

<https://doi.org/10.1038/s40494-025-01693-z>

# GuidePaint: lossless image-guided diffusion model for ancient mural image restoration

**Jialv Hu, Ying Yu✉ & Qixue Zhou**

Ancient murals, vital cultural heritage, suffer from damage due to natural erosion and human activities. Traditional restoration methods, relying on manual repair, have limitations, making virtual restoration an innovative solution. This paper proposes a virtual restoration method based on diffusion model. Using a lossless image-guided algorithm, we adapt diffusion model designed for image synthesis to restoration. Instead of feeding damaged images into the network, we use them to adjust the network's outputs directly, achieving unsupervised training. We also use random seeds to generate diverse outputs from a single image. Proposed similarity function ensures alignment of undamaged areas with the guiding image, and an interrupt sampling strategy removes subtle, dense degradations. Experiments on simulated and real damaged murals show our method yields results comparable to or better than other advanced methods for simple cases. For complex and severely damaged murals, it excels, outperforming others in both objective and subjective evaluations.

From the first spark that ignited the flames of civilization in ancient times to the industrial machines that reshape the world today, human has had a magnificent journey. Reflecting on the past, we possess precious treasures—chief among them is our history. Murals, as one of the earliest means of recording information, carry significant historical and cultural value. However, due to natural erosion and human activities, many ancient murals have suffered from issues such as cracks, flaking, mold, and fading. As a non-renewable resource, the protection and restoration of these murals have become urgent tasks.

Traditional mural restoration primarily relies on manual, physical methods, requiring specialists with comprehensive knowledge of history, humanities, arts, and archeology. Moreover, these methods may cause irreversible damage to the murals. As an alternative, the digital restoration of mural images has emerged as a novel approach in the field of ancient mural preservation. Without affecting the original artwork, this method uses computer technology to virtually reconstruct the murals. The digitally restored mural images can not only serve as references for physical restoration but also help build a replicable database, offering a more reliable means for the preservation and transmission of these cultural artifacts.

Using computer technology to restore digital images of ancient murals aims to virtually reconstruct the missing or degraded areas as accurately as possible. In recent years, many outstanding image restoration techniques have emerged, which can generally be categorized into two types: conventional image restoration methods and deep learning-based methods.

Conventional image restoration methods are primarily based on two techniques: patch-based and diffusion-based (not to be confused with the diffusion model discussed in this paper). The core idea behind patch-based methods is to find image patches similar to the damaged areas and use them to replace the missing regions. Yang et al.<sup>1</sup> proposed a method for restoring complex damaged areas in the Dunhuang murals using Dempster-Shafer evidence theory and its data fusion. This algorithm outperforms the Criminisi<sup>2</sup> algorithm and its improvements in visual quality, though it requires more processing time. Li et al.<sup>3</sup> introduced an automatic recognition and virtual restoration method for mud-spot damage in Tang dynasty tomb murals. This method automatically analyzes images to obtain mud-spot area masks, which are then used in a patch-based approach for restoration. Jiao et al.<sup>4</sup> proposed an improved patch-based algorithm to restore the Wutai Mountain murals. Compared to other methods, their algorithm shows better restoration results for textures, edges, and smooth areas, though it depends on the relative position and connectivity of the mask regions. Wang et al.<sup>5</sup> tackled the issue of large-scale structural damage in the murals of Yulin Caves and Mogao Caves in Gansu. They proposed using manually drawn lines to assist with restoration and combining global and local feature weighting based on structural guidance for restoration. The method surpasses other restoration methods for the targeted murals but heavily relies on artist guidance and line drawings. Cao et al.<sup>6</sup> developed an algorithm for restoring ancient temple murals that employs a local search algorithm of an adaptive sample block. Compared to the Criminisi

algorithm, the approach improves both the efficiency and effectiveness of restoration.

Diffusion-based methods fill missing areas by simulating the spread of pixels within an image, typically using gradient information and local features to guide the restoration process. Shen et al.<sup>7</sup> proposed a mural restoration algorithm based on Morphological Component Analysis (MCA), inspired by the manual restoration approach for Tang dynasty tomb murals, where the global structures are addressed first, followed by local textures. The algorithm shows significant results in restoring cracks in Tang tomb murals but does not extend to murals of other styles or damage types. Jaidilert et al.<sup>8</sup> introduced a computer aided semi-automatic repair framework. The framework involves user-provided seeds, followed by region-growing and morphological operations to identify the location of scratches, and finally applies various conventional image restoration methods such as Total Variation (TV)<sup>9</sup>, Curvature-Driven Diffusion (CDD)<sup>10</sup>, Lattice Boltzmann Method (LBM)<sup>11</sup>, and Group Sparsity Regularization (GSR)<sup>12</sup>. Chen et al.<sup>13</sup> proposed an improved adaptive curvature-driven model for restoring Dunhuang murals. The approach addresses issues such as false edges and staircase effects in the restoration results, and uses adaptive control strategies to reduce restoration time, but is less effective when dealing with extensive irregular damage and specific mural diseases.

