npj | heritage science

Article

https://doi.org/10.1038/s40494-025-02275-9

APLDiff: an adaptive perception-driven lightweight diffusion framework for digital mural inpainting

Check for updates

Dong Zhao[1,2,3,4,5], Xin Chen[1,2,3,5], Dan Zhang[1,2,3,4] ✉ & Jing Li[1,2,3]

Mural degradation presents significant challenges to cultural heritage preservation. To address this, a hierarchical mural inpainting model, APLDiff, based on a lightweight diffusion model, is proposed. A physics-based degradation simulation is introduced, which simulates real damage patterns by modeling material aging and environmental factors, thereby enhancing the model's generalization ability. An efficient diffusion network is constructed, with parameters reduced by 83% compare to the original Diffusion model, and an adaptive perception weight mechanism is incorporated to alleviate quality loss caused by model compression. The two-stage multi-scale sampling strategy allows for coarse structure restoration at low resolution, followed by high-fidelity detail enhancement in the latent space. These innovations provide a scientific foundation and practical solution for the digital inpainting of mural heritage, improving inference efficiency while maintaining visual authenticity.

Murals are ancient paintings that carry historical information and artistic value. They are valuable ancient artworks with historical, artistic, and scientific significance[1]. Murals are not only carriers of esthetic appreciation but also mirrors of social culture and civilization. These works often use materials such as rock walls, fabric, and silk, combined with natural mineral and plant pigments (such as cinnabar, azurite, and gamboge). However, with the passage of time, murals face significant preservation challenges due to the inherent properties of the materials and environmental erosion. On the other hand, modern and contemporary art uses disposable materials for the protection of cultural heritage, and these materials are considered not to last for a long time[2]. Disposable materials commonly used in contemporary art are not designed with long-term preservation in mind. They often degrade and lose their original appearance within a few years or, at most, a few decades due to degradation and fading. The short-term durability of disposable materials highlights the urgent value of preserving traditional mural materials, emphasizing the need for rescue efforts in the face of their inherent fragility.

Murals with different carriers face various damages: the paint layer of canvas murals is easy to peel off due to aging and shrinkage stress, and the carrier is also affected by temperature, humidity, and microorganisms. The paint of rock wall is peeled off due to light oxidation, and the wall cracking is caused by building displacement and thermal stress. These damages not only destroy their physical form but also threaten the integrity of art and cultural information.

With the enhancement of national cultural confidence and the improvement of public cultural awareness, the protection and restoration of murals has become the focus of cultural heritage research. The traditional manual repair process is complex, the threshold is high, the labor is time-consuming, and the experience is dependent, while the digital image repair technology (especially the breakthrough of deep learning) brings new opportunities for mural inpainting: it can deeply analyze the color texture of the mural, improve the efficiency of damage identification and reconstruction, shorten the repair cycle, and assist in the completion of high-precision detail repair, which injects technical impetus into the protection of cultural heritage.

Digital mural inpainting is essentially a task within the domain of image inpainting, initially proposed by Marcello Bertalmio et al.[3], with the core objective of algorithmically reconstructing missing, damaged, or occluded regions of an image in a way that ensures both semantic coherence and visual consistency with the original content. Current mainstream techniques include approaches based on generative adversarial networks (GANs)[4–7], variational autoencoders[8–10], and diffusion models[11–15]. Among them, diffusion models simulate a gradual noise corruption and reverse denoising process, enabling the generation of high-fidelity, detail-rich image content. These models are particularly effective in handling complex scenes and large missing regions due to their strong capacity for structural modeling and semantic reasoning. Compared to GANs, diffusion models offer more stable training and are less prone to mode collapse, but they typically

[1]School of Computer, Qinghai Normal University, Xining, China. [2]The State Key Laboratory of Tibetan Intelligence, Xining, China. [3]Research Base for Intangible Cultural Heritage of Qinghai Province, Xining, China. [4]Academy of Plateau Science and Sustainability, Xining, China. [5]These authors contributed equally: Dong Zhao, Xin Chen. ✉e-mail: danz@mail.bnu.edu.cn

require thousands of iterations to produce a single sample, resulting in high computational costs that limit their practical deployment.

Mural image inpainting. In recent years, the rapid development of digital technology and deep learning has brought revolutionary breakthroughs to the inpainting of mural images. Various neural network-based methods have been applied to mural inpainting tasks, significantly improving both inpainting efficiency and image quality. For example, ref. 16 proposed a mural virtual inpainting network combining global-local feature extraction and structural information guidance, which predicted the structure and coarse content of the missing area through the structure generator. Then, the content generator fused the global-local features of the BranchBlock module and the FFC convolution for fine repair and obtained an excellent repair effect. Ref. 17 uses an improved two-stage GAN to reduce the loss of feature information in the convolution process through the feature optimization fusion strategy, and uses the hole residual module to replace the hole convolution to increase the receptive field and reduce grid artifacts to realize the inpainting of murals. Ref. 18 proposed an ancient mural inpainting method based on improved GAN (consistency enhanced GAN). By combining global and local discriminators, dilated convolution, and a two-stage training strategy, the peak signal-to-noise ratio (SNR) and structural similarity (Structural Similarity Index (SSIM)) of mural inpainting were significantly improved in complex texture and large area missing scenes. Ref. 19 proposed a GAN model based on a dual attention mechanism and an improved generator. By fusing multi-scale features and piecewise loss optimization, it effectively solved the repair problem of complex diseases such as cracks and peelings caused by environmental erosion and artificial damage of tomb murals, and improved the accuracy and color coordination of repaired structures. Ref. 20 proposed a color inpainting method for Dunhuang murals based on a reversible residual network. Through automatic reference image selection, channel redundant information elimination, and an unbiased color transfer module, the color inpainting effect was significantly improved while maintaining the structural and texture integrity of the murals.

Image inpainting based on a diffusion model. Diffusion model[21] has attracted extensive attention in the field of image inpainting due to its powerful generation ability and flexibility. In the early stage, diffusion models were mainly used for unconditional image generation, and their potential in conditional generation tasks has been gradually exploited in recent years.

RePaint[13] first proposed using pre-trained unconditional DDPM for image inpainting. By introducing a mask condition in the reverse diffusion process, the proposed method directly uses the information of the known region for sampling, without training for a specific mask distribution. Experiments show that RePaint[13] performs well under extreme masks (such as sparse line masks and large area missing), but limited by the Markov chain structure of DDPM, it needs thousands of steps of iteration, resulting in high computational cost.

In order to improve reasoning efficiency, researchers have made improvements in many directions. The DDIM proposed by Song et al.[22] realizes hop-step sampling by constructing a non-Markov process, and reduces the number of inference steps to 50-250 while maintaining the generation quality, which lays a foundation for subsequent acceleration research. Rombach et al.'s[23]. Latent diffusion models migrate the diffusion process to the latent space, reduce the computational complexity while maintaining the high-resolution generation ability, and provide a new idea for processing large-sized images. Liu et al.[24] introduced the numerical integration method into the DDPM sampling process to reduce the number of iterations. The number of sampling steps is reduced to less than 50 while maintaining the quality of the generation. Salimans et al.[25] proposed a progressive distillation method, which iteratively distills the deterministic sampling process of the pre-trained diffusion model into a student model with half the number of steps, achieving an order-of-magnitude improvement in the sampling speed of the diffusion model.

In order to enhance the controllability of condition generation, researchers have proposed a variety of improvement strategies. Choi et al.[15] proposed P2 weighting scheme that redesigns the training objective function, maintains the generation quality through noise level-aware weighting when reducing model parameters, and provides theoretical support for lightweight diffusion model design. Zhang et al.[14] proposed the Copaint algorithm to solve the incoherence problem of existing diffusion inpainting methods through Bayesian joint optimization and stepwise error correction. DiffIR[26] proposed an efficient diffusion model, which extracted and fused image priors through compact prior coding + dynamic converter, combined with two-stage training (pre-trained reconstruction network + lightweight diffusion estimation), to achieve SOTA in super-resolution, deblurring, and other tasks, with extremely low computation and greatly improved reasoning speed. Meng et al.[27] proposed an image compositing and editing framework based on stochastic differential equations, which enables an efficient and controllable generation and editing process through a gradient-guided mechanism and an adaptive solver design. Nichol et al.[28] proposed a text-guided image generation and editing framework based on a diffusion model, which achieves high-fidelity and semantically controllable image synthesis by directly integrating text encoding into the conditional generation mechanism of the diffusion process.

Applications of diffusion models to other fields. Diffusion models have shown great potential in industry and healthcare: in industry, they have significantly improved production efficiency and product quality by generating optimal designs (e.g., chip layouts and molecular structures of new materials[29]), performing high-precision anomaly detection (e.g., product surface defects and equipment failure prediction[30]), and visual anomaly detection[31]. In the medical field, its applications include medical image synthesis and enhancement[32] (solving the problem of data scarcity), accurate segmentation and registration[33] (assisting disease diagnosis), generating the structure of novel molecules[34], and accelerating drug research and development. It provides a powerful tool for precision medicine, diagnosis and treatment automation, and new drug discovery.

Moreover, most existing inpainting methods rely on randomly generated masks to simulate missing regions. Although easy to implement, such masks fail to accurately reflect the physical characteristics of real mural damage, such as directional cracks from material aging, mold-induced spotting, gradual pigment peeling, or wrinkles caused by human interference. These unrealistic masks lack structural regularity and contextual correlation, leading to suboptimal inpainting performance when applied to real-world mural inpainting tasks.

To address these challenges, this paper proposes a comprehensive technical framework for digital mural inpainting, termed APLDiff. The main contributions of this paper are as follows: We construct a mural image dataset, DeMUDB, containing 30,000 samples, where damaged murals are generated through a physics-based degradation simulation framework. The dataset covers four typical types of mural damage, including pigment peeling, wrinkles, cracks, and mildew contamination. Compared with traditional artificial or random masks, the simulation process is consistent with material aging and environmental erosion, and the damage pattern is more realistic, which can improve the robustness and generalization ability of the inpainting model in the actual scene; We propose a lightweight diffusion model with an 83% reduction in parameters, integrated with an adaptive perception weighting mechanism to preserve inpainting quality. improve the existing P2 weighting strategy[15]. The strategy dynamically adjusts the perceptual loss weight—focusing on mid-frequency semantic structures in early training and enhancing high-frequency details in later stages—achieving a balance between structural coherence and texture fidelity. The resulting model is both computationally efficient and stable, making it well-suited for deployment in resource-constrained environments and scalable inpainting of high-resolution mural images. We introduce a two-stage multi-scale diffusion sampling strategy. It reconstructs structure at low resolution and refines details in a 256×256 latent space. This method enhances global semantics and local textures while reducing computation. Compared to single-stage sampling, it significantly improves both speed and quality, especially for high-resolution mural inpainting.

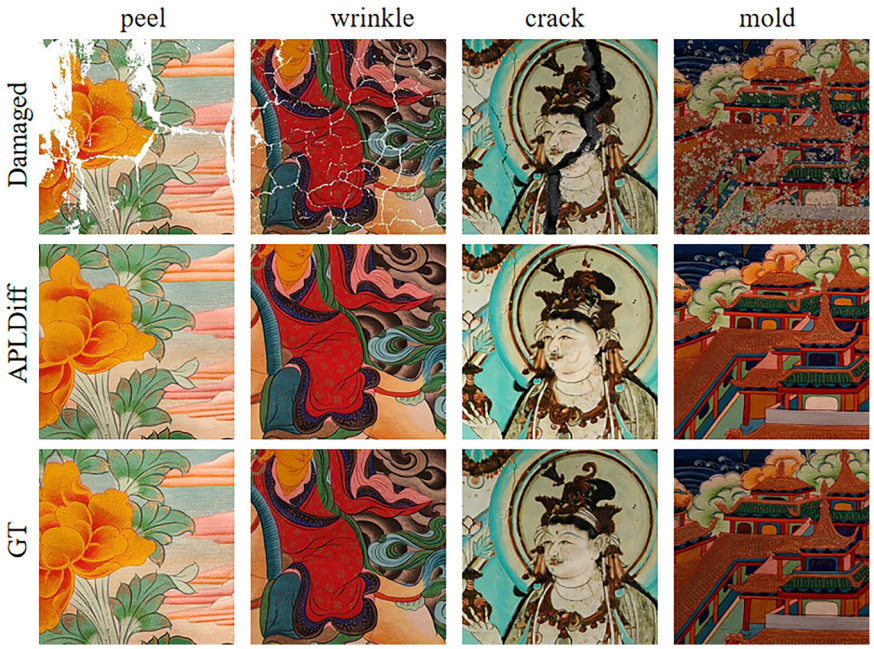

Fig. 1 | Visual examples of digital mural inpainting using the proposed APLDiff framework. Four types of simulated damage are shown—pigment peeling, canvas wrinkling, wall cracking, and surface mold—followed by their corresponding inpainting results and ground truth references.

Methods

Murals, after prolonged exposure to environmental factors such as humidity fluctuations, UV radiation, and microbial activity, gradually undergo various forms of physical degradation, including paint peeling, wall cracks, mold growth, and canvas wrinkling. These types of damage are inherently linked to material properties and aging processes, often displaying complex spatial and morphological characteristics. However, commonly used random or synthetic masks in existing image inpainting studies fail to reflect the true patterns of mural deterioration, resulting in limited generalization and stability of inpainting models when applied to real-world inpainting tasks.

To address this gap, this paper introduces a damage simulation approach grounded in physical degradation mechanisms. First, representative real-world damage samples are collected through field surveys and high-resolution image extraction. Then, a spatial mapping relationship is established between the damage samples and the target mural images to ensure that the degradation patterns are accurately aligned in terms of position, scale, and orientation. Specifically, the damage sample is denoted as $S(x,y) \in \mathbb{R}^{H\times W\times 3}$ and the target mural as $T(x,y) \in \mathbb{R}^{H\times W\times 3}$, where H and W represent the height and width of the image in pixels respectively, and the third dimension with size 3 corresponds to the three color channels (red, green, and blue) of the RGB image. A mapping function Ψ is defined to transfer the damaged regions from the sample domain to the target image space, enabling realistic simulation of mural deterioration:

$$\Psi : (x_s, y_s) \in S \rightarrow (x_t, y_t) \in T \tag{1}$$

The mapping rule is $x_t = \frac{x_s}{s_x}, y_t = \frac{y_s}{s_y}$. s_x and s_y are the size scaling factors of the sample and the mural image. When the resolutions of S and T are the same, $s_x = s_y = 1$.

This study focuses on simulating four typical and representative types of mural degradation: Pigment peeling, which replicates missing regions caused by the detachment of aged or damaged pigment layers; Canvas wrinkling, which reflects the creased textures resulting from material deformation or human interference; Wall cracking, which depicts fracture patterns induced by structural stress or climatic fluctuations; Mold spot contamination, which simulates mold stains caused by microbial erosion (Fig. 1).

