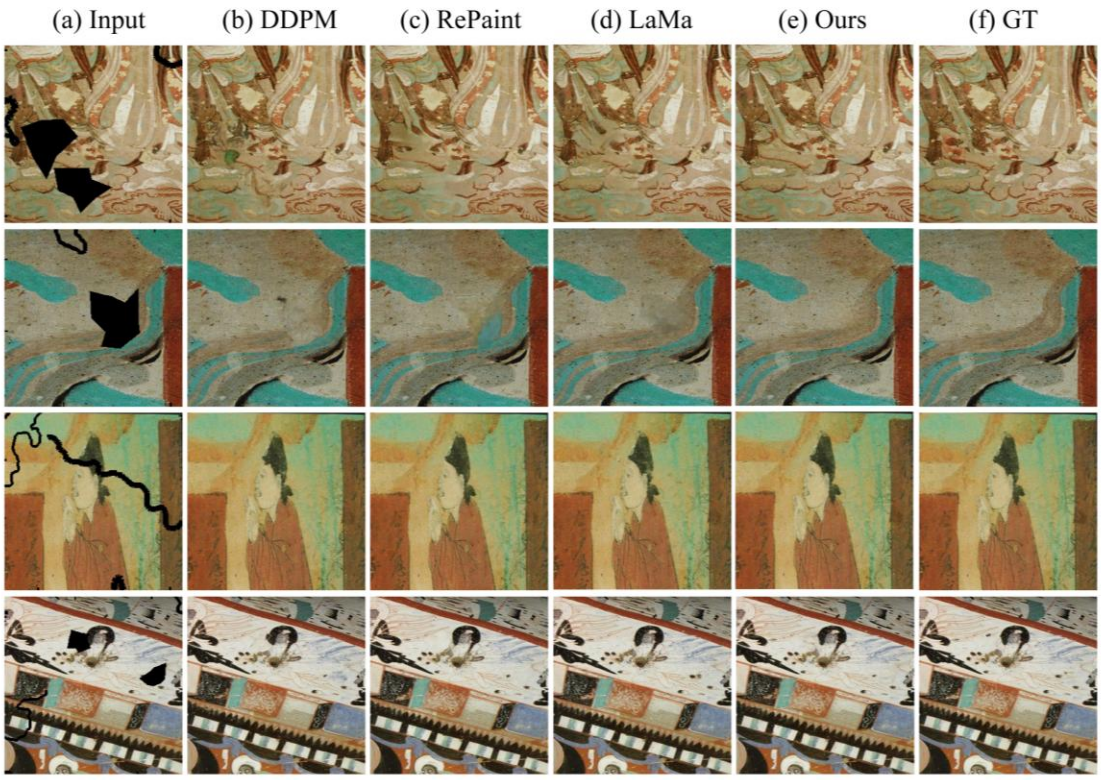


Fig. 6 | Qualitative comparison of different inpainting methods on MuralDH synthetic evaluation samples. Each row is one sample; columns are, in order: (a) damaged input, (b) DDPM, (c) RePaint, (d) LaMa, (e) CADDPM (Ours), (f) Ground Truth.

Fig. 6 shows the qualitative restoration results of each method on synthetic evaluation samples. This set of samples uses the same protocol as the quantitative evaluation in Table 2, facilitating intuitive comparison of reconstruction fidelity among methods with Ground Truth. Fig. 7 further shows the restoration results of each method on real damaged murals.