Despite many years of development, conventional image restoration methods still struggle to meet the demands of digital restoration for ancient murals, mainly for two aspects: (1) poor performance in restoring murals with large areas of damage, and (2) inability to generate new information based on existing content. Consequently, with the rise of deep learning techniques in the field of image processing, researchers have increasingly shifted their focus to deep learning-based methods. Yu et al.<sup>14</sup> proposed an end-to-end network based on U-Net for restoring Dunhuang murals, which demonstrates excellent performance in repairing highly non-rigid and irregular deteriorated regions. Cao et al.<sup>15</sup> introduced dilated convolutions to improve GANs<sup>16</sup> for the restoration of ancient Chinese murals. However, when dealing with complex structures and large missing areas, issues like blurriness and loss of structural information may arise. Wang et al.<sup>17</sup> proposed a Thanka mural restoration method based on multi-scale adaptive partial convolutions and stroke-like masks, but it also struggles with handling complex structures and large damaged areas. Ciortan et al.<sup>18</sup>, inspired by the modus operandi of an artist—edges first, then the color palette, and color tones at last—proposed a GAN-based restoration algorithm with two generators: one for edge and the other for color. The approach achieves satisfactory results in the restoration of Dunhuang murals. Li et al.<sup>19</sup> proposed a line drawing guided progressive mural inpainting method, dividing the inpainting process into two steps: structural reconstruction and color correction. The approach shows superior performance in mural image restoration, but it relies on line drawings manually created by professional artists as guidance. Schmidt et al.<sup>20</sup> proposed an image restoration method combining CAR<sup>21</sup> and HiNet<sup>22</sup> for the restoration of Dunhuang murals. The method performs well even with large damaged areas but may result in some loss of detail. Ge et al.<sup>23</sup> proposed a virtual restoration network for ancient murals based on global-local feature extraction and structural information guidance. The method achieves excellent results in terms of structural continuity, color harmony, and visual rationality in the restoration of mural images. Ren et al.<sup>24</sup> proposed a generative adversarial network model that combines a parallel dual convolutional feature extraction depth generator and a ternary heterogeneous joint discriminator, achieving effective restoration of Dunhuang murals.

Although current deep learning methods can handle some cases of large-scale damage in murals, several challenges remain: (1) Existing methods, mostly based on supervised learning, are highly influenced by the types of masks used during training. This leads to inconsistent performance when dealing with diverse types of damage. (2) Restoration of large-scale damage often results in blurriness or loss of fine details, especially in murals with intricate structures. This is a limitation common to existing methods. (3) The trained models in existing methods typically produce deterministic

outputs. If a specific mural cannot be well-restored, repeated attempts will yield the same unsatisfactory results. (4) Existing methods require masks to be obtained through marking and cannot handle subtle, dense degradation patterns that cannot be explicitly marked.

To address these challenges, we propose a mural image restoration method based on conditional diffusion model<sup>25</sup>. The main contributions of this paper are twofold: (1) We propose a similarity function that adapts the diffusion model—originally designed for image synthesis tasks—to the image restoration task in an indirect manner. The image synthesis capabilities of the diffusion model surpass those of existing deep learning frameworks, such as GANs. This contribution can result in extraordinary performance when applied to restoration tasks, particularly for images with large areas of damage and complex details (solve challenge 2). Moreover, the proposed similarity function is applied to the sampling algorithm rather than the backbone network, enabling unsupervised training and eliminating reliance on masks (solve challenge 1). Additionally, since the similarity function does not impose direct constraints on the unknown (damaged) areas during the sampling process, it allows the inherent randomness of the diffusion model in image synthesis tasks to be leveraged for restoration, leading to more diverse restoration outcomes (solve challenge 3). (2) We also propose an interrupted sampling strategy to address subtle, dense degradation types that cannot be explicitly marked, without requiring any masks throughout the whole process (solve challenge 4).

## Methods

### Diffusion model

Ancient mural images usually contain intricate structures, rich textures, and diverse color information, which place high demands on the generative capabilities of restoration models, especially when dealing with large areas of damage. Existing research<sup>25</sup> has demonstrated that diffusion models have surpassed GANs, becoming the leading models for high-quality image synthesis. Therefore, we introduce diffusion model and propose a lossless image-guided algorithm to better address the challenges of mural restoration tasks.

The diffusion model is a generative model that, like other generative models, learns the distribution of the training set and generates similar samples. However, unlike other generative models, the training of a diffusion model is not an end-to-end process. Instead, it progressively adds noise to the origin images until they become pure noise, while a neural network is trained to learn this noising process. During sampling, this noise process is reversed—i.e., denoising—allowing the model to generate a target sample from random noise. The framework of the model is illustrated in Fig. 1.

Let the original image be represented as  $\mathbf{x}_0$ , the image at step  $t$  as  $\mathbf{x}_t$ , and the image at the final step as  $\mathbf{x}_T$ . In the training of DDPM<sup>26</sup>, the diffusion process converts the image  $\mathbf{x}_0$  into Gaussian white noise  $\mathbf{x}_T \sim \mathcal{N}(0, 1)$  over  $T$  steps. Each step is defined as

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) \sim \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I}) \quad (1)$$

That is,  $\mathbf{x}_t$  is obtained by scaling  $\mathbf{x}_{t-1}$  by a factor of  $\sqrt{1 - \beta_t}$  and then adding Gaussian noise with mean 0 and variance  $\beta_t \mathbf{I}$ . Here,  $\beta_t$  is a predefined hyperparameter, and  $\sqrt{1 - \beta_t}$  decreases as  $t$  increases.