These damage types encompass the most common forms of deterioration encountered in mural inpainting and possess high realism and modeling value. The specific mapping rules for the four types of damage are detailed below, and the overall simulation process is illustrated in Figs. 2–5.

Simulation of pigment peeling

Define the set of pigment peeling points in damaged sample image S as:

$$\Omega = \left\{(x,y) \in S \mid \exists (G(x,y) > G_{threshold\_high} \vee G(x,y) < G_{threshold\_low}\right\} \tag{2}$$

$G(x,y)$ represents the gray value at this coordinate, $G_{threshold\_high}$ is the high threshold, corresponding to the position with a higher gray value in the damaged sample image S, and $G_{threshold\_low}$ is the low threshold, corresponding to the position with a lower gray value in the damaged sample image S. The pixels that meet the conditions form the pigment peeling point set Ω, and perform grayscale replacement on each peeling pixel in the target image T:

$$T_{peel}(x_t, y_t) = \begin{cases} G(x,y), & if\,(x_t, y_t) \in T = (x,y) \in \Omega \\ T(x_t, y_t), & otherwise \end{cases} \tag{3}$$

Simulation of canvas wrinkling

Based on the characteristics of sample $S \in \mathbb{R}^{H_s\times W_s}$, define a dual-threshold mask function:

$$M_f(x_s, y_s) = \begin{cases} 1, & if\left(S(x_s, y_s) \le \tau_{low}\right) \vee \left(S(x_s, y_s) \ge \tau_{high}\right) \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

τ_{low} is the black stripe decision threshold, and τ_{high} is the white stripe decision threshold. The region with $M_f(x_s, y_s) = 1$ is regarded as the wrinkle region, and the pixel value of each pixel in the wrinkle region is set to 255. The coordinates are mapped to the mural image space by affine mapping $\Psi(x_s, y_s) = (\left\lfloor x_s \cdot \frac{W_t}{W_s} \right\rfloor, \left\lfloor y_s \cdot \frac{H_t}{H_s} \right\rfloor)$ to generate the wrinkled mural

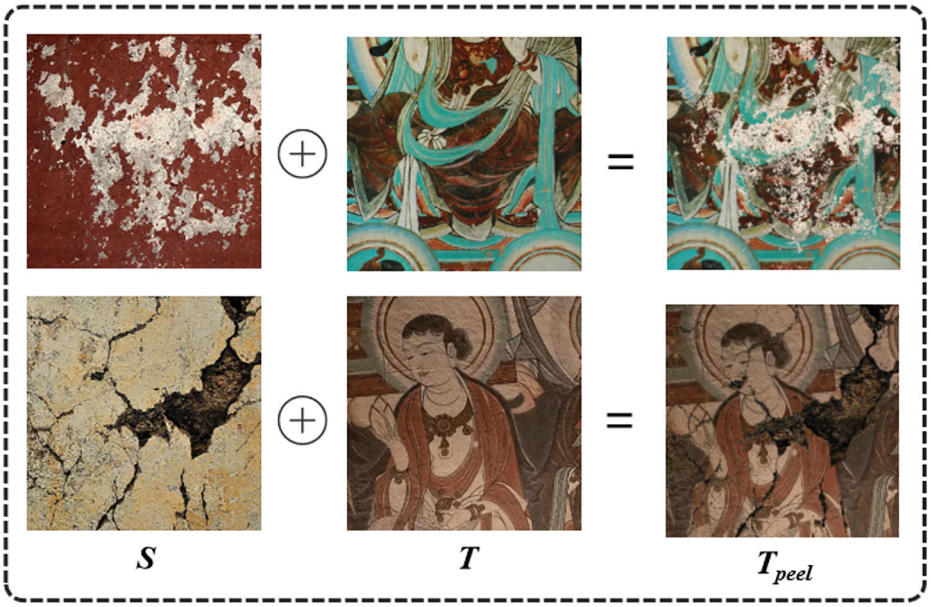

Fig. 2 | Examples of pigment peeling simulation.

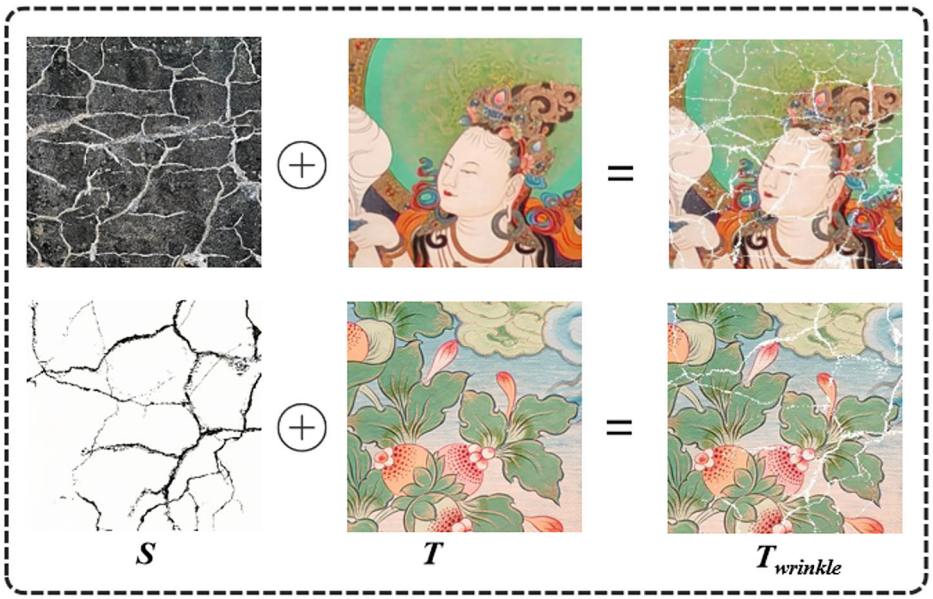

Fig. 3 | Examples of a canvas wrinkling simulation.

image:

$$T_{wrinkle}(x_t, y_t) = \begin{cases} 255, & if\ \exists(x_s, y_s) \in S\, s.t.\, \Psi(x_s, y_s) = (x_t, y_t) \wedge M_f(x_s, y_s) = 1 \\ T(x_t, y_t), & otherwise \end{cases} \tag{5}$$

Simulation of a wall crack

Extracting crack core regions and transition zones using triple-threshold segmentation:

$$M_c(x_s, y_s) = \begin{cases} 1, & if\ I_S(x_s, y_s) \le \tau_{core} (\text{Core crack}) \\ 0.5, & if\ \tau_{core} < I_S(x_s, y_s) \le \tau_{edge} (Transition) \\ 0, & otherwise\,(\text{Non} - \text{crack area}) \end{cases} \tag{6}$$

Among them, τ_{core} is the threshold, $\tau_{edge} = 0.8 * \tau_{core}$. I_S represents the cracked damage sample. The core crack region, transition region, and non-crack region of the cracked damage sample are identified using threshold values τ_{core} and τ_{edge}. Injecting crack features via adaptive blending strategy:

$$T_{crack}(x_t, y_t) = (1 - \lambda) \cdot T + \lambda \cdot \left[I_S(x_s, y_s) \cdot \alpha + \beta \right] \tag{7}$$

Next, pixel fusion is performed between the target image T and the crack sample image I_S to obtain the cracked mural image T_{crack}.

Dynamic weight function:

$$\lambda = \begin{cases} 0.9, M_c(x_s, y_s) = 1 \\ 0.6, M_c(x_s, y_s) = 0.5 \\ 0, \text{otherwise} \end{cases} \tag{8}$$

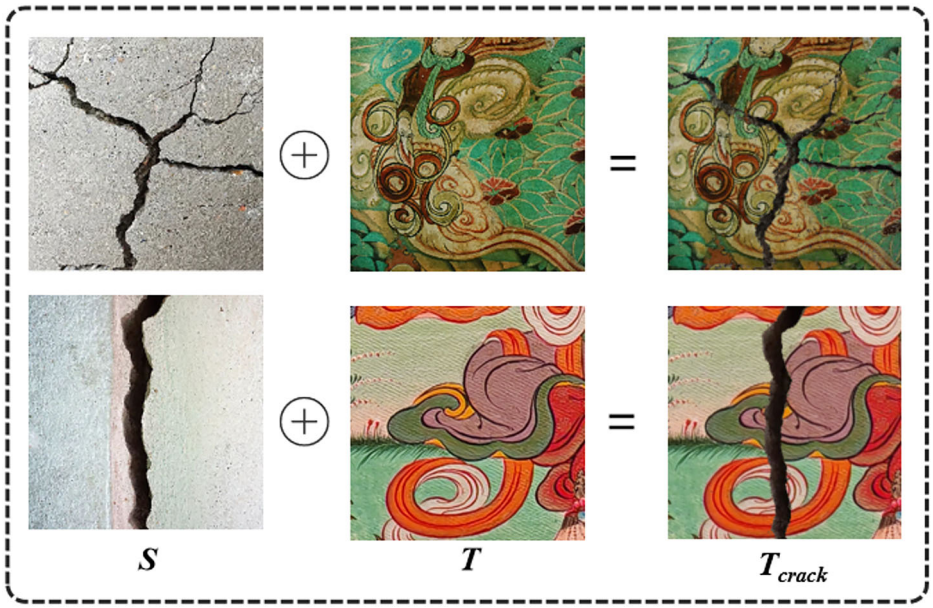

Fig. 4 | Examples of wall cracking simulation.

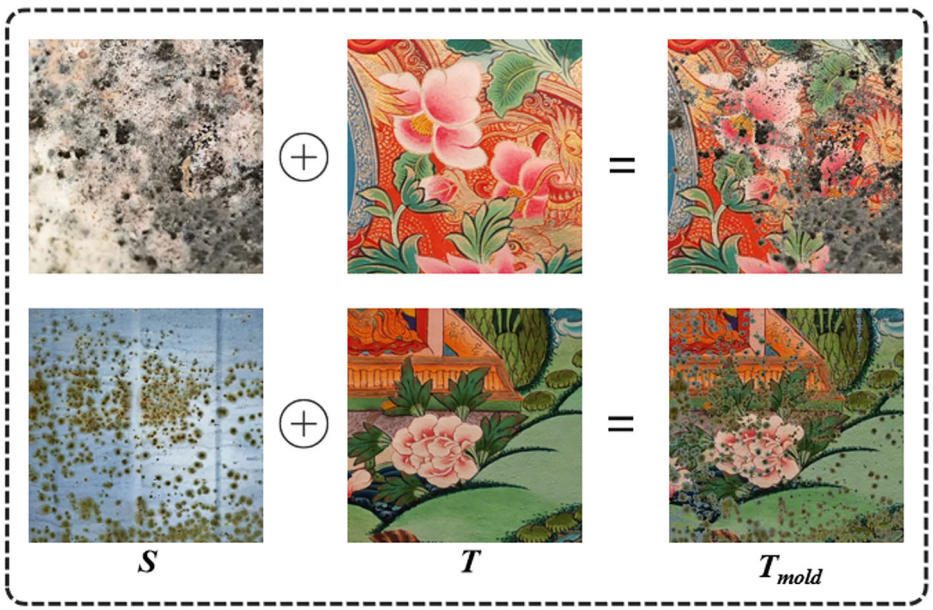

Fig. 5 | Example of mold spot contamination simulation.

In the formula, $\alpha = 0.8$ and $\beta = 10$ are used to adjust the contrast of the cracks and match the optical properties of the mineral pigments in the murals.

Simulation of mold spot contamination

Record the coordinates and pixel values of the mold spots in the mold spot samples:

$$B = \left(x_b, y_b, S_R\left(x_b, y_b\right), S_G\left(x_b, y_b\right), S_B\left(x_b, y_b\right)\right) \tag{9}$$

Where (x_b, y_b) is the mildew point coordinate,S_R,S_G,S_B are the RGB channel values at the corresponding positions. Performing dual-mapping operations on murals:

$$\begin{aligned} &\forall(x_b, y_b, R, G, B) \in B : T_{mold}(\Psi(x_b), \Psi(y_b)) \\ &= (R, G, B) \odot M_{blend} + T(R, G, B) \odot (1 - M_{blend}) \end{aligned} \tag{10}$$

Where $\Psi(\cdot)$ is the coordinate mapping function, $\odot$ represents the element-wise multiplication, and M_{blend} is the color blending weight, which controls the degree of fusion between the mold and the original image, which is calculated as follows:

$$M_{blend} = \frac{\left|\left|S\left(x_b, y_b\right) - \mu_{mold}\right|\right|_2}{255} \cdot \alpha \tag{11}$$

μ_{mold} is the average RGB value of the mildew point, numerator: the Euclidean distance between the current mold spot color and the average mold spot color μ_{mold} (quantifies the degree of color abnormality), denominator 255: normalized to the range [0,1], parameter α controls the migration intensity and adjusts the mold spot color concentration.

Mask data

The mural damage simulation method based on physical degradation characteristics not only simulates four common types of mural degradation

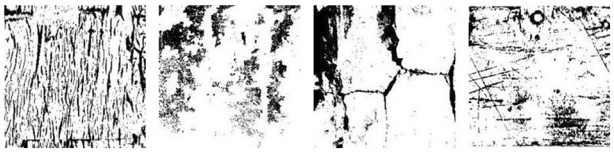

Fig. 6 | Examples of the simulated pigment peeling masks.

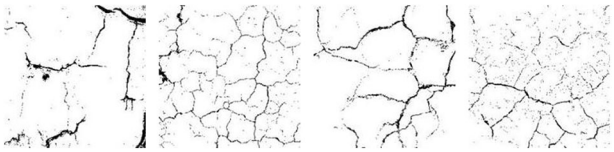

Fig. 7 | Examples of the simulated canvas wrinkling masks.

Fig. 8 | Examples of the simulated wall cracking masks.

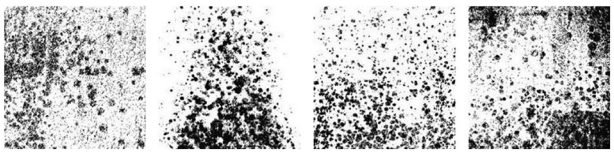

Fig. 9 | Examples of the simulated surface mold masks.

but also generates corresponding masks for each type, as shown in Figs. 6–9. These masks are designed according to the physical degradation mechanisms and accurately reproduce the different damage patterns found in real murals. This provides more representative training and testing data for subsequent inpainting models and effectively addresses the challenge of lacking "clean" ground truth in damaged mural datasets, enabling reliable quantitative evaluation of inpainting results.