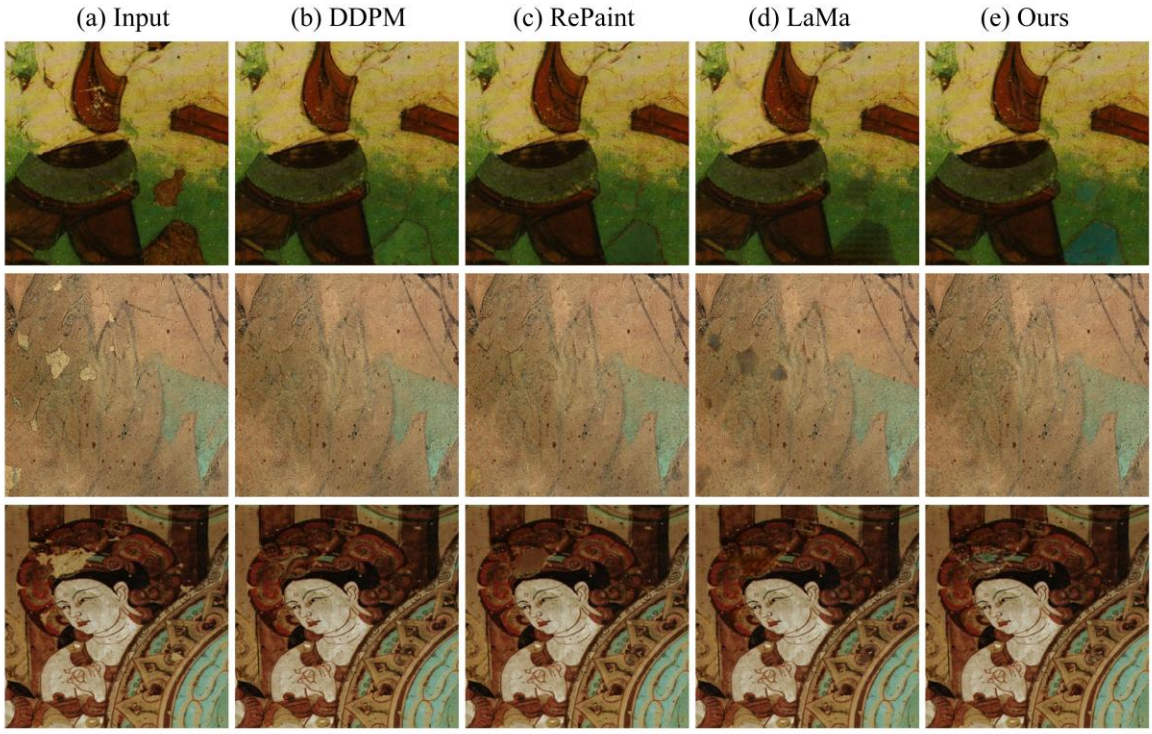


Fig. 7 | Restoration results of each method on real damaged murals. Each row is one sample; columns are, in order: (a) real damaged input, (b) DDPM, (c) RePaint, (d) LaMa, (e) CADDPM (Ours). Real damaged murals have no clean reference images, so this is only a qualitative comparison.

From the qualitative comparison in Fig. 6 and Fig. 7, the performance differences among methods are significant. DDPM produces obvious artifacts at the boundaries of restored regions because standard covering sampling lacks a smooth transition

mechanism at the junction. RePaint improves boundary transition, but texture recovery inside large-area damage regions is insufficient, and the restoration results are overall blurry. LaMa performs relatively well at the edges of restored regions, but internal textures exhibit a smoothing effect, with considerable loss of high-frequency details such as mural lines and color gradients. In contrast, CADDPM performs best in structural integrity, color harmony, and boundary transition. The global contour condition helps stabilize the global layout, the style loss ensures hue consistency, and the feathered soft constraint together with the edge alignment loss eliminates seam artifacts.

Ablation Study

To verify the independent effectiveness of the four improvements, this paper sets up six ablation configurations: Base uses standard DDPM, hard covering sampling, and MSE loss as the baseline; −Contour removes the contour condition; −ASPP removes the ASPP multi-scale feature enhancement; −Loss replaces the multi-task loss with MSE; −PSC replaces progressive sampling with hard covering; Full is the complete model. All configurations are trained on the same dataset and under the same training settings.

Table 3 | Ablation results of each improvement module in CADDPM. Note: ΔPSNR is the difference between each configuration and the full model Full; negative values indicate lower than Full.