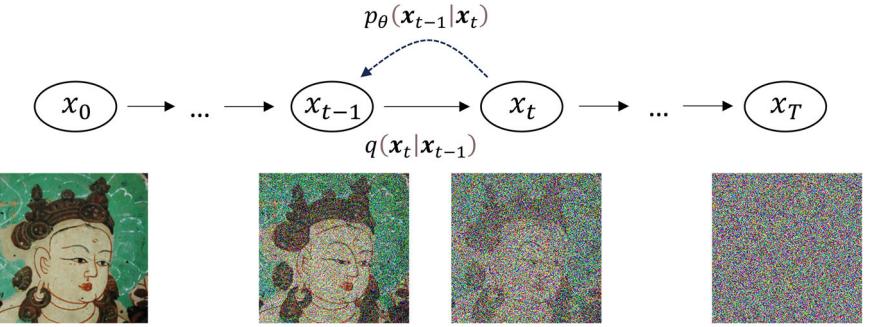
By the Markov assumption, the joint probability of all the latent variables can be factorized as a product of conditional probabilities over all the steps. This allows us to rewrite the forward diffusion process as

$$q(\mathbf{x}_t | \mathbf{x}_0) \sim \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t} \mathbf{x}_0, (1 - \bar{\alpha}_t) \mathbf{I}) \quad (2)$$

where  $\alpha_t = 1 - \beta_t$  and  $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$ . Equation (2) shows that we can sample  $\mathbf{x}_t$  at any timestep  $t$  directly from  $\mathbf{x}_0$  without going step by step, effectively scaling and adding noise in one shot.

Using a neural network to obtain the mean  $\mu_\theta(\mathbf{x}_t, t)$  and variance  $\Sigma_\theta(\mathbf{x}_t, t)$ , we can define the reverse denoising process, which generates an image by progressively removing noise. The reverse process is modeled as a

**Fig. 1 | The framework of diffusion model.** The forward process  $q(x_t|x_{t-1})$ , the model training stage, progressively adds noise to the original image. The reverse process  $p_\theta(x_{t-1}|x_t)$ , the model prediction stage, removes the noise step by step.



parameterized Gaussian distribution:

$$p_\theta(x_{t-1}|x_t) \sim \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \quad (3)$$

where  $\mu_\theta(x_t, t)$  and  $\Sigma_\theta(x_t, t)$  are the mean and variance obtained from the neural network, conditioned on the noisy image  $x_t$  and the timestep  $t$ .

Combining the forward process defined by Equation (2) and the reverse process defined by Equation (3), the mean  $\mu_\theta(x_t, t)$  can be obtained as

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left( x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}} \right) \epsilon_\theta(x_t, t) \quad (4)$$

where  $\epsilon_\theta(x_t, t)$  is the noise predicted by the neural network for timestep  $t$ .

To train the target network, we need to derive the optimization objective based on the Variational Lower Bound (VLB) of the log-likelihood of the data:

$$\begin{aligned} \mathbb{E}[-\log p_\theta(x_0)] &\leq \mathbb{E}_q \left[ -\log \frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)} \right] \\ &= \mathbb{E}_q \left[ \underbrace{D_{KL}(q(x_T|x_0) \| p(x_T))}_{L_T} + \sum_{t>1} \underbrace{D_{KL}(q(x_{t-1}|x_t, x_0) \| p_\theta(x_{t-1}|x_t))}_{L_{t-1}} - \underbrace{\log p_\theta(x_0|x_1)}_{L_0} \right] \\ &= L_{\text{VLB}} \end{aligned} \quad (5)$$

By simplifying the  $L_{t-1}$  term, we can derive the final optimization objective for training the diffusion model:

$$L_{\text{simple}} = \mathbb{E}_{t, x_0, \epsilon} \left[ \|(\epsilon - \epsilon_\theta(x_t, t))\|^2 \right] \quad (6)$$

The resulting simplified objective is the mean squared error (MSE) between the true noise  $\epsilon$  and the noise predicted by the neural network  $\epsilon_\theta(x_t, t)$ .

### Proposed lossless image-guided sampling algorithm

The original diffusion model generates samples directly from noise. To apply it to image inpainting or restoration tasks, we need to introduce control conditions. Dhariwal et al.<sup>25</sup> modified the model's sampling process to achieve classifier-guided generation. Liu et al.<sup>27</sup> further expanded the classifier to various guiding forms, such as text guidance and image guidance. With the introduction of control conditions, the forward process remains unaffected and is still defined by Eq. (1), while the reverse process is defined as

$$p_\theta(x_{t-1}|x_t, y) \sim \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t) + \gamma \Sigma_\theta(x_t, t) \nabla_{x_t} F(x_t, y, t), \Sigma_\theta(x_t, t)) \quad (7)$$

where  $y$  represents the condition,  $F(x_t, y, t)$  is a similarity measure that quantifies the relationship between the sample  $x_t$  and the condition  $y$ . Equation (7) can be interpreted as using the gradient of  $F(x_t, y, t)$  to adjust the mean  $\mu_\theta(x_t, t)$  (including both known and unknown regions), causing it

to shift towards areas with higher values of  $F(x_t, y, t)$ . The coefficient  $\gamma$  controls how much the gradient influences the mean.

The specific similarity measure  $F(x_t, y, t)$  depends on the task. For example, in their image content guidance task, Liu et al.<sup>27</sup> passed  $x_t$  and  $y$  through a pre-trained image encoder for noisy images to obtain feature embeddings, then calculated the distance between these feature embeddings. But here is an important issue regarding image inpainting: when using an image to guide the restoration, the lossy compression process during feature embedding can result in a significant difference between the details of the known regions in the original image and the output of the model. To address this problem, we propose a new method for calculating the similarity measure  $F(x_t, y, t)$ . By examining Equation (2), we find that reversing it yields

$$x_0(x_t, t) = \frac{1}{\sqrt{\alpha_t}} (x_t - \sqrt{1 - \alpha_t} \epsilon_\theta(x_t, t)) \quad (8)$$

In other words, at each step of the sampling process, we can obtain a rough estimate of  $x_0(x_t, t)$ . Therefore, we define  $F(x_t, y, t)$  as the distance between  $x_0(x_t, t)$  and  $y$ . Let  $m$  denote the mask,  $y$  the conditional image, and the known region as  $(1 - m) \odot y$ , the function  $F(x_t, y, t)$  we define can be expressed as

$$F(x_t, y, t) = \|((1 - m) \odot x_0(x_t, t) - (1 - m) \odot y)\|^2 \quad (9)$$

Since we directly use  $y$  for calculations without any lossy compression, the generated results can achieve a high level of detail consistency with the reference image in the known regions. Our algorithm is outlined in Fig. 2 and Algorithm 1.