Verification of the authenticity of physical degradation simulations

Firstly, due to the challenges of accurately extracting damaged areas from real damaged murals, most of the damaged samples in this study are taken from regions with the same material as the mural but without any patterns. This allows us to use our method to extract the damaged areas. This ensures a high degree of authenticity and consistency when simulating the damage. Furthermore, to validate the reliability of the physical degradation simulation method used and to clarify the subtle differences between real mural damage and simulated samples, we compare our damaged samples with real damaged murals, showing a high degree of similarity in color statistics and crack geometric features. We present some cases in Figs. 10 and 11. (Note: The mask for the real damaged murals is manually annotated, while the mask for the damaged samples is extracted using the method proposed in this study.)

Figure 10 shows the RGB channel color histograms of the mask regions for real damaged murals and damaged samples. Considering that different wall base colors exist, directly comparing color histograms lacks rigor. Therefore, only the color histogram of a wall with a brownish-yellow base color is shown. It can be seen that the color histogram of damaged sample C (the lower half of each group) is quite similar to that of real sample A, and the color histogram of damaged sample D is also similar to that of real sample B, especially in terms of the high mean match in the R and G channels. Moreover, all the figures exhibit the characteristic that "the red channel has the widest distribution, followed by the green channel, and the blue channel is relatively concentrated". This indicates that the damaged samples in this study are able to well simulate the color effects of real damaged murals.

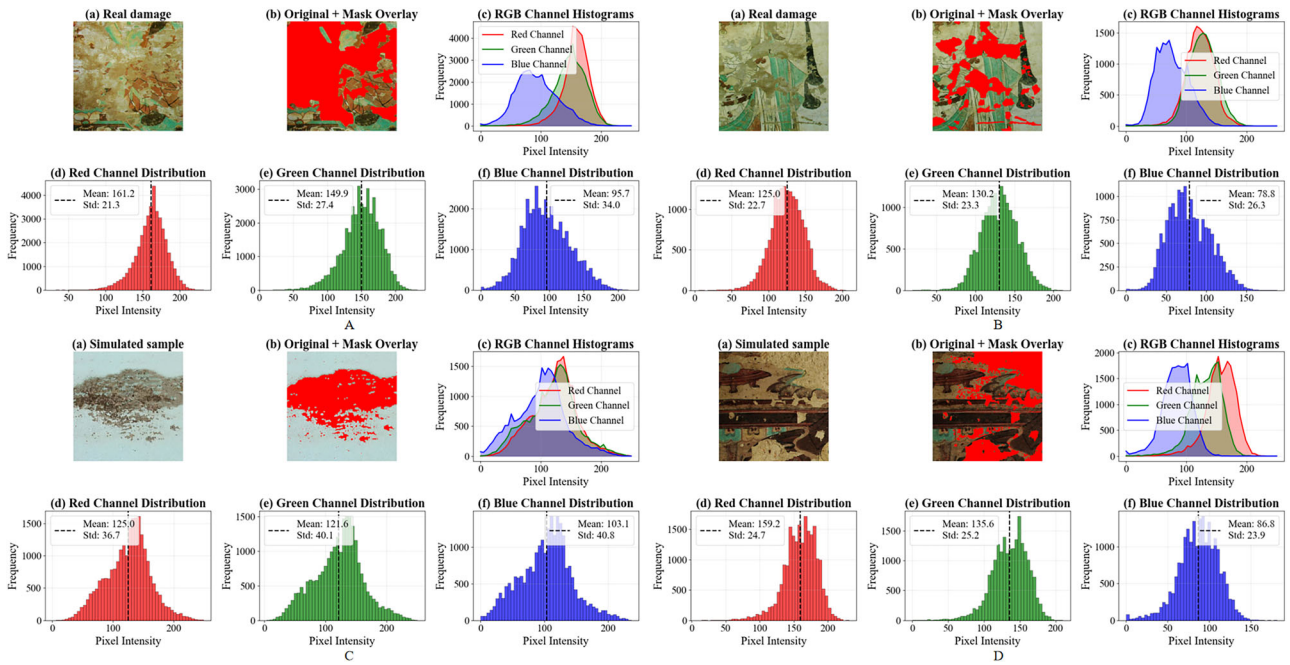

Fig. 10 | Comparison of color distribution features of real damaged murals (A, B) and simulated samples (C, D). Compare the color histogram (**c**) of the damaged area (**a**, **b**) and the descriptions of each channel (**d**–**f**).

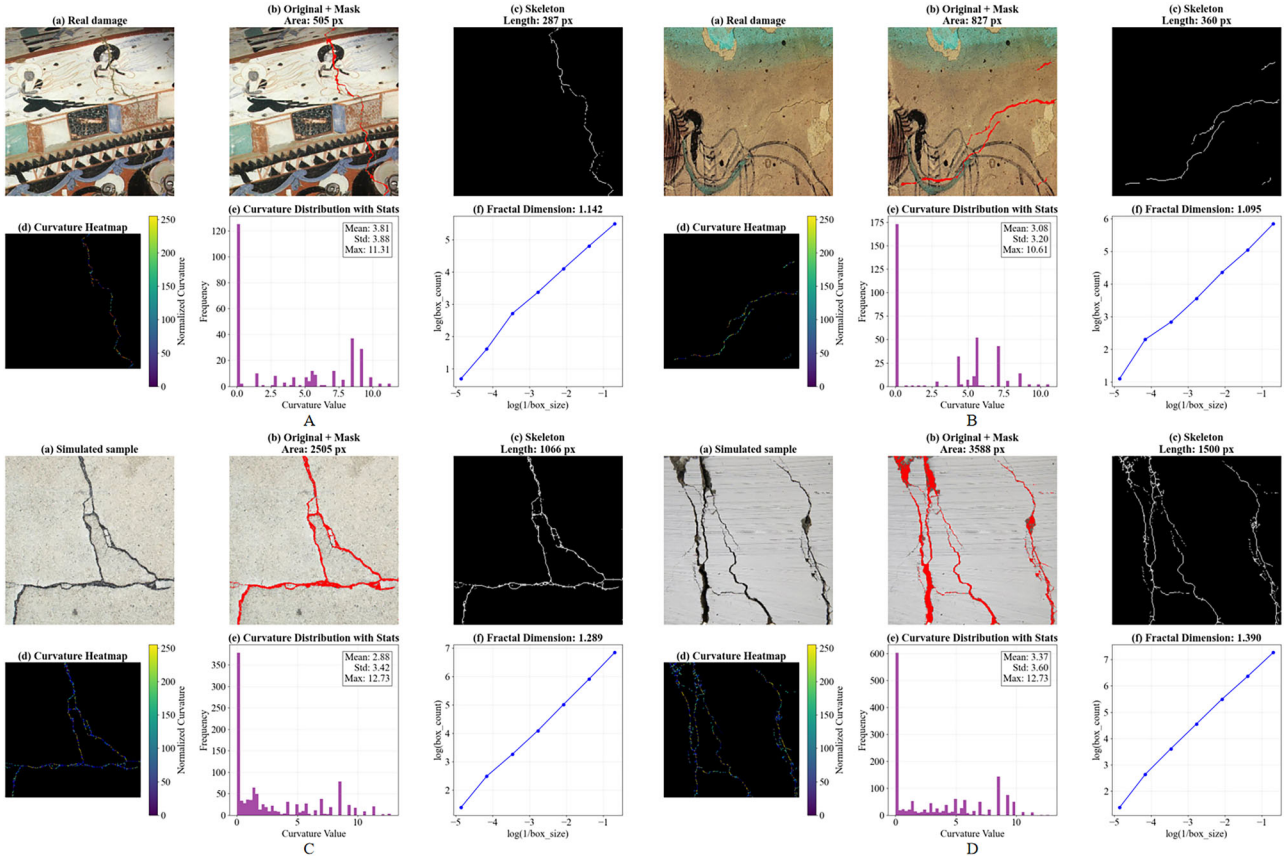

Fig. 11 | Comparison of crack geometric features between real damaged murals (A, B) and simulated samples (C, D). Compare the curvature heatmap (**d**) of the crack area (**a**, **b**), the skeletal length (**c**), the curvature distribution (**e**) and the fractal dimension (**f**).

Figure 11 shows a comparison of the crack geometric features between real damaged murals and simulated damaged samples. Due to significant differences in the area and length of different cracks, the focus of the study is on comparing the curvature and fractal dimension, which better reflect the geometric characteristics of the cracks. Curvature is used to describe the degree of bending of a curve. In Fig. 11d is the curvature heatmap, with the color range from dark (low curvature) to bright yellow (high curvature). The higher the brightness, the sharper the curvature of the crack at that location. Figure 11e is the curvature distribution map, showing the frequency of different curvature values. Figure 11f is the fractal dimension, which is used to measure the roughness and irregularity of complex shapes, with a range from 1.0 (simple straight line) to 2.0 (extremely complex). The average curvature of real samples A and B is 3.81 and 3.08, respectively, while the average curvature of simulated samples C and D is 3.37 and 2.88, respectively, with only minor differences between the two. Additionally, the fractal dimensions also show only small differences. It can be observed that the simulated cracks accurately reproduce the physical characteristics of the real mural cracks in terms of fine curvature features, complexity, self-similarity, and other aspects.

To address damage conditions in mural images such as pigment peeling, wrinkling, cracking, and mold spots, we have designed an accelerated inference process while employing a multi-scale sampling mechanism to ensure both high efficiency and high inpainting fidelity in mural conservation.

Preliminary knowledge

The mathematical foundation of the diffusion model framework can be traced back to the diffusion probabilistic model proposed by Sohl-Dickstein et al.[35]. Diffusion models are a class of generative models based on non-equilibrium thermodynamics, whose core idea involves gradually adding noise to transform the data distribution $p(x_0)$ into a simple Gaussian distribution $\mathcal{N}(0, I)$, making the image at the final time step x_T approach pure noise, and then learning the reverse denoising process to generate data. The forward process, which gradually adds noise, is defined as:

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right) \tag{12}$$

Through the accumulation of noise coefficients $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$ and the reparameterization trick, noise data at any time step x_t can be directly sampled at arbitrary time step t:

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon \tag{13}$$

Where $\epsilon \sim N(0, I)$.

The reverse process gradually recovers pure noise data x_T into real images x_0 by learning a conditional probability $p_\theta(x_{t-1}|x_t)$ and predicting noise through a parameterized denoising network $\epsilon_\theta(x_t, t)$, ultimately generating sample reconstruction data through iterative denoising. The mathematical expression is as follows:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}\left(x_{t-1}; \mu_\theta(x_t, t), \Sigma_t I\right) \tag{14}$$

$$x_{t-1} = \frac{1}{\sqrt{1-\beta_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right) + \sigma_t z, z \sim \mathcal{N}(0, I) \tag{15}$$

The goal of the diffusion model is to obtain the real as accurately as possible. The objective function aims to minimize the mean squared error

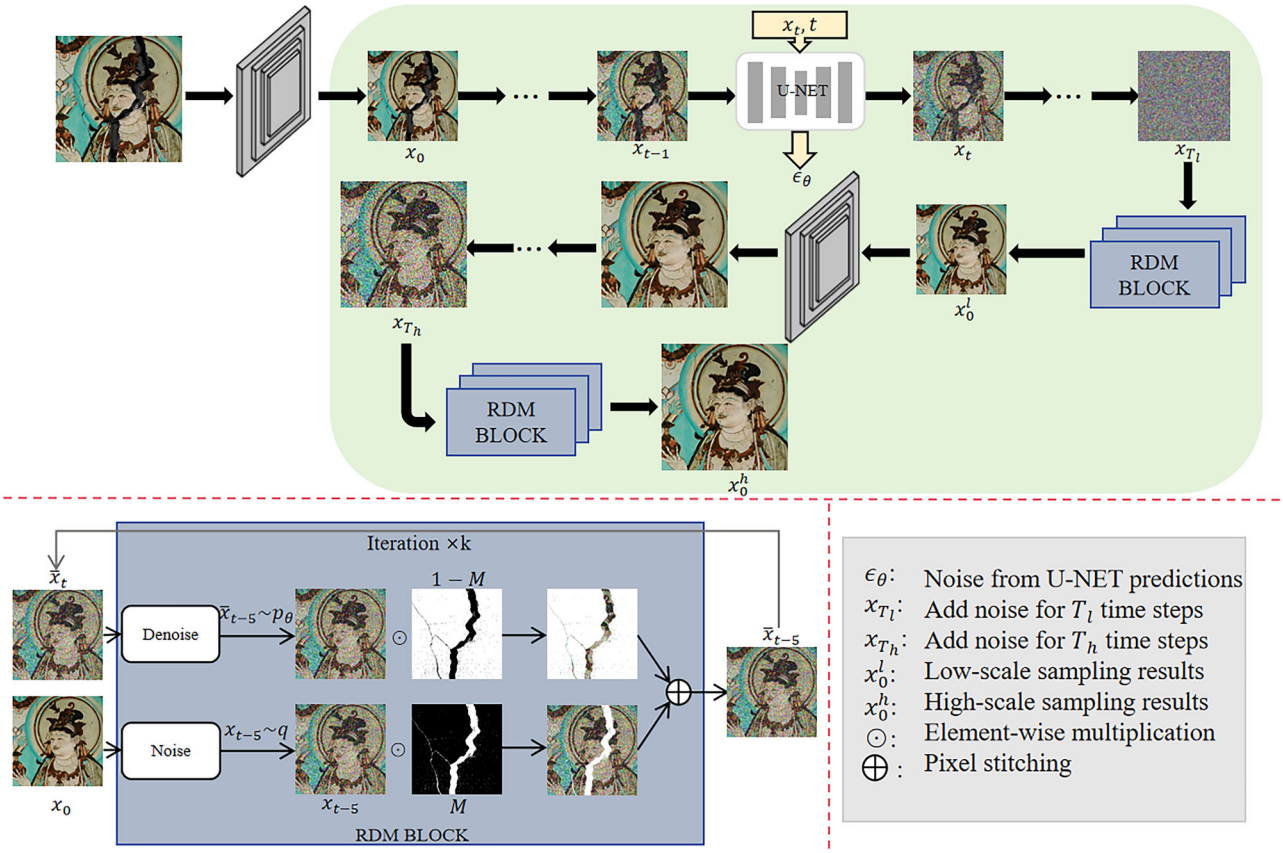

Fig. 12 | Schematic diagram of mural inpainting by the model.

(MSE) between the predicted noise ϵ_θ and the true noise ϵ:

$$L_{simple} = \sum_{t=1}^{T} \mathbb{E}_{x_0,\epsilon}\left[\left|\left|\epsilon - \epsilon_\theta(x_t, t)\right|\right|^2\right] \tag{16}$$

This objective implicitly assigns a weight of $\lambda_t = \frac{(1-\beta_t)(1-\alpha_t)}{\beta_t}$ to the loss terms for different noise levels t.