Configuration	PSNR (dB)↑	Δ PSNR vs Full	SSIM↑	LPIPS↓
Base	32.95	−1.54	0.946	0.050
− Contour	32.75	−1.74	0.940	0.051
− ASPP	32.72	−1.77	0.939	0.052
− Loss	33.68	−0.81	0.950	0.047
− PSC	33.39	−1.10	0.949	0.044
Full（Ours）	**34.49**	**—**	**0.956**	**0.041**

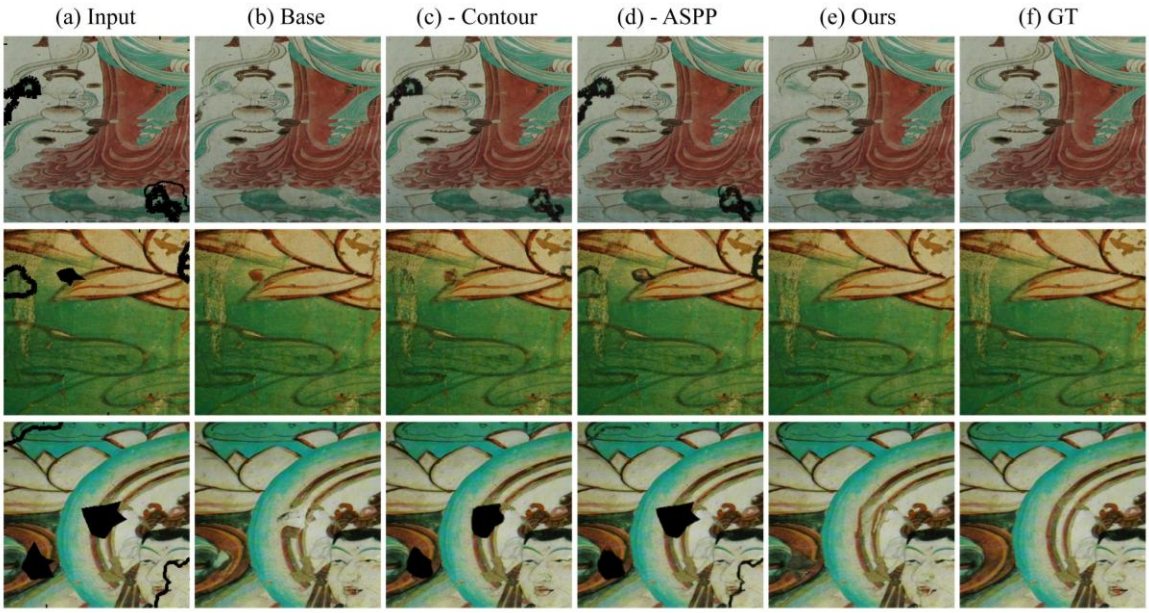


Fig. 8 | Visual comparison of ablation experiments. Each row is one sample; columns are, in order: (a) damaged input, (b) Base, (c) −Contour, (d) −ASPP, (e) Full (Ours), (f) Ground Truth.

Taking the full model Full as the reference, the effects of removing each improvement based on the results in Table 3 are as follows: (1) After removing the contour condition, PSNR decreases by 1.74 dB, SSIM by 0.016, and LPIPS increases by 0.010, indicating that contour information plays the most critical role in constraining the global layout of restored regions, and its LPIPS degradation is the largest among the four, making it the main contributor to perceptual quality. (2) After removing ASPP, PSNR decreases by 1.77 dB, SSIM by 0.017, and LPIPS increases by 0.011, indicating that the multi-scale receptive field helps simultaneously recover large-scale structures and fine textures. (3) After replacing the multi-task loss with MSE, PSNR decreases by 0.81 dB, SSIM by 0.006, and LPIPS increases by 0.006, indicating that perceptual and style losses contribute significantly to restoration quality. (4) After replacing progressive sampling with hard covering, PSNR decreases by 1.10 dB, SSIM by 0.007, and LPIPS increases by 0.003, indicating that progressive constraints during inference help eliminate boundary artifacts.

Overall, the four improvements synergistically enhance restoration quality across different dimensions: the contour condition provides structural priors, ASPP enhances

multi-scale feature representation, the multi-task loss constrains the optimization direction from pixel, perceptual, style, and edge levels, and progressive sampling ensures structural consistency during inference.

Discussion

Analysis of the Role of Global Contour Condition Modulation

Global contour condition modulation is one of the core improvements of our method. Combining the ablation metrics in Table 3 with the visual comparison in Fig. 8, it can be observed that the benefits of the contour condition are more pronounced on large-area damage. This is because the contour statistics of large-area damaged regions better reflect the overall shape and scale of the damage; after being injected into the noise prediction network through the contour encoder, they can help stabilize the global layout of large-area damage restoration. For narrow damage such as fine cracks, the gain from the contour condition is relatively small, because the restoration of such damage mainly relies on texture propagation from surrounding known regions.