### Algorithm 1. Proposed Lossless Image-Guided Sampling Algorithm

```

1:  $x_T \sim \mathcal{N}(0, 1)$ 
2: for  $t = T, \dots, 1$  do
3:    $x_0(x_t, t) = \frac{1}{\sqrt{\alpha_t}} (x_t - \sqrt{1 - \alpha_t} \epsilon_\theta(x_t, t))$ 
4:    $\nabla_{x_t} F(x_t, y, t) = \nabla_{x_t} \|((1 - m) \odot x_0(x_t, t) - (1 - m) \odot y)\|^2$ 
5:    $\epsilon \sim \mathcal{N}(0, 1)$ 
6:    $\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} (x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}} \epsilon_\theta(x_t, t))$ 
7:    $x_{t-1}(x_t, t) = \mu_\theta(x_t, t) + \gamma \Sigma_\theta(x_t, t) \nabla_{x_t} F(x_t, y, t) + \sqrt{\Sigma_\theta(x_t, t)} \epsilon$ 
8: end for
9:  $x = (1 - m) \odot y + m \odot x_0(x_1, t)$ 
10: return  $x$ 

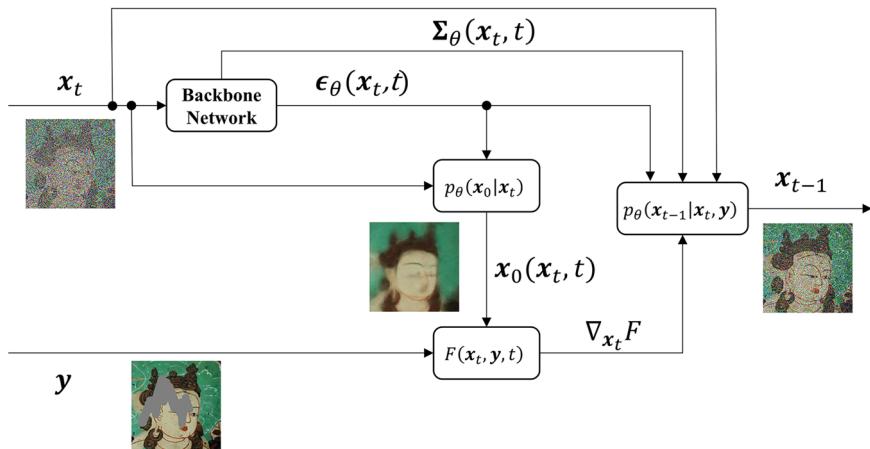
```

## Results

### Dataset

We collect a number of ancient mural images with minimal or no damage to construct a mural dataset. The sources of these murals are the Dunhuang murals from Gansu and the Baisha murals from Yunnan, which encompass different periods, styles, and subjects. The Dunhuang murals span a long period, beginning in the Northern Wei dynasty (386–534) and lasting until the Qing dynasty (1644–1911), with each period exhibiting its own distinct

**Fig. 2** | Proposed lossless image-guided sampling algorithm.



style. The Baisha murals are from the Ming (1368–1644) and Qing (1644–1911) dynasties, representing a fusion of various religious cultures, including Taoism, Tibetan Buddhism, Han Buddhism, and Dongbaism. The original images are cropped into small sub-images with minimal overlap and augmented to produce 10,314 images of size  $256 \times 256$ , with 8251 images used for training and 2063 for simulated damage testing. Additionally, we selected some images from severely damaged murals that cannot be used for training or simulated testing. These images are cropped into 105 sub-images of size  $256 \times 256$ , which are used for a real damage restoration experiment.

### Simulated data experiment

In this section, we present our experiments on simulated damaged data. We compare our method with three advanced image restoration approaches—AOT<sup>28</sup>, EdgeConnect<sup>29</sup>, and LaMa<sup>30</sup>—using the Structural Similarity Index Measure (SSIM) and Learned Perceptual Image Patch Similarity (LPIPS) as quantitative evaluation metrics. In our method, the backbone network for noise prediction is based on the network proposed by Dhariwal et al.<sup>25</sup>. Our model is trained with 1000 diffusion steps, and the sampling steps are also set to 1000 during restoration, with a guidance weight  $\gamma$  of 0.001 and a learnable variance. For AOT, EdgeConnect, and LaMa, apart from resizing the images to  $256 \times 256$ , we keep the default hyperparameters as defined in their respective open-source implementations.

Our method, being unsupervised, does not rely on masks during training. For LaMa, as its mask generation strategy is integral to its algorithm, we use its proprietary mask generation strategy. For AOT and EdgeConnect, we use irregular masks from Liu et al.<sup>31</sup>. To ensure fairness and evaluate the models' generalization to different mask types, we test with stroke-like masks from Yu et al.<sup>32</sup>, which none of the models have encountered during training.

Figure 3 shows a visual comparison of six examples from the four methods, while Table 1 provides the objective evaluation metrics for the images in Fig. 3. In Fig. 3, the first column shows the ground truth, the second column the masked image, the third column AOT's results, the fourth column EdgeConnect's results, the fifth column LaMa's results, and the sixth column our method's results. In Table 1, the numbers in the first column correspond to the image numbers in Fig. 3. Each row displays the SSIM and LPIPS restoration results of our method and three comparison approaches.