Choi et al.[15] noticed that the traditional weighting scheme λ_t does not distinguish between high, medium, and low SNR(t) stages, which may cause the model to excessively focus on detail repair and ignore key semantic information. A perceptually first weighting scheme is also proposed to adjust the loss weights to emphasize medium SNR(t) tasks. Weight λ_t' is:

$$\lambda_t' = \frac{\lambda_t}{(k + \mathrm{SNR(t)})^\gamma} \tag{17}$$

Free-form image inpainting: RePaint[13] was the first to apply DDPM to image inpainting, using a pre-trained unconditional diffusion model as a generative prior, enabling free-form inpainting without the need for specific mask fine-tuning. In the reverse diffusion time step t, the noise state of the known region (M) is first calculated, then denoising is performed to predict the content of the masked region ($1 - M$), and finally, the result is completed through pixel-level stitching.

$$x_{t-1} = M \odot x_t^{known} + (1 - M) \odot x_t^{unknown} \tag{18}$$

Adaptive perception weighted training

In ref. 15, Choi et al. divided the noise level into three stages: coarse-grain stage, content stage, and clean stage, corresponding to high noise, medium noise, and low noise, respectively. The noise level is indexed by time step t, where t ranges from 1 to T (T is the total number of time steps). The noise level is closely related to the SNR, which describes the ratio between the signal (that is, the image content) and the noise. The higher the SNR is, the clearer the image content is. The lower the SNR, the more seriously the image content is corrupted by noise. High SNR clean stage (early diffusion): $(k + \mathrm{SNR(t)})^\gamma$ value in Eq. (17) is high, λ_t' is significantly suppressed, and the model allocates minimal weight to learn "imperceptible detail inpainting"; Medium SNR content stage (mid-diffusion): $(k + \mathrm{SNR(t)})^\gamma$ value is moderate, λ_t' remains at a high level, and the model focuses on learning "perceptible content restoration" (e.g., object structure, color consistency)—this is the noise weighting range that is most critical to generation quality; Low SNR coarse-grain stage (late diffusion): $(k + \mathrm{SNR(t)})^\gamma$ value is low, λ_t' maintains reasonable weights, and the model learns global coarse features.

When training the diffusion model, the model needs to learn to recover the original image from noisy images with different noise levels. The traditional training objective is to uniformly weight the loss across all noise levels, but this approach may not fully exploit the learning potential of each noise level. P2 weighting[15] provides a good inductive bias for learning rich visual concepts by lifting the weights of the coarse-grained and content phases and suppressing the weights of the cleaning phase.

Building upon the perception prioritized training (P2 weighting) framework proposed by Choi et al.[15], we further designed a dynamic γ scheduling scheme to address potential optimization imbalance issues caused by fixed-weight strategies during training. As shown in Eq. (19):

$$\gamma_t = \gamma_{final} + \left(\gamma_{initial} - \gamma_{final}\right) \cdot \max\left(0, 1 - \frac{t}{T_{decay}}\right) \tag{19}$$

Where T_{decay} is a hyperparameter set to $0.5 * T$. $\gamma_{initial} = 1$, $\gamma_{final} = 0.5$.

The core idea of this method is to dynamically adjust the γ parameter in P2 weighting[15] according to the training stage, so that the weight of the model to the content stage (medium SNR) and coarse-grained stage (low

SNR) is further enhanced in the early training ($\gamma \approx 1$), and the suppression of high SNR (low noise stage) tasks is further enhanced. Prioritizing learning semantic content at medium SNR forces the model to prioritize building semantic consistency. The core of P2 weighting is to suppress the weights of the cleaning stage (high SNR), but removing them altogether may lead to noise artifacts. Therefore, as the training progresses, we will gradually decrease the value of γ according to linear decay, gradually reduce the suppression of high SNR tasks, and gradually release the constraint of detail repair tasks. Therefore, the second half of training will strengthen the learning balance between low SNR global features and high SNR detail repair tasks, so that the model optimizes the details while maintaining semantic rationality, and finally realizes the balance between global and local quality. This design does not require additional computational overhead but improves the generation. The adaptive perception weight training is shown in Algorithm 1.

Multi-scale sampling

As shown in Fig. 12, this approach reconstructs the single-stage inpainting into a two-stage inpainting system. In the low-scale processing stage, the mural image is used as input, relying on a multi-dimensional downsampling module that integrates convolution and average pooling. During the downsampling operation, a balance is maintained to preserve features, resulting in a 64×64 image. The downsampled image is then subjected to noise addition for T_l time steps, yielding x_{T_l}, and subsequently sampled from a standard Gaussian distribution to obtain the low-scale result x_0^l. Then, through a multi-dimensional upsampling module that integrates nearest-neighbor interpolation and optional convolution, the image is restored to a 256×256 size, ensuring balanced feature optimization during the upsampling process. In the high-scale stage, the 256×256 image undergoes noise addition for T_h time steps, producing x_{T_h} as the input for the high-scale stage. The final result x_0^h is obtained through the reverse diffusion process to remove noise.

Algorithm 1. Adaptive Perception Weight Training

Require: $p_{\text{data}}(x_0), T, \{\beta_t\}_{t=1}^T, \gamma_{\text{initial}} = 1, \gamma_{\text{final}} = 0.5, T_{\text{decay}} = 0.5T$
Ensure: $\epsilon_\theta(x_t, t)$
$\beta_t = \text{LinearSchedule}(t, T)$
while True **do**
  $x_0 \sim p_{\text{data}}$
  $\epsilon \sim \mathcal{N}(0, \mathbf{I})$
  $t \sim \text{Uniform}\{1, \ldots, T\}$
  $\alpha_t = \prod_{s=1}^{t}(1-\beta_s)$
  $x_t = \sqrt{\alpha_t}x_0 + \sqrt{1-\alpha_t}\epsilon$
  $\gamma_t = \gamma_{final} + (\gamma_{initial} - \gamma_{final}) \cdot \max\left(0, 1 - \frac{t}{T_{\text{decay}}}\right)$
  $\text{SNR}(t) = \frac{\alpha_t}{[}$
  $\lambda_t' = \frac{(1-\beta_t)(1-\alpha_t)}{\beta_t} \cdot \frac{1}{(1+\text{SNR}(t))^{\gamma_t}}$
  $\epsilon_\theta = \epsilon_\theta(x_t, t)$
  $\mathscr{L} = \lambda_t' \cdot ||\epsilon - \epsilon_\theta||^2$
  $\theta \leftarrow \theta - \eta\nabla_\theta\mathscr{L}$
end while
for $t = T, T-1, \ldots, 1$ **do**
  if $t \neq 1$ **then**
    $z \sim \mathcal{N}(0, \mathbf{I})$
  else
    $z = 0$
  end if
  $x_{t-1} = \frac{1}{\sqrt{1-\beta_t}}(x_t - \frac{\beta_t}{\sqrt{1-\alpha_t}}\epsilon_\theta(x_t, t)) + \sigma_t z$
end for

The core of multi-scale sampling is the Reverse Diffusion Module. Through the collaborative design of DDIM sampling[22] and resampling[13], efficient and high-quality image inpainting is achieved. As shown in Fig. 10, the module first performs step-sampling through a lightweight pre-trained model: using the DDIM algorithm with a step size of $s = 5$ for denoising, rapidly propagating the noise latent variable x_t to x_{t-5}. At the same time, a mask fusion technique dynamically combines the known region information with the generated region, completing a 5-step leap in a single iteration, reducing the total number of iterations by 80%. Resampling optimizes detail consistency through a local noise-addition-denoising loop: the current state x_{t-s} is first noise-added for s steps, and then denoised k times back to the original time step. This process is executed eight times in the low-scale stage (64×64 resolution) and 10 times in the high-scale stage (256×256 resolution), focusing on enhancing the boundary transition between the generated and known regions. The two-stage division strategy (low-scale stage steps/high-scale stage steps) combined with resolution-level processing, along with a lightweight diffusion model trained with adaptive perceptual weights (83% reduction in parameters), achieves an inference speed of 12.73 s per image on the mural dataset (RTX 4060Ti 8G), while maintaining competitive inpainting quality with SSIM > 0.897 and Learned Perceptual Image Patch Similarity (LPIPS) < 0.049. Multi-scale sampling is shown in Algorithm 2.

In summary, the low-scale sampling stage operates at low resolution, reducing memory usage and capturing the overall structure by processing low-frequency signals. The high-scale sampling stage supplements high-frequency information, focusing on local details. This approach improves inpainting efficiency while ensuring high inpainting quality.

Algorithm 2. Multi-scale sampling

Require: $I_{\text{damaged}} \in \mathbb{R}^{H \times W \times 3}$
Ensure: I_{restored}
Scale 1 (64×64):
$I_{\text{low}} \leftarrow \text{AvgPool}(\text{Conv}(\text{I}, 3 \times 3), 4)$
$x_{T_l} \leftarrow \sqrt{\bar{\alpha}_{250}}I_{\text{low}} + \sqrt{1-\bar{\alpha}_{250}}\epsilon\{\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)\}$
for $k \leftarrow 1 \text{to} \lfloor 250/5 \rfloor$ do
  DDIM : $x_{t-5} \leftarrow \sqrt{\frac{\bar{\alpha}_{t-5}}{\bar{\alpha}_t}}x_t + \sqrt{\bar{\alpha}_{t-5} - \frac{\bar{\alpha}_{t-5}}{\bar{\alpha}_t}}\epsilon_\theta(x_t, t)$
  $x_{\text{mask}} \leftarrow M \odot I_{\text{known}} + (1-M) \odot x_{t-5}$
  for $r \leftarrow 1$ to 8 **do**
    $x_{\text{noise}} \leftarrow \sqrt{1-\beta_1}x_{\text{mask}} + \sqrt{\beta_1}\epsilon$
    $x_{\text{mask}} \leftarrow \frac{1}{\sqrt{1-\beta_1}}(x_{\text{noise}} - \beta_1\epsilon_\theta(x_{\text{noise}}, t))$
end for
end for
Scale 2 (256×256):
$I_{\text{mid}} \leftarrow \text{Conv}(\text{NNUp}(x_0^{\text{low}}, 4), 5 \times 5)$
$x_{T_n} \leftarrow \sqrt{\bar{\alpha}_{75}}I_{\text{mid}} + \sqrt{1-\bar{\alpha}_{75}}\epsilon$
for $k \leftarrow 1$ to $\lfloor 75/5 \rfloor$ do
  DDIM:(3)
  for r $\leftarrow 1$ to 10 **do**
    (Repeat steps 5-6)
  end for
end for*
$I_{\text{restored}} \leftarrow \mathscr{P}(x_0^{\text{high}})$

Results

Dataset and experimental setup

The mural dataset DeMUDB used in this study is sourced from the Tibetan Culture Museum, the Henan Ancient Mural Museum, the Lanzhou Dunhuang Art Museum, and the Dunhuang Mogao Caves. A total of 2876 original mural images were collected with a resolution of 4096×3072. For training and validation purposes, these high-resolution images were cropped into approximately 30,000 smaller images of size 512×512 and then losslessly scaled down to 256×256. The dataset covers a wide variety of themes, including Buddha statues, religious scenes, flora and fauna, and more. It features complex texture characteristics and fine structural details, providing a comprehensive reflection of the diversity and complexity of murals. To simulate real mural damage scenarios, we manually applied damage to some of the images in the test set. The damage processing is based on the physical degradation mechanism simulation method introduced in the section "Methods". This approach not only preserves the artistic characteristics of the mural images but also effectively evaluates the inpainting

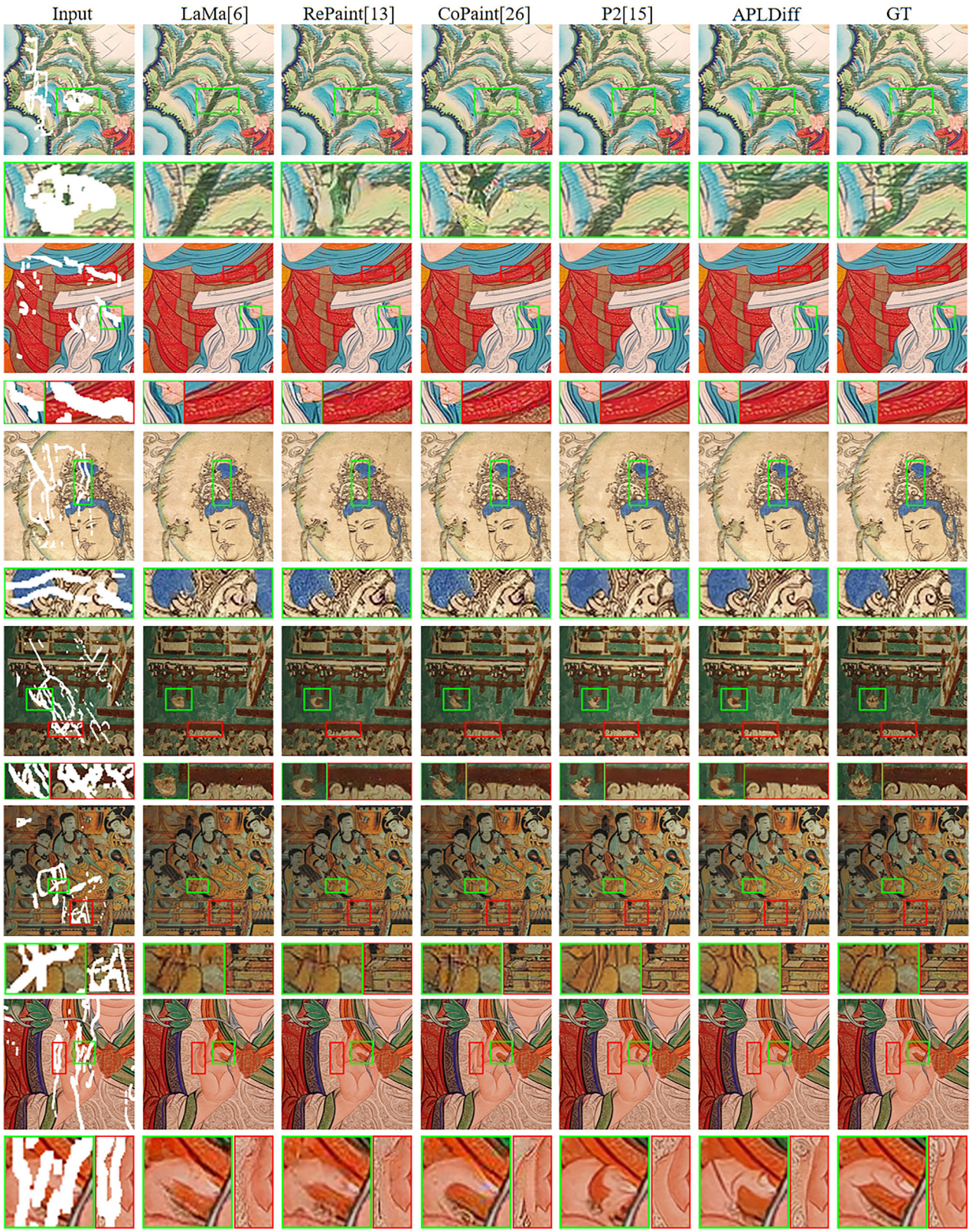

Fig. 13 | Mural inpainting results for each model in random masks.

capability and robustness of the proposed algorithm under various types of damage.