Roles of the Components in the Multi-Task Loss

The multi-task joint loss adopted in this paper contains four components, each playing a different role. The MSE loss has a weight of 1.0, ensuring basic pixel-level denoising quality; the perceptual loss has a weight of 0.1, using pretrained VGG features to ensure semantic consistency; the style loss has a weight of 0.5, the highest among the four, and is used to maintain color and texture harmony between restored regions and surrounding regions, which is particularly important for mural restoration because large-area color gradients and texture patterns require overall stylistic consistency; the edge alignment loss has a weight of 0.2, specifically eliminating seam artifacts at the boundaries between restored regions and known regions. In the ablation experiments, after replacing the multi-task loss with a single MSE, PSNR decreases by 0.81 dB, indicating that the multi-task loss still makes a considerable contribution to pixel-level quality; the roles of the perceptual and style components are mainly reflected at the perceptual quality level.

Design Principle of the Progressive Sampling Constraint Strategy

The design of the progressive sampling constraint strategy originates from observations of the diffusion denoising process. In the high-noise stage, where tt is large, the image structure is not yet determined; imposing strict constraints on known regions too early would interfere with the diffusion process's exploration of restoration content. In the low-noise stage, where tt is small, the image structure is largely determined, and precise restoration of details is required. PSC applies stage-wise constraint weights to known regions according to the denoising progress t/T: (1) Initial stage (t/T>rsoft=0.6t/T>rsoft=0.6): the constraint weight is 0, the restoration region is freely generated by the model, and known regions maintain pixel fidelity using the noised version of the original image; (2) Transition stage (r_{hard}=0.3 < $t/T \leq r_{soft}$): the constraint weight increases linearly from 0 to 1, and feathering is introduced at the boundaries of known regions. The feathering radius adaptively decreases as the constraint strength increases, enabling a smooth transition of the blend from the model output to the original image. This is the key mechanism for eliminating seam artifacts between restored and known regions; (3) Final stage ($t/T \leq r_{hard}$=0.3): the constraint weight is 1, strictly covering known regions and refining details. Compared with direct hard covering, PSC provides a smooth transition from free generation to constraint in the middle and late stages, avoiding abrupt artifacts at boundaries.

Comparative Analysis with Existing Methods

Compared with the GAN-based method (LaMa), the stepwise denoising mechanism of CADDPM is better suited to the fine-grained control required for mural restoration. GAN-generates restoration results through a single forward pass; although inference is fast, it is prone to texture smoothing or mode degradation when handling the complex color gradients and fine lines commonly found in murals. Diffusion models perform multi-step iterative denoising, where each step has the opportunity to correct deviations from the previous step, giving them a natural advantage in fine restoration.

Compared with RePaint, which is also a diffusion model, CADDPM's advantages come from two aspects: first, global contour condition modulation provides additional shape and scale guidance, compensating for the lack of global structure in purely data-

driven restoration of large-area damage; second, the multi-task loss constrains restoration quality from four dimensions—pixel, perceptual, style, and edge—whereas RePaint relies only on MSE loss.

In addition, several new diffusion-based methods have emerged in the field of Dunhuang mural restoration in recent years. Among them, DiffuMural [18] is most relevant to this work. Published in npj Heritage Science, it adopts multi-scale convergence and cooperative guidance mechanisms, combines ControlNet with a cycle consistency loss, and is trained on approximately 650,000 multi-scale samples from 25 Tang-dynasty caves. Currently, this work has not released its source code or training data, and its evaluation protocol is also inconsistent with the synthetic mask protocol used in this paper, making direct comparison under fully equivalent conditions difficult; even if its published metrics are cited, the numbers are not directly comparable due to differences in training data scale and evaluation protocol. This paper and DiffuMural have different methodological emphases: DiffuMural emphasizes data-driven multi-scale cooperative generation, whereas CADDPM centers on lightweight global contour condition modulation, achieving structural prior modeling under conditions of relatively limited annotated data. Early mural diffusion restoration methods such as MuralDiff also face similar situations. After the above methods release their code or a unified evaluation benchmark is established, this work will conduct further systematic comparisons.