### Real data experiment

In this subsection, we conduct restoration experiments on real damaged murals. It should be noted that real damaged murals do not have a ground truth for quantitative evaluation, and any feasible evaluation methods inherently involve subjectivity. We evaluate the model's performance through voting and collect voting results from 100 volunteers on the

Internet. The voting process follows an exclusive voting system, where participants are asked to select the one they consider the best in each group of restoration results. Figure 4 shows visual comparisons of restoration results for six real damaged murals, and Fig. 5 presents the corresponding voting results.

### Diverse outputs

In this subsection, we present a unique feature that previous methods lack: non-deterministic restoration results. In existing approaches, once a model is trained, the restoration outcome for a particular mural image is fixed; no matter how many times the restoration process is repeated, the result remains the same. This implies that if the restoration fails once, the model inherently struggles with that image, and adjustments to the dataset and training process, followed by retraining, are required—an expensive and time-consuming task. Our method, however, can produce different restoration results by simply setting different random seeds, without the need for dataset adjustments or retraining. If one restoration attempt does not achieve the desired outcome, multiple attempts can be made, and the best result can be selected from these outputs. Additionally, the diverse outputs generated by our method can offer restoration experts valuable references and inspiration, thereby aiding in the manual restoration of murals.

The idea behind this functionality is that, although our method is applied to a restoration task, it essentially performs image synthesis. The generated results remain consistent with the reference image in the known (intact) regions, achieving restoration indirectly. However, in the unknown (damaged) regions, no direct constraints are applied, resulting in varied outcomes.

Figure 6 illustrates restoration results with two different random seeds using our method. The first column shows the ground truth, the second column shows the masked image, and the third and fourth columns display restoration results with different random seeds. Table 2 presents the statistical information of the metrics for 100 restoration attempts corresponding to the four examples shown in Fig. 6. The “Max” represents the maximum value, the “Min” represents the minimum value, the “Avg” denotes the average value, and the “Var” refers to the variance.

### Unmarkable degradations

Murals exhibit not only continuous damage but also subtle, dense degradation types that cannot be explicitly marked. Existing methods require marking the damaged area to create a mask for restoration, which limits their capability in unmarkable cases. Our method, however, can eliminate this type of degradation without requiring any marking.

As described in Section “Proposed Lossless Image-Guided Sampling Algorithm”, our approach allows obtaining a coarse  $x_0(x_t, t)$  at each step of the restoration process. By interrupting the sampling process at an

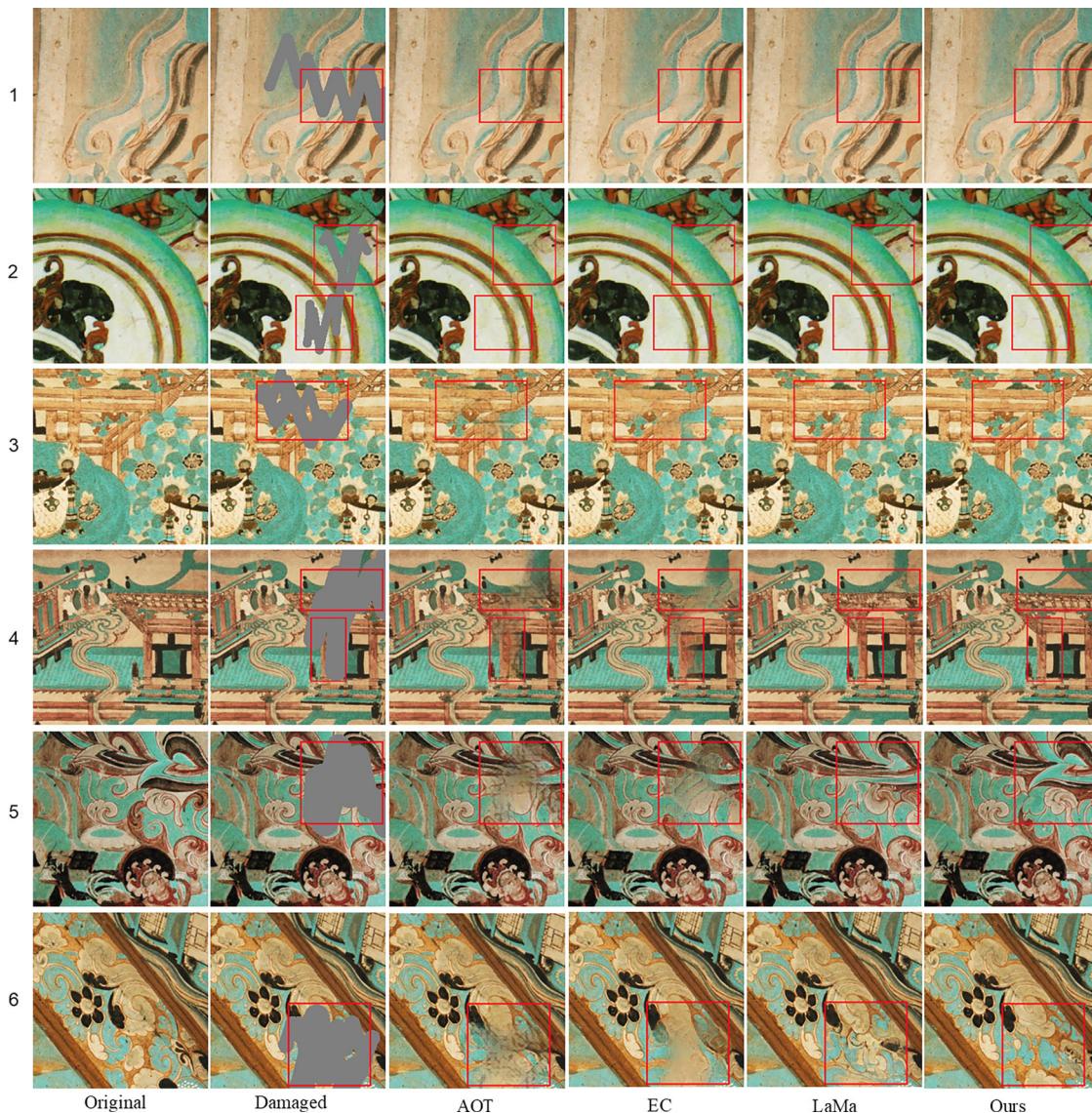


Fig. 3 | Visual comparison of simulated data experiment.