A total of 27,000 images were selected from the DeMUDB dataset as the training set for 500,000 iterations of inpainting training. while the remaining 3000 images were used as a test set for algorithm performance validation and evaluation. In this study, a systematic efficiency evaluation of five image inpainting algorithms was conducted in an NVIDIA RTX 4060Ti (8 GB) GPU environment.

This study comprehensively evaluates the performance of different methods in mural inpainting tasks through four systematic experiments.

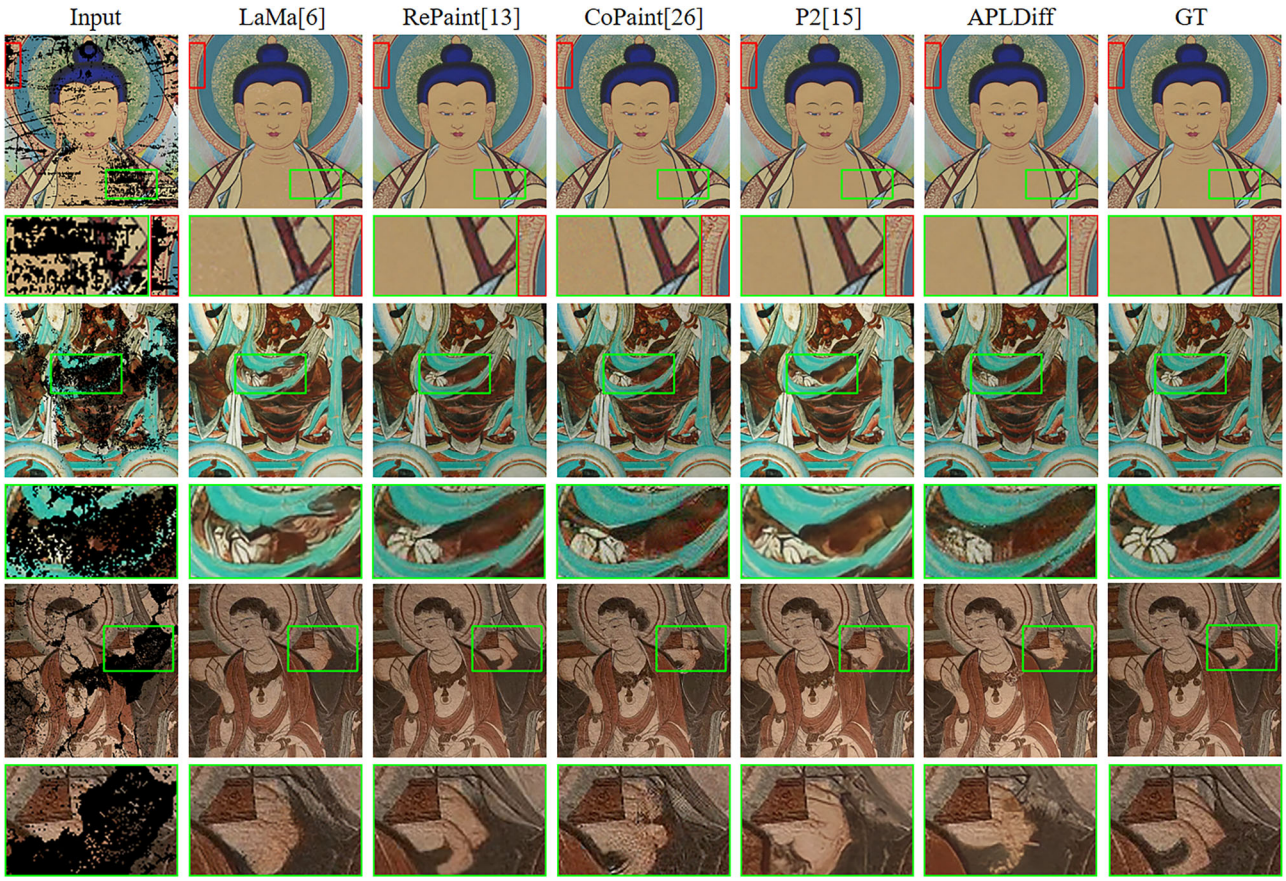

Fig. 14 | Qualitative comparison of the inpainting results on the masks with pigment peeling.

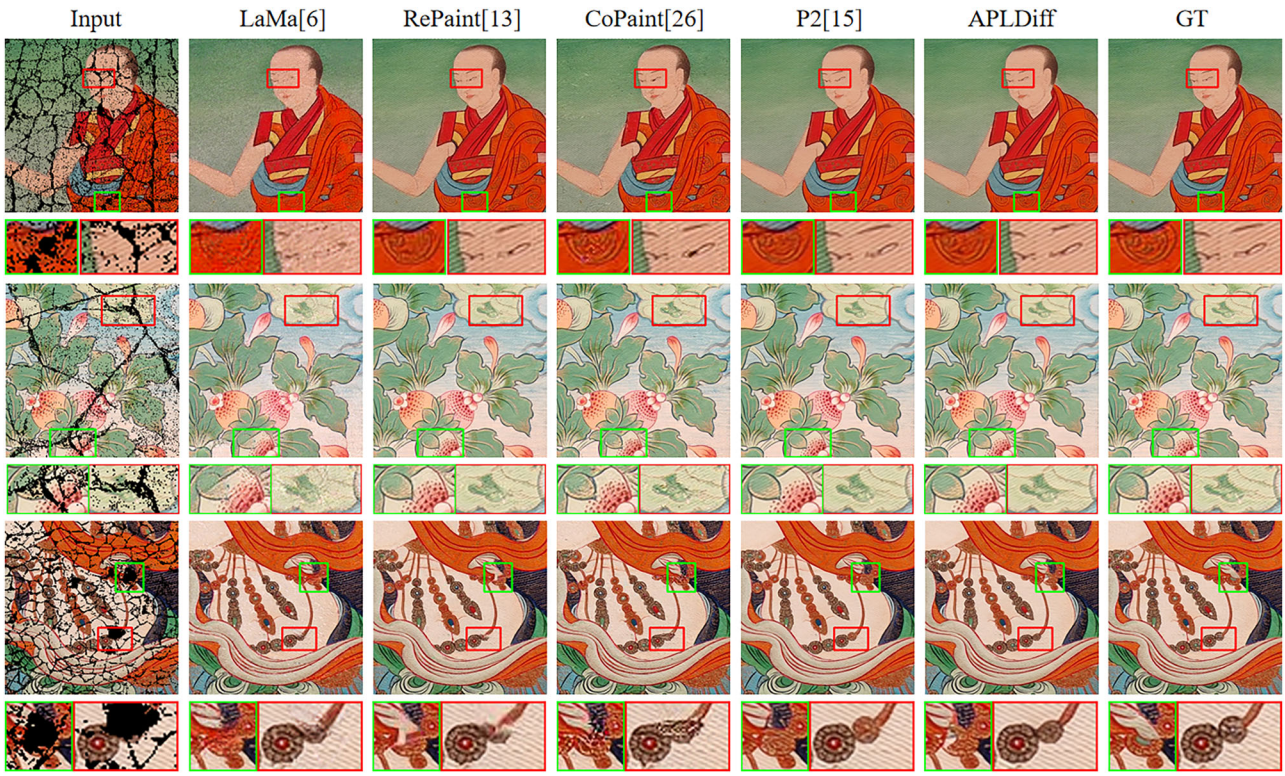

Fig. 15 | Qualitative comparison of the inpainting results on the masks of the canvas wrinkle.

The experimental setups are as follows: The input is the complete original image with a randomly generated mask, used to test the algorithm's ability to inpainting randomly missing areas; The input is the original image with simulated masks for four types of damage, to assess the algorithm's adaptability to different types of damage; The input is a simulated damaged mural with a simulated mask, to verify the algorithm's inpainting performance in damaged scenarios; The input is a real damaged mural with its corresponding mask, to test the algorithm's performance in real-world applications. For the first three experimental setups, we conducted both qualitative and quantitative comparison analyses. Since the fourth experiment lacks real reference images, we adopted two evaluation schemes: first, we invited 30 volunteers to subjectively rate the inpainting results (on a scale of 1–10), and secondly, we used no-reference image quality assessment (NR-IQA) methods for objective quantitative comparison, thus providing a comprehensive evaluation of the actual inpainting performance of each method.

Evaluation metrics

In the evaluation of mural image inpainting tasks, we use SSIM[36], LPIPS[37], Universal Quality Index (UQI)[38], Gradient Magnitude Similarity Deviation (GMSD)[39], and Gray Level Co-occurrence Matrix Correlation Difference (GLCM_Correlation_Diff)[40] to construct a multi-dimensional quantitative evaluation system, comprehensively assessing the inpainting results from perspectives such as structural fidelity, perceptual consistency, and statistical regularity.

SSIM[36] is based on modeling brightness, contrast, and structural similarity, making it suitable for detecting the coherence between the inpainting area and the original mural in terms of overall structure. LPIPS[37], based on deep neural networks, extracts high-level semantic features and captures detail differences sensitive to human vision. It is ideal for evaluating the fusion of local textures, colors, and edges after mural inpainting, especially sensitive to style inconsistencies in the inpainting region. UQI[38] combines brightness and structural information to achieve lightweight computation for rapid global quality screening, making it suitable for analyzing the balance of multi-region inpainting effects in murals. GMSD[39] evaluates distortion by calculating image gradient differences, used to detect edge sharpness and detail retention in mural inpainting. GLCM_Correlation_Diff[40] calculates the difference in gray-level co-occurrence matrix correlation features between two images, reflecting changes in texture regularity after inpainting. The smaller the value, the higher the correlation, indicating stronger linear dependency of textures and better preservation of the mural's original graininess.

Effectiveness of our methods

For the types of mural damage, we compare the inpainting effects of LaMa[6], RePaint[13], CoPaint[14], P2[15], and the adaptive perception weighted training method proposed in this paper on the dataset constructed in this paper (Note that P2[15] is not designed for image inpainting tasks but for image generation. We can achieve image inpainting by incorporating the weight file obtained from training the P2 model on our mural dataset into our inpainting method. Figure 13 shows the inpainting results of different models on random masks. Figures 14–17 demonstrate the inpainting results for four representative damage types. Other methods have been trained on our dataset.

This is shown in Fig. 13. In most scenarios, LaMa[6] can closely match the original image style and perform reasonable inpainting; however, in areas with complex textures, it occasionally appears stiff and blurry. In the second row, RePaint[13] and CoPaint[14], while completing the structure, show blurry details in the ribbon and noticeable color spots. P2[15] and Ours achieve better inpainting results. In the fourth row, within the green box, the inpainting results of LaMa[6], CoPaint[14], and P2[15] are severely distorted. In the sixth row, the inpainting results for the fingers from each model are quite blurry, while Ours' result is relatively clear.

Figure 14 shows the visual inpainting results of different models in the case of pigment peeling and damage. Through observation and comparison, it can be seen that LaMa[6] has a smooth inpainting effect and insufficient detail reduction. The inpainting effect in the second and third rows is also relatively simple in texture and color, and does not fully restore the original details. CoPaint[14] and P2[15] have shortcomings in details and overall coordination. Compared with GT, the texture accuracy of the second row of clothing Wrinkles and the area near the shoulder of the third row is not good. RePaint[13] and Ours perform well with other models in detail processing, such as a high degree of detail reduction of character mouth and clothing texture, which makes the repaired image more realistic.

Figure 15 shows the visual results of different models for the inpainting of canvas wrinkles. The comparison shows that LaMa[6] has a relatively basic performance in detail and texture inpainting, and the repaired images are blurred and lack realism. CoPaint[14] is relatively insufficient in detail processing, such as the eyes of characters, the details of ornaments, and the fineness of plant patterns. RePaint[13] in the third row, the degree of inpainting of ornaments is not as good as P2[15] and Ours. In terms of coherence of details, P2[15] and Ours are superior to other models, and the visual effect is more natural.

Figure 16 presents visual comparisons of inpainting results for wall crack scenarios across different models. The results demonstrate that LaMa[6], CoPaint[14], and P2[15] exhibit noticeable gaps in detail accuracy and color consistency compared to GT. Specifically, the first row shows inadequate pattern details, the second row reveals suboptimal contour and accessory inpainting for human figures, while the third row displays unnatural color matching, collectively resulting in compromised realism. In contrast, both RePaint[13] and our method achieve superior performance, closely approximating ground truth in texture reproduction and color fidelity. The complex patterns in the first row and smooth color transitions in the third row demonstrate exceptional consistency with authentic images.

Figure 17 presents the visual comparison of inpainting results under mold contamination scenarios. The first row demonstrates that LaMa[6] and CoPaint[14] exhibit noticeable color block artifacts and blurring effects. In the second row's vegetation inpainting around animals, CoPaint[14], P2[15], and our method show texture discrepancies compared to GT, while RePaint[13] achieves higher fidelity. For architectural and botanical details in the third row, our method outperforms others in both detail accuracy and overall visual coherence.

To validate the model's inpainting capability for damaged murals, we use the simulated damaged murals and corresponding masks as input. The resulting inpainting outcomes are then used for quantitative evaluation rather than relying on subjective human opinions to determine the quality of the inpainting. This is possible because the simulated damaged murals have ground truth. Figure 18 presents a qualitative comparison of the inpainting results of different models on the simulated damaged murals.

LaMa[6] has a decent overall structure inpainting, but the color transitions are somewhat stiff (as seen in the blocky color phenomenon in the floral area in row 4). When inpainting with RePaint[13], it does a good job of preserving the original colors and textures, but some unnatural inpainting artifacts remain in certain areas (such as in rows 1 and 6). The inpainting results of CoPaint[14] in rows 1, 5, and 6 are somewhat average. P2[15] shows noticeable differences in texture and color compared to the original image (as seen in row 3). APLDiff restores complex details such as cracks in the clothing in row 3 and paint peeling in row 6, with fine accuracy.

In this study, for each type of damage to the mural, we compared each original image with the restored image and presented the results, respectively, to analyze the performance of each model in different damage scenarios more accurately. SSIM is used to evaluate the similarity of image structure, LPIPS to evaluate the similarity of perception, UQI to evaluate the overall quality of the image, GMSD to evaluate the edges and details of the image, and GLCM_Correlation_Diff to evaluate the correlation of image texture. The detailed results are shown in Table 1.