Conclusion

This study investigates key challenges in mural image inpainting, focusing on core issues such as insufficient utilization of geometric structure information in damaged regions, stylistic inconsistency of restored content, and boundary artifacts. To this end, this paper proposes a contour-aware denoising diffusion probabilistic model (CADDPM), constructing a unified restoration framework from four dimensions: structural prior injection, multi-scale feature enhancement, multi-task joint optimization, and inference constraints. First, a damage-contour-aware conditional modulation is designed. The Sobel operator extracts geometric contours from the damage mask, and a lightweight convolutional encoder generates a contour feature vector, which is injected into the U-Net time-step embedding to provide shape and scale priors for missing regions. Second,

an atrous spatial pyramid pooling module is introduced at the U-Net bottleneck to enhance context modeling for large-area damage through multi-scale atrous convolution. Meanwhile, a multi-task joint loss function integrating pixel, perceptual, style, and edge-alignment terms is constructed to jointly constrain restoration quality from four dimensions: structure, semantics, style, and boundary. Finally, a progressive sampling constraint strategy is proposed, dividing reverse diffusion into three stages: free generation, soft blending, and hard constraint, effectively alleviating boundary artifacts in restoration. Experimental results on the MuralDH dataset show that CADDPM outperforms representative methods such as DDPM, RePaint, and LaMa on PSNR, SSIM, and LPIPS, and ablation studies further verify the independent contributions and synergistic effects of each improvement. Future research will focus on introducing self-supervised pretraining strategies to make full use of unlabeled data, exploring accelerated sampling techniques such as DDIM to shorten inference time, and incorporating subjective ratings from cultural heritage conservation experts to build a more comprehensive evaluation system, further enhancing the cultural authenticity and long-term preservation value of restoration results.

Data availability

No datasets were generated or analyzed during the current study.

References

1. Li, J. et al. Investigation of the renewed diseases on murals at Mogao Grottoes. Herit. Sci. 1, 31 (2013). https://doi.org/10.1186/2050-7445-1-31

2. Li, N., Su, B., Wang, X. & Ding, S. Investigation of the soot-covered murals in the caves of the Dunhuang Mogao Grottoes. Res. Conserv. Cave Temples Earthen Sites 4, 17 – 29 (2025). https://doi.org/10.20182/j.cnki.ISSN2097-1370.202504002

3. Hua, S. et al. A review of deep learning-based digital restoration methods for murals. In Advanced Intelligent Computing Technology and Applications (Springer, 2026). https://doi.org/10.1007/978-981-92-3429-5_13

4. Buades, A., Coll, B. & Morel, J.-M. Non-local means denoising. Image Process. On Line 1, 208 – 212 (2011). https://doi.org/10.5201/ipol.2011.bcm_nlm

5. Criminisi, A., Pérez, P. & Toyama, K. Region filling and object removal by exemplar-based image inpainting. IEEE Trans. Image Process. 13, 1200 – 1212 (2004). https://doi.org/10.1109/TIP.2004.833105

6. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T. & Efros, A. A. Context encoders: feature learning by inpainting. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) 2536 – 2544 (IEEE, 2016). https://doi.org/10.1109/CVPR.2016.278

7. Yu, J. et al. Generative image inpainting with contextual attention. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) 5505 – 5514 (IEEE, 2018). https://doi.org/10.1109/CVPR.2018.00577

8. Liu, G. et al. Image inpainting for irregular holes using partial convolutions. In Proc. Eur. Conf. Comput. Vis. (ECCV) 85 – 100 (Springer, 2018). https://doi.org/10.1007/978-3-030-01252-6_6

9. Zeng, Y., Fu, J., Chao, H. & Guo, B. Learning pyramid-context encoder network for high-quality image inpainting. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) 1486 – 1494 (IEEE, 2019). https://doi.org/10.1109/CVPR.2019.00158

10. Wan, Z., Zhang, J., Chen, D. & Liao, J. High-fidelity pluralistic image completion with transformers. In Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV) 4692 – 4701 (IEEE, 2021). https://doi.org/10.1109/ICCV48922.2021.00465

11. Lugmayr, A. et al. RePaint: inpainting using denoising diffusion probabilistic models. In Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) 11461 – 11471 (IEEE, 2022). https://doi.org/10.1109/CVPR52688.2022.01114