Table 1 | SSIM and LPIPS of simulated data experiment

Murals	SSIM ↑				LPIPS ↓			
	AOT	EC	LaMa	Ours	AOT	EC	LaMa	Ours
1	0.93	<b>0.94</b>	<b>0.94</b>	<b>0.94</b>	0.0291	0.0220	<b>0.0187</b>	0.0192
2	0.96	0.96	<b>0.97</b>	<b>0.97</b>	0.0151	0.0116	0.0090	<b>0.0089</b>
3	0.92	0.93	0.93	<b>0.94</b>	0.0457	0.0433	0.0282	<b>0.0217</b>
4	0.86	0.86	0.88	<b>0.89</b>	0.0872	0.0799	0.0489	<b>0.0444</b>
5	0.85	0.85	0.85	<b>0.87</b>	0.0742	0.0723	0.0600	<b>0.0375</b>
6	0.84	<b>0.85</b>	<b>0.85</b>	<b>0.85</b>	0.0738	0.0777	0.0661	<b>0.0658</b>

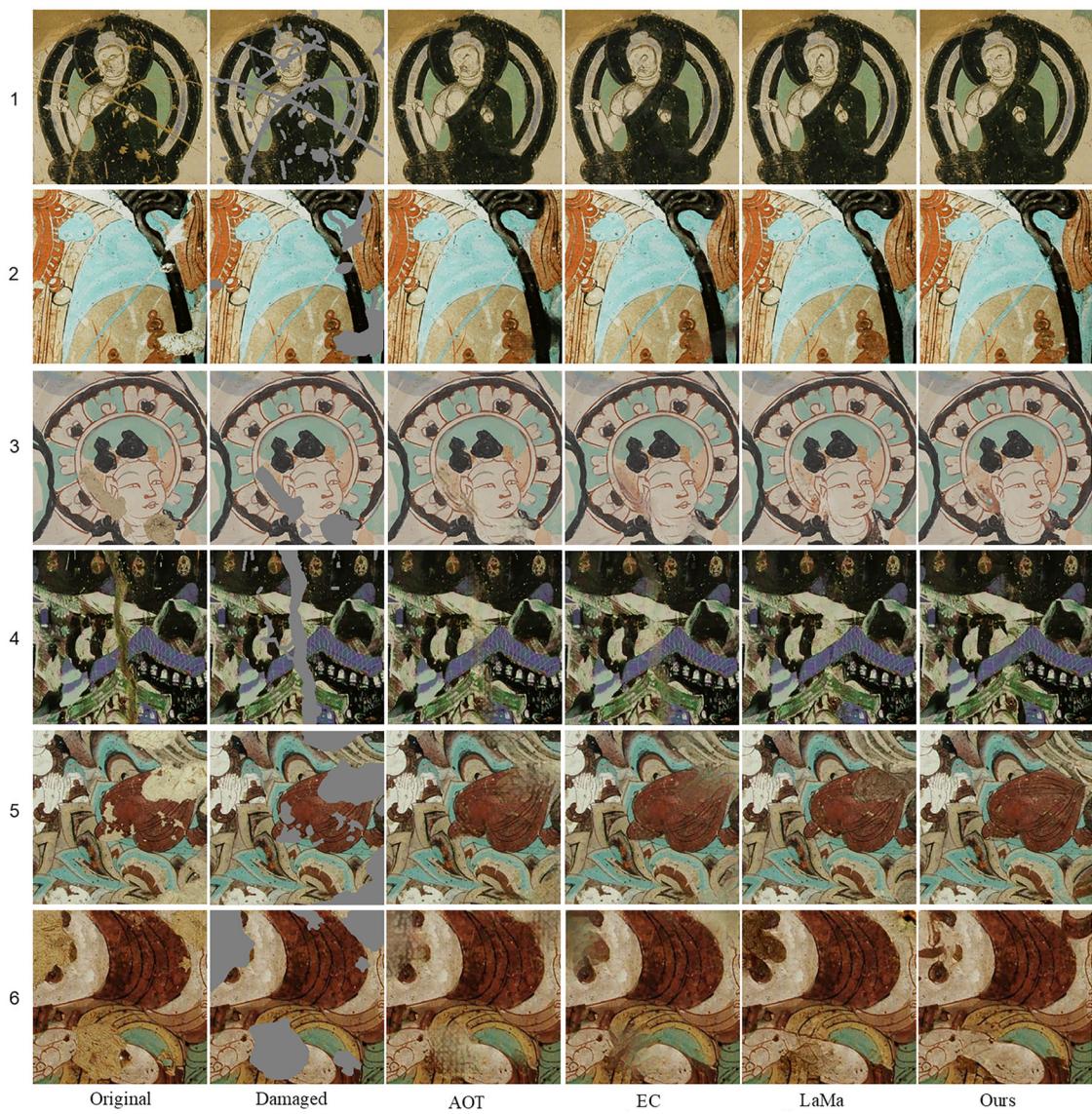
Bold values indicate the best performance among all methods.

intermediate step—rather than completing the full sampling procedure—we can directly generate  $x_0(x_b, t)$  at this stage, achieving degradation removal. The underlying idea is that  $x_0(x_b, t)$  is obtained by guiding the known region. This guiding process first restores the overall outline and then recovers the local details. By interrupting the process at an appropriate step, we obtain a result with only partial guidance, discarding certain details. For these degraded mural images, the details discarded by this interruption are exactly those representing the degradation.