From Table 1, it can be seen that each model presents different characteristics in the effect of mural inpainting. RePaint[13] has a relatively outstanding overall performance. In various key indicators, it has excellent performance against multiple damage types. (Since P2[15] and Ours use

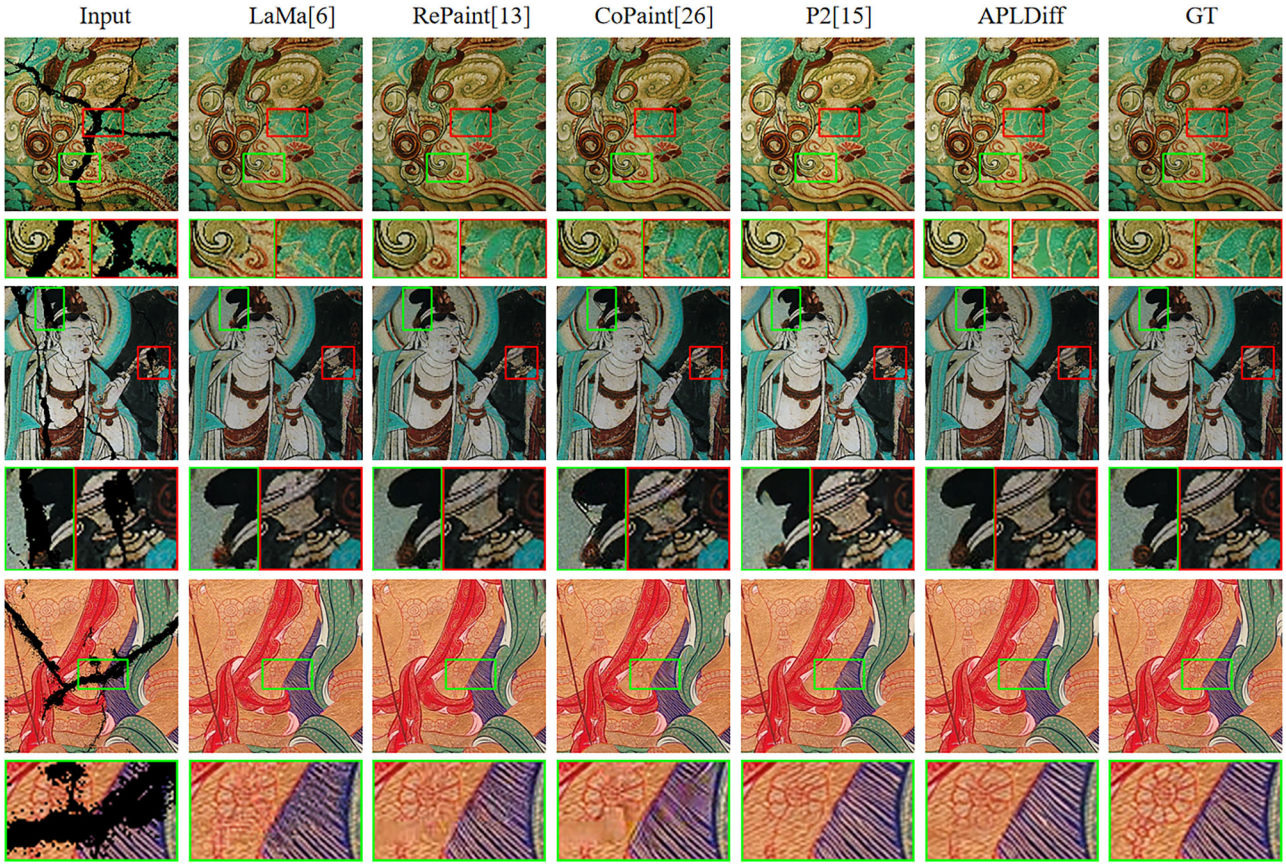

Fig. 16 | Qualitative comparison of the inpainting results on the masks of wall crack.

lightweight models, the model parameters have been reduced by 83%, so ours did not achieve the most outstanding score). Although Ours did not achieve the most outstanding results, it ranked second in the vast majority of scores. The restored images were highly consistent with the original images in terms of structural similarity, human perception similarity, comprehensive quality, edge details, and texture correlation. LaMa[6] and CoPaint[14] performed relatively stably, and the values of each index were mostly at a medium level. The various indicators of P2[15] are relatively ordinary and do not show outstanding advantages in different damage types, which also reflects the effectiveness of the adaptive perception weight training we proposed.

Lightweight of the model

As shown in Table 2, our model has the minimum number of parameters and model size. Fig. 19 illustrates the efficiency distribution of different models for inpainting 100 images, which can be seen the LaMa[6] algorithm, based on fast Fourier convolution, shows an outstanding real-time advantage, and its processing time stably maintains 0.31 s, which is 2–5 orders of magnitude faster than other methods. However, the performance of the algorithm is not good in GLCM_Correlation_Diff, SSIM, and LPIPS. The computational cost of RePaint[13] (467.69 s) and CoPaint[14] (2701.78 s) based on the diffusion model is significantly increased due to the iterative sampling mechanism, and the fluctuation range of CoPaint[14] is especially obvious (±13.6%). The quality assessment of our method (12.73 s) was higher than that of P2[15] (12.59 s), while being relatively equal to P2[15]. At the same time, although RePaint[13] has a slight advantage over ours in the above quality evaluation, our repair efficiency is 37 times higher than that of RePaint[13]. Our evaluation metric is 212 times more efficient than CoPaint[14] in most cases. It can be seen that we achieve a better balance between computational efficiency and inpainting quality.

Experiment on real damage

To validate the model's inpainting ability on real damaged murals, we conducted experiments on the MuralDH[41] dataset. A total of 201 murals from the test set were selected for comparison experiments, with the damaged areas manually annotated to generate corresponding masks. Figure 20 presents the inpainting results of different methods. As shown in the figure, LaMa[6] exhibits significant issues such as insufficient color filling and obvious repair traces when inpainting real damaged murals. CoPaint[14] can complete the basic structure but lacks fine detail. RePaint[13] and APLDiff show better repair results, with fainter repair traces; however, in the green box in the first row, the Repaint inpainting result has noticeable distortion and deformation. APLDiff provides the most natural inpainting results in the second and third rows.

Since there is no ground truth for real damaged murals, quantitative comparison is not possible. To evaluate the inpainting results of each method, we invited 30 volunteers to score the inpainting results of each method, with a maximum score of 10. Figure 21 shows the average scores achieved by each method. It can be seen that our model received the maximum score.

To comprehensively evaluate the inpainting performance of each model, we employed various NR-IQA methods, including ARNIQA[42], BRISQUE[43], LAR-IQA[44], and DBCNN[45], for comparison. As shown in Table 3, we systematically scored the inpainting results in Fig. 18. The experimental results demonstrate that, out of 24 evaluation metrics, our model achieved the best performance in 17 of them. This significant advantage fully validates the superiority and robustness of the proposed method.

At the same time, to enhance the objectivity of the evaluation, we also present inpainting failure cases. As shown in Fig. 22, none of the methods has achieved satisfactory inpainting results. The possible reasons are as follows: First, as a cultural heritage, murals feature extremely high image

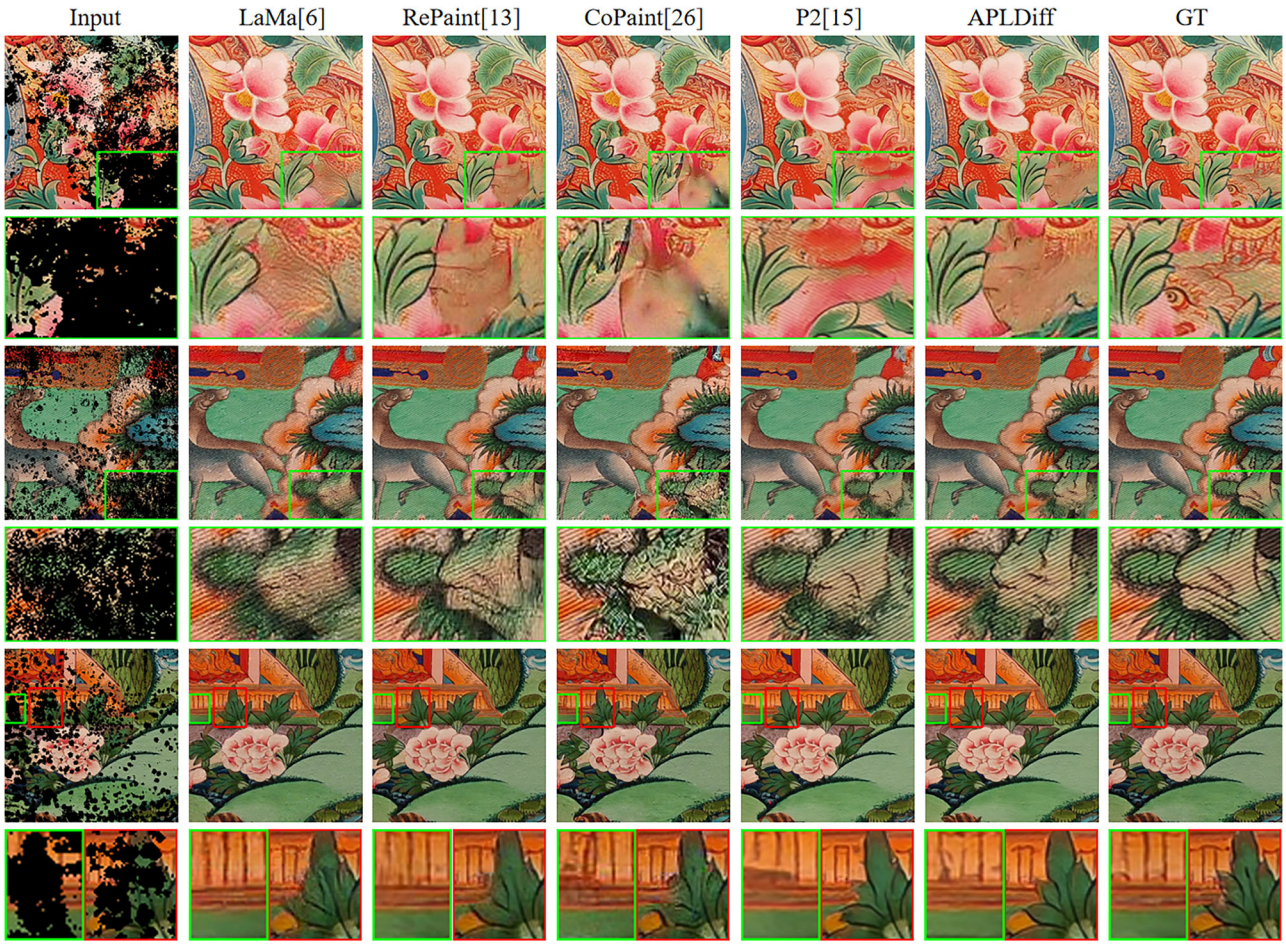

Fig. 17 | Qualitative comparison of the inpainting results on the masks of mold spot.

complexity. Inpainting not only requires visual coherence but also needs to strictly match the artistic style of specific historical periods. Although we have trained the model on our self-constructed mural dataset, inpainting for murals may require a high-quality dataset with wider coverage. Second, murals with large areas and disordered damage suffer from global semantic discontinuity. When the known regions fail to provide sufficient contextual clues, the model tends to generate content that is semantically reasonable but irrelevant to the original image, making it difficult to restore the original information of the murals.

Ablation study

First, we analyze the roles of multi-scale sampling (referred to as MSS) and adaptive perceptual weighting training (referred to as APW) through ablation experiments. In Fig. 23 and Table 4, $\gamma = 1$ represents a fixed value set for formula 17. We validate the effect of multi-scale sampling by comparing $\gamma = 1$ with MSS + ($\gamma = 1$), and then we verify the effect of adaptive perceptual weighting training by comparing MSS + ($\gamma = 1$) with MSS + APW. Note that in the $\gamma = 1$ experimental group, we also used a lightweight model, removing MSS and only using the model trained on the 256 × 256-sized DeMUDB dataset.

As shown in Fig. 23, when we add the MSS module on top of $\gamma = 1$, the inpainting effect slightly decreases. The inpainting results for MSS + ($\gamma = 1$) in the first column still exhibit some residual flaws, and there are noticeable artifacts in the inpainting results of the second column. However, as indicated in Table 4, the inpainting efficiency improved by 3.7 times with the addition of the MSS module based on $\gamma = 1$. Then we remove $\gamma = 1$ and add APW. As shown in Fig. 23, the inpainting results show a significant improvement, achieving quality that is comparable to or even better than the results with only $\gamma = 1$. This also demonstrates the effectiveness of adaptive perceptual weighting training. Table 4 provides a quantitative comparison of the three test combinations, where we conducted experimental comparisons with random masks and four types of damage masks. Compared to the baseline MSS, the inpainting efficiency significantly improved, and APW effectively enhanced the inpainting quality.

Generalization experiment

To fully demonstrate the generalization capability of our method, we additionally selected tapestries as a typical cultural heritage image for validation experiments. In the experiments, we first simulated damage to the original tapestry image, and then used it as input for the inpainting process. From the inpainting results in Fig. 24, it is evident that our method can accurately restore the texture details and cultural features, showcasing excellent inpainting performance. This result not only further validates the effectiveness of the method but also provides a reliable solution for inpainting tasks involving similar cultural heritage images, such as murals and tapestries.