12. Rombach, R., Blattmann, A., Lorenz, D., Esser, P. & Ommer, B. High-resolution image synthesis with latent diffusion models. In Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) 10684 – 10695 (IEEE, 2022). https://doi.org/10.1109/CVPR52688.2022.01042

13. Chen, M., Feng, M., Lin, H. & Xing, W. A priori-based adaptive repair method for mural images. In Proc. 3rd Int. Conf. Comput. Sci. Appl. Eng. (CSAE) 1 – 6 (ACM, 2019). https://doi.org/10.1145/3331453.3361295

14. Xiao, H., Zheng, H. & Meng, Q. Research on deep learning-driven high-resolution image restoration for murals from the perspective of vision sensing. IEEE Access 11, 71472 – 71483 (2023). https://doi.org/10.1109/ACCESS.2023.3295253

15. Chen, Y., Fan, Z., Zhang, S. et al. A novel Brownian bridge diffusion-based generative inpainting algorithm for ancient murals. Eng. Appl. Artif. Intell. 165, 113367 (2026). https://doi.org/10.1016/j.engappai.2025.113367

16. Ren, H., Zhu, X., Zhang, C. & Li, Z. Inpainting of figure murals based on a dual-branch diffusion model. In Proc. 2025 IEEE/ACIS 29th Int. Conf. Softw. Eng. Artif. Intell. Netw. Parallel/Distrib. Comput. (SNPD) 944 – 949 (IEEE, 2025). https://doi.org/10.1109/SNPD65828.2025.11254471

17. Jia, R. et al. ST-DiffNet: diffusion-based inpainting of Dunhuang murals with structural and textural guidance. Multimed. Syst. 31, 467 (2025). https://doi.org/10.1007/s00530-025-02056-1

18. Han, P. et al. DiffuMural: a diffusion model for Dunhuang murals restoration via multi-scale convergence and cooperative guidance. npj Herit. Sci. 14, 44 (2026). https://doi.org/10.1038/s40494-025-02274-w

19. Li, Y. et al. An improved mural image restoration method based on diffusion model. npj Herit. Sci. 13, 347 (2025). https://doi.org/10.1038/s40494-025-01914-5

20. Li, Y. et al. DMEA-Net: a diffusion-based network for mural image restoration using enhanced multi-head contextual attention. npj Herit. Sci. 14, 348 (2026). https://doi.org/10.1038/s40494-026-02591-8

21. Zhang, L. et al. Mural inpainting for structural distortion and color deviation: edge-constrained diffusion and adversarial optimization strategy. npj Herit. Sci. 14, 363 (2026). https://doi.org/10.1038/s40494-026-02724-z

22. Peng, X. et al. DCADif: decoupled conditional adaptive time-dynamic fusion diffusion inpainting of traditional Chinese mural paintings. npj Herit. Sci. 14, 02327 (2026). https://doi.org/10.1038/s40494-026-02327-8

23. Zhao, D., Chen, X., Zhang, D. & Li, J. APLDiff: an adaptive perception-driven lightweight diffusion framework for digital mural inpainting. npj Herit. Sci. 14, 02275 (2026). https://doi.org/10.1038/s40494-025-02275-9

24. Ren, N. & Huang, K. MAPGR: multi-agent prompt-guided residual diffusion for ancient mural restoration. npj Herit. Sci. 14, 319 (2026). https://doi.org/10.1038/s40494-026-02607-3

25. Li, G., Pan, Q. & Yang, P. K2Mural: text-guided pre-trained model for mural image inpainting. In Advanced Intelligent Computing Technology and Applications (Springer, 2025). https://doi.org/10.1007/978-981-96-9869-1_40

26. Grechka, A., Couairon, G. & Cord, M. GradPaint: gradient-guided inpainting with diffusion models. Comput. Vis. Image Underst. 240, 103928 (2024). https://doi.org/10.1016/j.cviu.2024.103928

27. Chu, E. & Chu, X. Yongle Palace murals' image restoration using deep learning optimization algorithms. npj Herit. Sci. 13, 683 (2025). https://doi.org/10.1038/s40494-025-02199-4

28. Lv, G. et al. Virtual restoration method of Kizil Grotto murals based on multimodal controlled diffusion models. npj Herit. Sci. 13, 554 (2025). https://doi.org/10.1038/s40494-025-01981-8