Figure 7 shows the effect of interrupting the sampling process at  $t = 249$ . In the figure, the first row displays the original images, while the second row shows the modified outputs. The modified results appear smoother and clearer than the original, with fewer irrelevant details, highlighting our method's capability to eliminate such degradation without a mask.

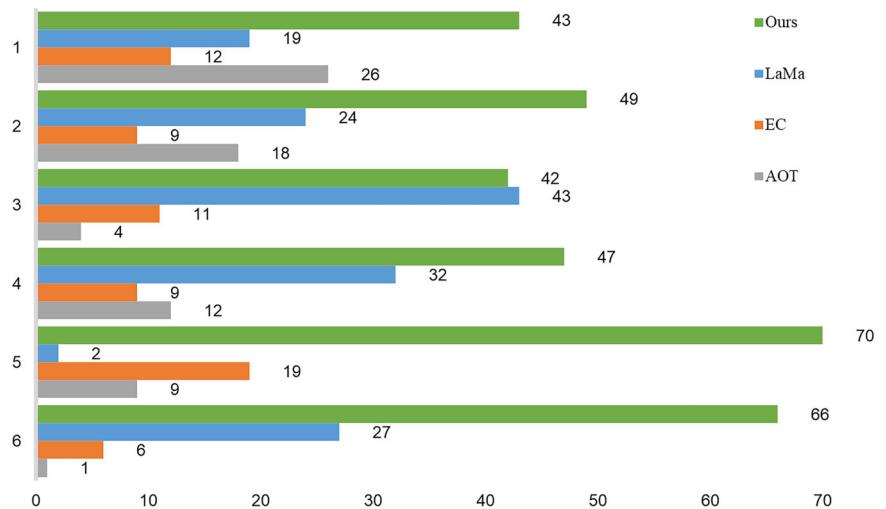
#### Time consumption

Our method offers superior restoration quality as compared to existing approaches, at the cost of higher time consumption. Unlike existing deep learning approaches, our method is not an end-to-end restoration process; it requires multiple iterations to produce the final result, leading to higher computational costs. Moreover, for images with unsatisfactory restoration results, our method can use different random seeds to generate multiple restorations. This will increase time consumption. We attempt to reduce time consumption by decreasing the number of sampling steps. However, this will lead to insufficient guidance and cause the output inconsistent with the reference image in the known (intact) regions. This is not an ideal solution. Table 3 presents the average time cost per image for restoring 100 images of size  $256 \times 256$  on an NVIDIA RTX 3090.