Discussion

This paper addresses common issues of detail loss and efficiency bottlenecks in the digital inpainting of murals, proposing a lightweight diffusion model based on adaptive perceptual weight training. It also innovatively designs a damage mask generation framework based on physical degradation features. The method dynamically adjusts semantic weights under different noise stages, effectively mitigating the performance degradation caused by model lightweight. At the same time, a multi-scale sampling strategy is employed, enabling fast global structure generation at low resolution and fine detail inpainting at high resolution, effectively balancing inpainting quality and computational efficiency. The experiment shows that this

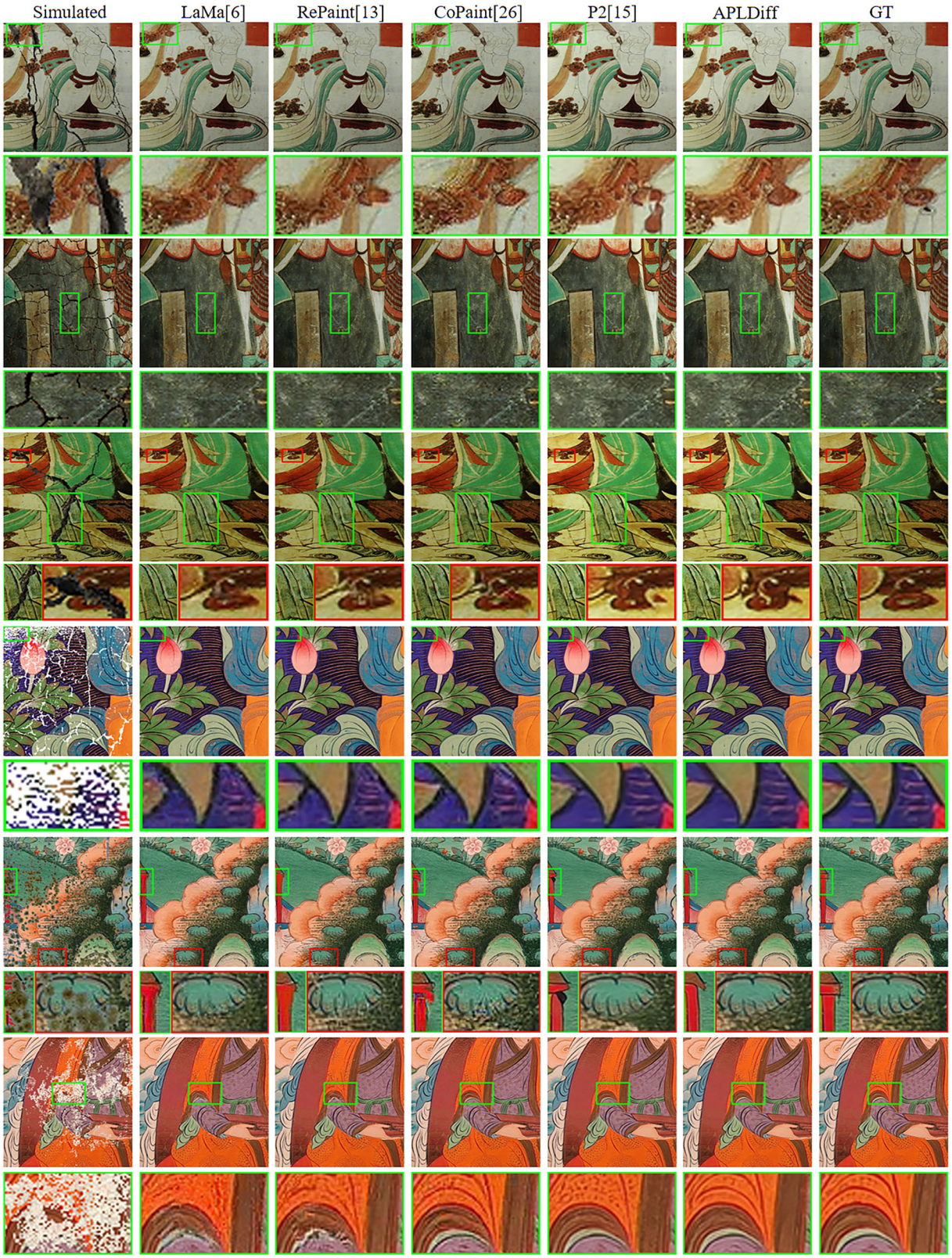

Fig. 18 | Qualitative comparison results of inpainting simulated damaged murals.

strategy reduces the inpainting time for a single mural to 12.73 s, which is a 37-fold efficiency improvement compared to RePaint[13] (467.69 s per image). Additionally, it outperforms SOTA methods such as LaMa[6] and CoPaint[14] in core metrics like SSIM and LPIPS, achieving the dual goals of "efficient inpainting" and "high-quality restoration".

The mask generation algorithm based on real physical degradation mechanisms makes the training data more aligned with the actual damage characteristics of murals, significantly improving the model's adaptability and generalization to diverse damage types. At the same time, the damaged mural images obtained through simulation are accompanied by ground

Table 1 | Quantitative comparison of inpainting results is presented from top to bottom, including random masks, four types of damage masks, and murals with simulated real damage

Condition	Metrics	LaMa[6]	RePaint[13]	CoPaint[14]	P2[15]	APLDiff
Random	SSIM↑	0.889	0.906	0.903	0.897	0.906
	LPIPS↓	0.036	0.039	0.037	0.043	0.039
	UQI↑	0.965	0.977	0.972	0.968	0.974
	GMSD↓	0.429	0.426	0.435	0.443	0.428
	GLCM_Correlation_Diff↓	0.016	0.016	0.020	0.022	0.018
Peel	SSIM↑	0.895	0.913	0.894	0.892	0.897
	LPIPS↓	0.033	0.027	0.032	0.034	0.030
	UQI↑	0.963	0.981	0.959	0.956	0.966
	GMSD↓	0.409	0.384	0.420	0.416	0.403
	GLCM_Correlation_Diff↓	0.017	0.012	0.019	0.018	0.014
Wrinkle	SSIM↑	0.959	0.972	0.962	0.964	0.969
	LPIPS↓	0.015	0.008	0.011	0.011	0.010
	UQI↑	0.987	0.994	0.990	0.993	0.995
	GMSD↓	0.234	0.217	0.227	0.232	0.228
	GLCM_Correlation_Diff↓	0.004	0.003	0.006	0.004	0.004
Crack	SSIM↑	0.940	0.955	0.938	0.933	0.937
	LPIPS↓	0.019	0.014	0.019	0.021	0.016
	UQI↑	0.980	0.987	0.978	0.975	0.979
	GMSD↓	0.264	0.236	0.267	0.262	0.254
	GLCM_Correlation_Diff↓	0.014	0.009	0.008	0.016	0.012
Mold	SSIM↑	0.853	0.901	0.845	0.891	0.897
	LPIPS↓	0.059	0.043	0.061	0.055	0.049
	UQI↑	0.959	0.967	0.951	0.955	0.949
	GMSD↓	0.457	0.412	0.466	0.442	0.431
	GLCM_Correlation_Diff↓	0.010	0.008	0.018	0.007	0.006
Simulation	SSIM↑	0.929	0.940	0.927	0.926	0.933
	LPIPS↓	0.022	0.021	0.022	0.026	0.024
	UQI↑	0.970	0.976	0.967	0.965	0.978
	GMSD↓	0.304	0.285	0.301	0.298	0.294
	GLCM_Correlation_Diff↓	0.011	0.007	0.013	0.012	0.010

Table 2 | The number of parameters and model size correspond to the results of training on 256 × 256-sized images for different methods

Models	LaMa[6]	RePaint[13]	CoPaint[14]	P2[15]	APLDiff
Number of parameters	124,130,425	552,814,086	552,814,086	93,563,910	93,563,910
Inpainting Efficiency	0.31	467.69	2701.78	12.59	12.73
Model size	391 MB	2159 MB	2159 MB	367 MB	367 MB

The inpainting efficiency is the average time in seconds for inpainting one image out of 100 test images.

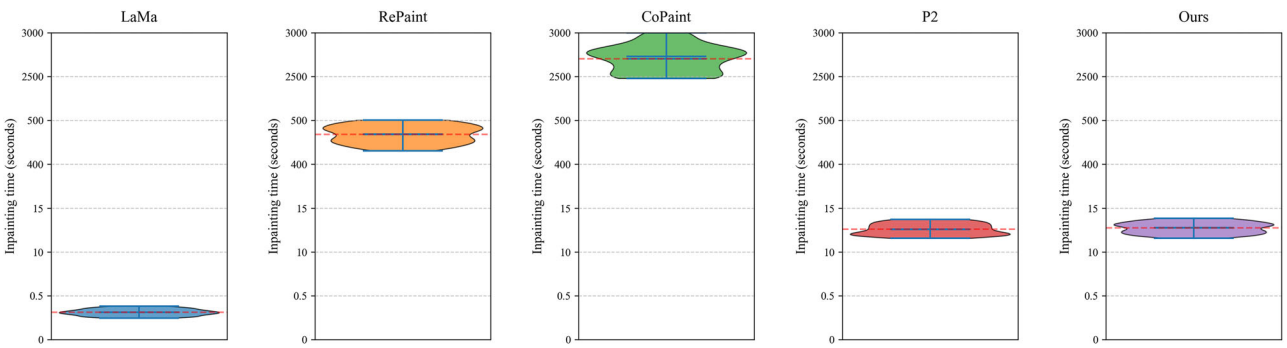

Fig. 19 | A violin graph showing the inpainting efficiency of 100 test images by different models.

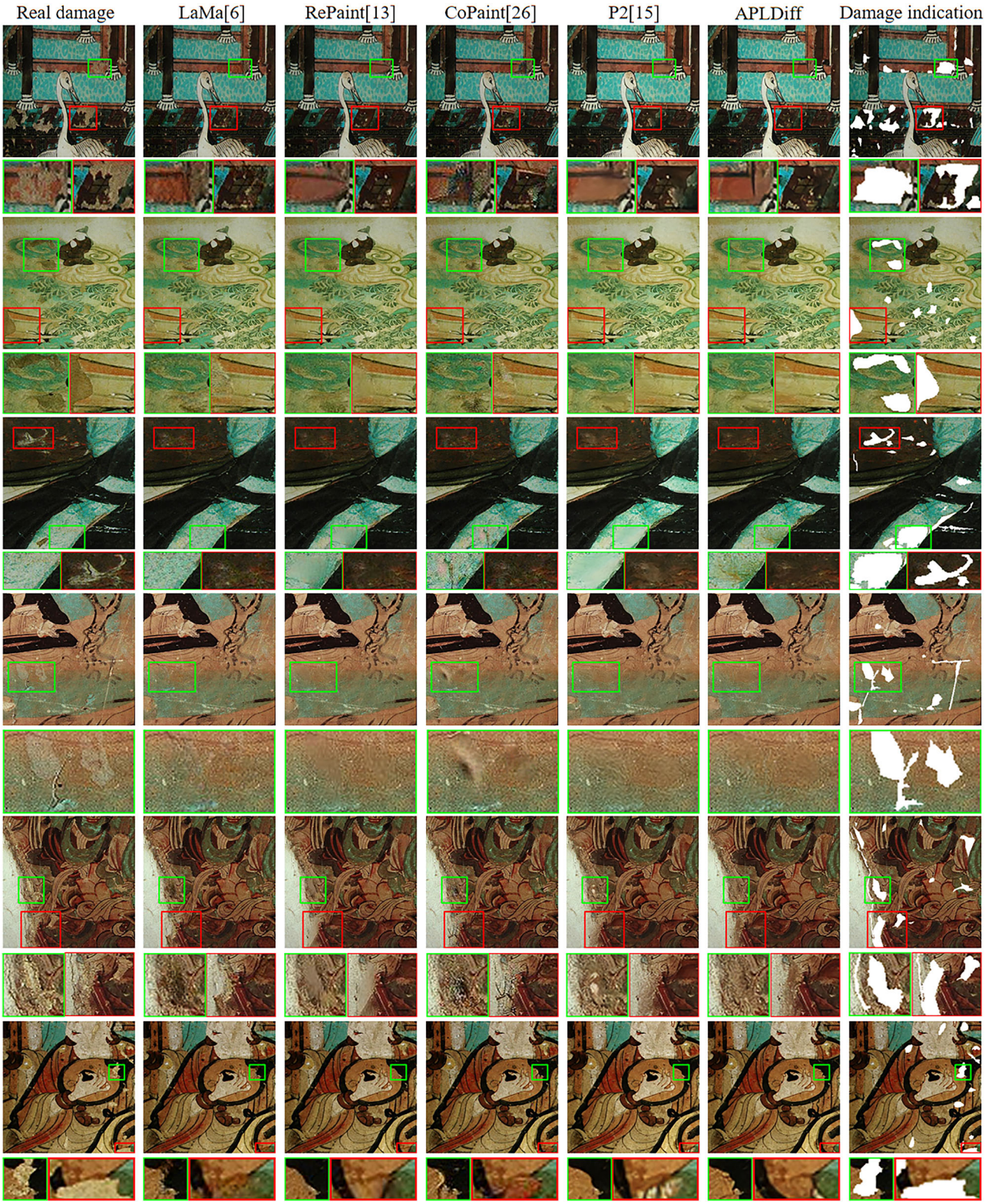

Fig. 20 | Qualitative comparison of inpainting results of different models on real damaged mural images.

truth references, which enable us to conduct quantitative analysis and comparisons of the inpainting results, overcoming the limitations of purely qualitative assessments. Experimental results show that this method achieves excellent quantitative evaluation metrics in the inpainting tasks of four typical damage types: pigment peeling, wrinkles, cracks, and mold spots, validating the practical value and technical advancement of the proposed model in mural digital inpainting.

The APLDiff framework proposed in this study achieves a balance between authenticity and efficiency in digital mural inpainting. However, there remains a key limitation in practical applications: the inpainting of damaged areas relies on a pre-labeled damage mask, which, to some extent, limits the framework's automation and practicality. The core of this limitation lies in the fact that the framework has not yet achieved an end-to-end integration of "damage detection -

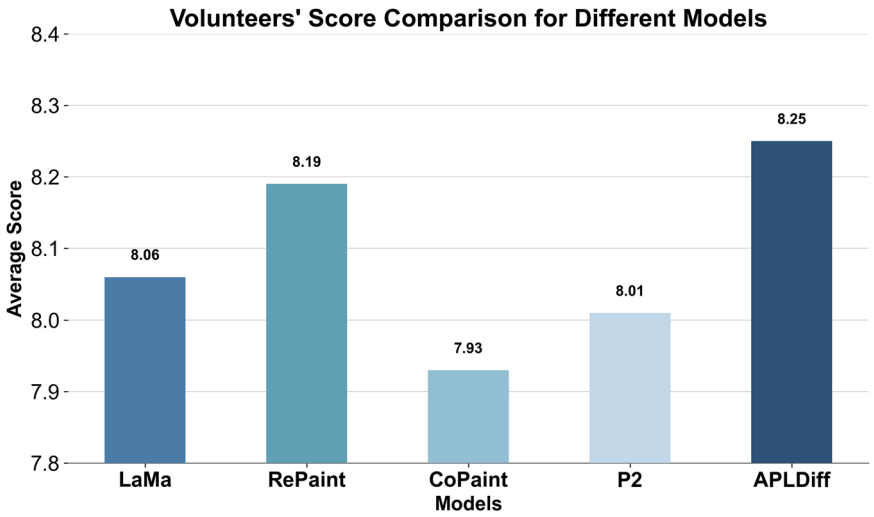

Fig. 21 | The inpainting results of different models received the average scores from 30 volunteers.