29. Liu, Z., Liu, Y. & Hu, W. Mural restoration via the fusion of edge-guided and multi-scale spatial features. npj Herit. Sci. 13, 491 (2025). https://doi.org/10.1038/s40494-025-02073-3

30. Zhang, J. et al. Supporting historic mural image inpainting by using coordinate attention aggregated transformations with U-Net-based discriminator. npj Herit. Sci. 13, 305 (2025). https://doi.org/10.1038/s40494-025-01891-9

31. Liu, Z., Liu, Y. & Hu, W. Mural restoration via joint gated attention and residual dense transformer. Infrared Laser Eng. 54, 20250052 (2025). https://doi.org/10.3788/IRLA20250052

32. Lan, S., Zhang, S., Zhong, H., Li, R. & Zhang, Y. AMAN: attention-modulated adversarial network for blind mural image completion. Displays 86, 103296 (2026). https://doi.org/10.1016/j.displa.2025.103296

33. Hsieh, T. et al. Text and edge guided Thangka image inpainting with diffusion model. In Proc. IEEE Int. Conf. Multimed. Expo (ICME) 1 – 10 (IEEE, 2024). https://doi.org/10.1109/ICME57554.2024.10687698

34. Tang, X., Sui, Y., Sun, K. & Xiang, L. DiffInpaint: line drawing guided murals restoration with diffusion model. Measurement 258, 119223 (2026). https://doi.org/10.1016/j.measurement.2025.119223

35. Lei, T., Li, Y., Cheng, Z. & Zhou, W. Low-rank structure guided diffusion for Shaanxi temple mural restoration. npj Herit. Sci. 13, 295 (2025). https://doi.org/10.1038/s40494-025-01827-3

36. Yang, C., Liu, Y. & Cai, Y. Digital restoration of ancient Jiangnan murals via proxy learning and structural guidance. npj Herit. Sci. 14, 182 (2026). https://doi.org/10.1038/s40494-026-02369-y

37. Qiu, Y. et al. SGCM-Net: structure-guided mural inpainting with state space model. Sci. Rep. 16, 23429 (2026). https://doi.org/10.1038/s41598-026-53498-0

38. Liu, H. et al. Inpainting of Dunhuang murals based on progressive reference guidance and sparse feature matching. npj Herit. Sci. 14, 02971 (2026). https://doi.org/10.1038/s40494-026-02971-0

39. Liu, H., Wang, Y., Qian, B., Wang, M. & Rui, Y. Structure matters: tackling the semantic discrepancy in diffusion models for image inpainting. In Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) 8038 – 8047 (IEEE, 2024). https://doi.org/10.1109/CVPR52733.2024.00768

40. Xu, Z. et al. A comprehensive dataset for digital restoration of Dunhuang murals. Sci. Data 11, 955 (2024). https://doi.org/10.1038/s41597-024-03785-0

41. Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. In Advances in Neural Information Processing Systems 33, 6840 – 6851 (2020).

42. Johnson, J., Alahi, A. & Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Computer Vision – ECCV 2016 694 – 711 (Springer, 2016).

43. Gatys, L. A., Ecker, A. S. & Bethge, M. Image style transfer using convolutional neural networks. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) 2414 – 2423 (IEEE, 2016).

44. DiffuMural murals dataset. Hugging Face Datasets https://huggingface.co/datasets/steven191/murals (2024).

45. Suvorov, R. et al. Resolution-robust large mask inpainting with Fourier convolutions. In Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV) 2149 – 2159 (IEEE, 2022).

Acknowledgements

The authors gratefully acknowledge the financial support from the National Social Science Fund of China (Art Studies) (Grant No. 21BG107) and the Shaanxi Sanqin Talents Special Support Program—Provincial First-Class Team (Silk Road Culture and Art Design Innovation Team, Shaanxi University of Science and Technology).

Author contributions

All authors have contributed to the current work. W.Q. conceived the overall structure and writing direction of the manuscript, drafted and revised the complete manuscript, and provided technical support. Z.Q.C. supervised the entire research process and contributed to manuscript editing. L.Y.H. contributed to manuscript editing and organized the literature. D.G.L. and L.R.C. were responsible for data collection and partial data curation. All authors read and approved the final manuscript.

Competing interests

The Authors declare no Competing Financial or Non-Financial Interests.