Table 3 | The no-reference image quality assessment of the inpainting results from different models on real damaged murals is presented

Murals	NR-IQA	LaMa[6]	RePaint[13]	CoPaint[14]	P2[15]	APLDiff
1	ARNIQA↑	0.7756	0.7469	0.7510	0.7669	0.7924
	BRISQUE↓	50.5666	47.7507	54.3044	36.7359	38.3254
	LAR-IQA↑	0.5296	0.5394	0.5273	0.5261	0.5521
	DBCNN↑	0.5969	0.6331	0.5804	0.6698	0.6706
2	ARNIQA↑	0.6699	0.6655	0.6580	0.7029	0.7065
	BRISQUE↓	41.3019	37.6335	46.4474	26.9343	24.7873
	LAR-IQA↑	0.5863	0.5837	0.5871	0.5919	0.5920
	DBCNN↑	0.6709	0.6857	0.6482	0.7031	0.7101
3	ARNIQA↑	0.5883	0.6459	0.6211	0.6935	0.6622
	BRISQUE↓	25.0256	23.0905	29.8975	18.8332	18.5807
	LAR-IQA↑	0.5337	0.5437	0.5615	0.5383	0.5520
	DBCNN↑	0.5152	0.5450	0.5159	0.6162	0.6079
4	ARNIQA↑	0.5263	0.5325	0.5175	0.5484	0.5602
	BRISQUE↓	57.2189	55.2648	57.3430	50.4377	51.2956
	LAR-IQA↑	0.5949	0.5898	0.6004	0.6058	0.6066
	DBCNN↑	0.6098	0.6230	0.6115	0.6394	0.6564
5	ARNIQA↑	0.6311	0.6842	0.6311	0.6720	0.6697
	BRISQUE↓	40.9831	39.6657	44.6379	38.1403	34.4787
	LAR-IQA↑	0.5736	0.5900	0.5834	0.5577	0.5593
	DBCNN↑	0.6378	0.6551	0.6213	0.6623	0.6664
6	ARNIQA↑	0.7388	0.7505	0.7276	0.7587	0.7614
	BRISQUE↓	19.9890	16.4025	25.3151	13.7722	11.4401
	LAR-IQA↑	0.5397	0.5405	0.5432	0.5458	0.5466
	DBCNN↑	0.6523	0.6590	0.6406	0.6707	0.6815

inpainting." The current design treats the localization of damaged areas and the inpainting process as separate steps, lacking dynamic awareness of damage features.

Future work will further enrich the physical mechanism of damage simulation, combine multi-modal data to improve the accuracy of inpainting, for example, by integrating infrared imaging data (which captures underlying textures invisible to the naked eye) and hyperspectral imaging data (which analyzes the chemical composition and distribution of pigments, and identifies the spectral characteristics of pigments), the inpainting process can not only better restore the visual appearance but also align more closely with the original creative logic and material properties. The introduction of a "damage feature self-supervised learning" module utilizes a large amount of unlabeled damaged mural data to train the model to automatically learn visual features of damage, such as pigment loss and wall cracks, enabling the direct generation of high-precision masks from the original image for inpainting. And explore efficient processing strategies for high-resolution images to meet the needs of more complex mural

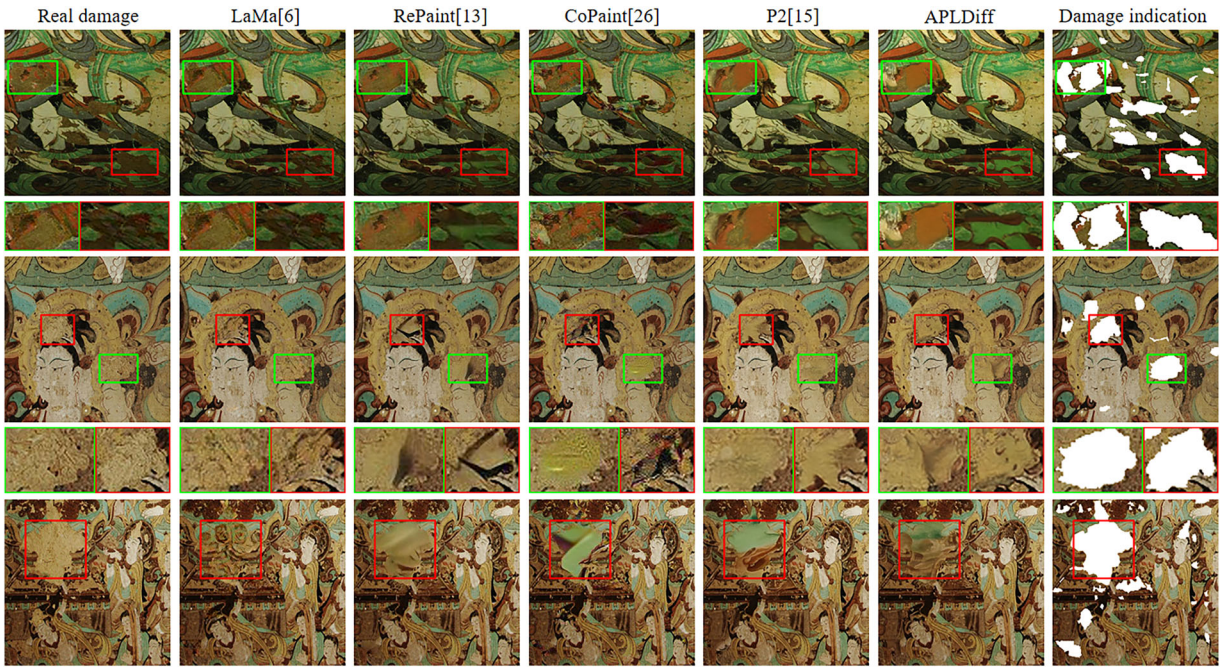

Fig. 22 | Failure case results of each method for inpainting real damaged murals.

Fig. 23 | Qualitative comparison of inpainting results of different strategies.

Table 4 | Comparison of different inpainting strategies

Strategies			Inpainting Efficiency	Random			Simulation		
MSS	$\gamma = 1$	APW		SSIM↑	LPIPS↓	GMSD↓	SSIM↑	LPIPS↓	GMSD↓
	√		46.68	0.898	0.036	0.421	0.897	0.029	0.363
√	√		12.55	0.887	0.029	0.434	0.881	0.039	0.395
√		√	12.62	0.991	0.033	0.426	0.890	0.035	0.375

The inpainting efficiency is the tie time to inpaint a single image.

Fig. 24 | Inpainting results of our model on the tapestries.

inpainting. By continuously optimizing the algorithm and enhancing the interactivity, the digital inpainting technology of murals is promoted to a more intelligent and practical direction.

Data availability

The datasets used and presented in this study are available from the corresponding author upon reasonable request.

Code availability

Not applicable. If the code is open-sourced in the future, it will be made available via GitHub or the corresponding author's email.

Received: 9 July 2025; Accepted: 15 December 2025;
Published online: 16 January 2026

References

1. Wang, Y. H. & Wu, X. D. Current progress on murals: distribution, conservation and utilization. *Herit. Sci.* **11**, 61 (2023).
2. Baglioni, M., Poggi, G., Chelazzi, D. & Baglioni, P. Advanced materials in cultural heritage conservation. *Molecules* **26**, 3967 (2021).
3. Bertalmio, M., Sapiro, G., Caselles, V. & Ballester, C. Image inpainting. In *Proc. 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2000)* 417–424 (ACM, 2000).
4. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T. & Efros, A. A. Context encoders: feature learning by inpainting. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 2536–2544 (IEEE, 2016).
5. Yu, J. H., Lin, Z., Yang, J. M., Shen, X. H. & Huang, T. S. Generative image inpainting with contextual attention. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 5505–5514 (IEEE, 2018).
6. Suvorov, R. et al. Resolution-robust large mask inpainting with Fourier convolutions. In *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* 3172–3182 (IEEE, 2022).
7. Zheng, H. et al. Image inpainting with cascaded modulation GAN and object-aware training. In *Proc. European Conference on Computer Vision* 277–296 (Springer, 2022).
8. Peng, J. L., Liu, D., Xu, S. C. & Li, H. Q. Generating diverse structure forimage inpainting with hierarchical VQ-VAE. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 10770–10779 (IEEE, 2021).
9. Mishra, A., Reddy, M. S. K., Mittal, A. & Murthy, H. A. A generative model for zero shot learning using conditional variational autoencoders. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops* 2269–2277 (IEEE, 2018).
10. Lin, X. M., Li, Y. K., Hsiao, J., Ho, C. M. & Kong, Y. Catch missing details: image reconstruction with frequency augmented variational autoencoder. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition* 1736–1745 (IEEE, 2023).
11. Saharia, C. et al. Palette: image-to-image diffusion models. In *Proc. ACM SIGGRAPH 2022 Conference* (ACM, 2022).

12. Zhang, L. M., Rao, A. & Agrawala, M. Adding conditional control to text-to-image diffusion models. In *Proc. IEEE/CVF International Conference on Computer Vision* 3813–3824 (IEEE, 2023).
13. Lugmayr, A. et al. RePaint: inpainting using denoising diffusion probabilistic models. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition* 11451–11461 (IEEE, 2022).
14. Zhang, G. H. et al. Towards coherent image inpainting using denoising diffusion implicit models. In *Proc. 40th International Conference on Machine Learning* Vol. 202, 41164–41193 (PMLR, 2023).
15. Choi, J. et al. Perception prioritized training of diffusion models. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition* 11462–11471 (IEEE, 2022).
16. Ge, H., Yu, Y. & Zhang, L. A virtual restoration network of ancient murals via global–local feature extraction and structural information guidance. *Herit. Sci.* **11**, 264 (2023).
17. Zhang, S. & Yang, F. Digital mural inpainting model based on improved two-stage generative adversarial network. *Electron. Meas. Technol.* **46**, 123–129 (2024).
18. Cao, J. F., Zhang, Z. B., Zhao, A. D., Cui, H. Y. & Zhang, Q. Ancient mural restoration based on a modified generative adversarial network. *Herit. Sci.* **8**, 7 (2020).
19. Wu, M., Chang, X. & Wang, J. Fragments inpainting for tomb murals using a dual-attention mechanism GAN with improved generators. *Appl. Sci.* **13**, 3972 (2023).
20. Xu, Z. G. & Geng, C. P. Color restoration of mural images based on a reversible neural network: leveraging reversible residual networks for structure and texture preservation. *Herit. Sci.* **12**, 351 (2024).
21. Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. In *Proc. 34th International Conference on Neural Information Processing Systems* 6840–6851 (2020).
22. Song, J., Meng, C., & Ermon, S. Denoising diffusion implicit models. In *Proc. International Conference on Learning Representations* (ICLR, 2021).
23. Rombach, R., Blattmann, A., Lorenz, D., Esser, P. & Ommer, B. High-resolution image synthesis with latent diffusion models. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 10674–10685 (IEEE, 2022).
24. Liu, L. P., Ren, Y., Lin, Z. j. & Zhao, Z. Pseudo numerical methods for diffusion models on manifolds. In *Proc. International Conference on Learning Representations* (ICLR, 2022).
25. Samlimans, T. & Ho, J. Progressive Distillation for fast sampling of diffusion models. In *Proc. International Conference on Learning Representations* (ICLR, 2022).
26. Xia, B. et al. DiffIR: efficient diffusion model for image restoration. In *Proc. IEEE International Conference on Computer Vision (ICCV)* 13049–13059 (IEEE, 2023).
27. Meng, C. L. et al. SDEdit: guided image synthesis and editing with stochastic differential equations. In *Proc. International Conference on Learning Representations* (ICLR,2022).
28. Nichol, A. Q. et al. GLIDE: Towards photorealistic image generation and editing with text-guided diffusion models. In *Proc. International Conference on Machine Learning (ICML)* Vol. 162, 16784–16804 (PMLR, 2021).
29. Schneuing, A. et al. Structure-based drug design with equivariant diffusion models. *Nat. Comput. Sci.* **4**, 899–909 (2024).
30. Wang, X. B., Li, W. J. & Lu, C. A mask guided cross data augmentation method for industrial defect detection. *Future Gener. Comput. Syst.* **166**, 107676 (2025).
31. Wang, X. B., Li, W. J. & He, X. J. MTDiff: Visual anomaly detection with multi-scale diffusion models. *Know. Based Syst.* **302**, 112364 (2024).
32. Yu, Y. R., Gu, Y. N., Zhang, S. T. & Zhang, X. F. MedDiff-FM: a diffusion-based foundation model for versatile medical image applications. Preprint at https://doi.org/10.48550/arXiv.2410.15432 (2024).
33. Kim, B., Oh, Y. j. & Ye, J. C. Diffusion adversarial representation learning for self-supervised vessel segmentation. In *Proc. International Conference on Learning Representations* (ICLR, 2023).
34. Wu, L. M., Gong, C. Y., Liu, X. C., Ye, M. & Liu, Q. Diffusion-based molecule generation with informative prior bridges. In *Proc. 36th International Conference on Neural Information Processing Systems (NIPS)* 36533–36545 (Curran Associates Inc, 2022).
35. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N. & Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proc. 32nd International Conference on International Conference on Machine Learning (ICML)* Vol. 37, 2256–2265 (PMLR, 2015).
36. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).
37. Zhang, R., Isola, P., Efros, A. A., Shechtman, E. & Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In *Proc. 2018 IEEE Conference on Computer Vision and Pattern Recognition* 586–595 (IEEE, 2018).
38. Wang, Z. & Bovik, A. C. A universal image quality index. *IEEE Signal Process. Lett.* **9**, 81–84 (2002).
39. Xue, W., Zhang, L., Mou, X. & Bovik, A. C. Gradient Magnitude Similarity Deviation: a highly efficient perceptual image quality index. *IEEE Trans. Image Process.* **23**, 684–695 (2014).
40. Haralick, R. M., Shanmugam, K. & Dinstein, I. Textural features for image classification. *IEEE Trans. Syst. Man Cybern*. **SMC-3**, 610–621 (1973).
41. Xu, Z. S. et al. A comprehensive dataset for digital restoration of Dunhuang murals. *Sci. Data.* **11**, 955 (2024).
42. Agnolucci, L., Galteri, L., Bertini, M. & Del Bimbo, A. ARNIQA: learning distortion manifold for image quality assessment. In *Proc. 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* 188–197 (IEEE, 2024).
43. Mittal, A., Moorthy, A. K. & Bovik, A. C. Blind/referenceless image spatial quality evaluator. In *Proc. 2011 Asilomar Conference on Signals, Systems and Computers* 723–727 (IEEE, 2011).
44. Jamshidi Avanaki, N., Ghildyal, A., Barman, N. & Zadtootaghaj, S. LAR-IQA: a lightweight, accurate, and robust no-reference image quality assessment model. In *Proc. European Conference on Computer Vision (ECCV)* 328–345 (Springer, 2025).
45. Zhang, W. X., Ma, K., Yan, J., Deng, D. X. & Wang, Z. Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Trans. Circuits Syst. Video Technol.* **30**, 36–47 (2020).

Acknowledgements

The authors gratefully acknowledge the financial support provided by the Youth Fund of the Natural Science Foundation of Qinghai Province (Grant No. 2023-ZJ-947Q) and the National Natural Science Foundation of China (Grant Nos. 6246070542, 62262056).

Author contributions

D.Z. (author) and X.C. wrote the main manuscript text. D.Z. (author) mainly wrote the methods and experiments in the manuscript, and X.C. mainly wrote the introduction and related work in the manuscript. D.Z. (corresponding author) mainly tests the code. J.L. drew the illustrations in the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Dan Zhang.