**Fig. 4** | Visual comparison of real data experiment.

**Fig. 5** | Voting results of real data experiment.



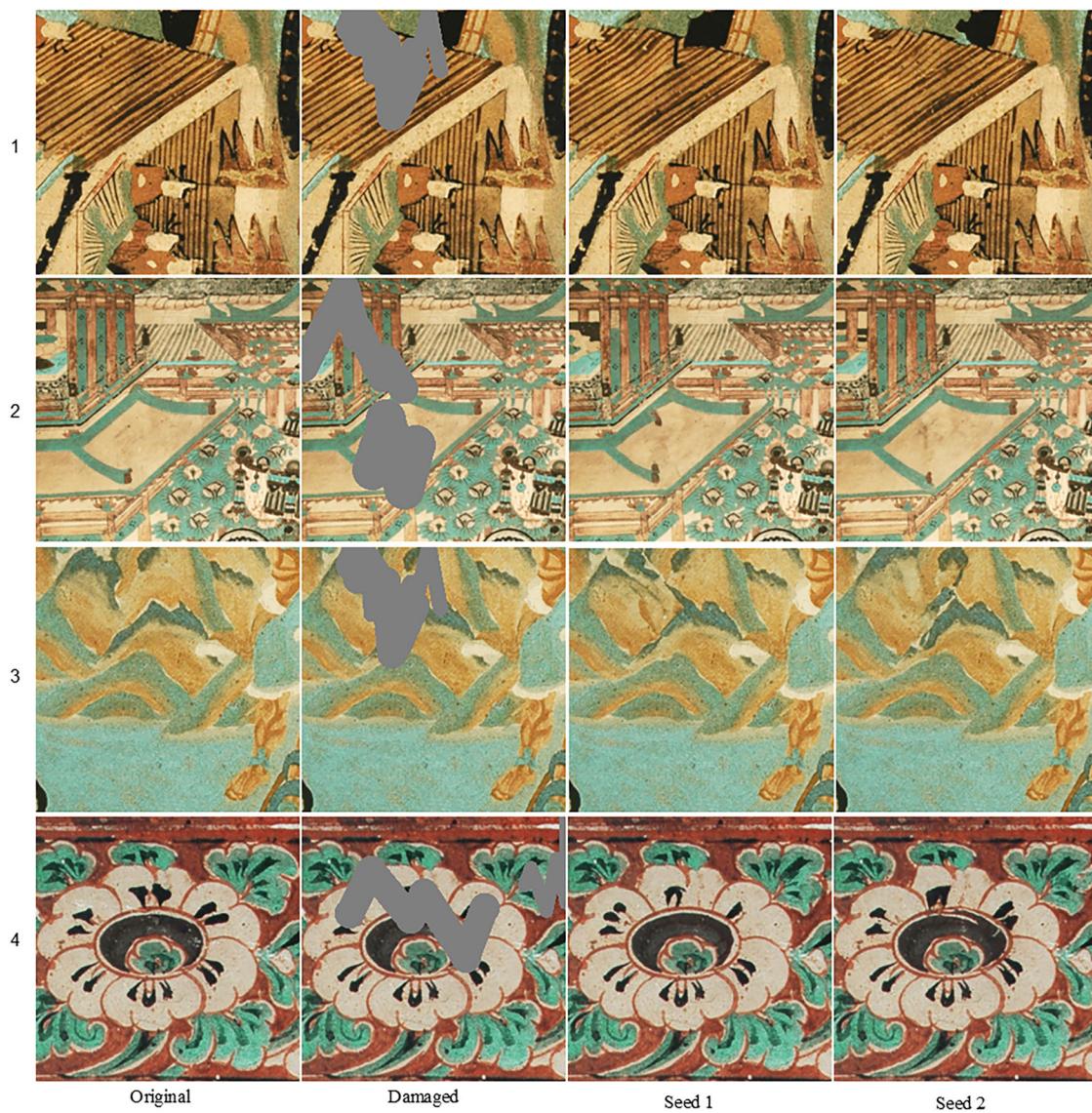


Fig. 6 | Results with different random seeds.

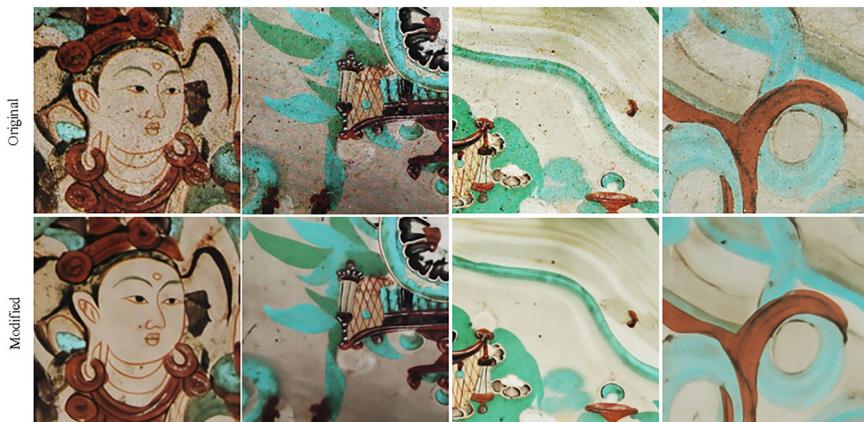
Table 2 | Statistical information of 100 restoration attempts

Murals	SSIM ↑				LPIPS ↓			
	Min	Max	Avg	Var	Min	Max	Avg	Var
1	0.920	0.952	0.935	$4.60 \times 10^{-5}$	0.0135	0.0519	0.0338	$8.13 \times 10^{-5}$
2	0.883	0.916	0.901	$5.07 \times 10^{-5}$	0.0209	0.0634	0.0385	$1.06 \times 10^{-4}$
3	0.910	0.921	0.916	$6.78 \times 10^{-6}$	0.0319	0.0683	0.0461	$3.82 \times 10^{-5}$
4	0.885	0.916	0.903	$3.65 \times 10^{-5}$	0.0325	0.0562	0.0414	$2.53 \times 10^{-5}$

## Discussion

In Fig. 3, rows 1 and 2 feature simple structures and smaller damaged areas. In the first row, all four methods restore the left part of the rectangular region, where color transitions are simple, effectively. However, in the middle part of the two black curves on the right, AOT and EdgeConnect encounter color and edge blurring issues, with AOT being particularly affected. LaMa and our method restore this area well, reflected in superior evaluation metrics. In the second row, all four models perform well in the lower rectangular region, but AOT and EdgeConnect fail to recover the red curve in the top rectangle, while LaMa and our method do so effectively. Metrics in Table 1 also show LaMa and our method outperforming the

others. Rows 3 and 4 represent architectural images with complex lines and extensive damage, the third row shows that AOT and EdgeConnect approximate the outline of the damaged region but suffer from significant blurring and detail loss. LaMa does not experience this level of blurring, but its ability to recover detail is inferior to ours. Our method outperforms in terms of the evaluation metrics as well. In the fourth row, AOT and EdgeConnect fail entirely to restore a basic outline. LaMa restores a rough outline but fails in reconstructing finer details, whereas our method successfully restores fine details, again achieving superior performance on the evaluation metrics. Rows 5 and 6 representing intricate structures with complex curves and extensive damage; both AOT and EdgeConnect fail to

Fig. 7 | Visual results interrupted at  $t = 249$ .**Table 3 | Average time cost per image**

	<b>AOT</b>	<b>EC</b>	<b>LaMa</b>	<b>Ours</b>
Time(s)	0.0449	0.0666	0.226	162

restore these examples. LaMa's results display blurred edges, causing the restored patterns to blend and lose definition, which compromises the finer details. In contrast, our method achieves superior detail restoration, preserving the complex structures without blurring or blending issues. Corresponding evaluation metrics also show that our method performs best.

In Fig. 4, rows 1 to 3 representing cases with smaller damaged areas. In the first row, the damage consists of thin, elongated scratches. Visually, all four methods perform well in restoration, but in the voting results shown in Fig. 5, our method wins with 43 votes. The second and third rows represent typical small-area damage. Both AOT and EdgeConnect exhibit color and edge blurring, with EdgeConnect performing better. LaMa and our method produce more satisfactory restoration results. In the second row, our method wins with 49 votes, while in the third row, LaMa slightly leads with 43 votes, compared to our method's 42 votes. Rows 4 to 6 representing cases with larger damaged areas. In the fourth row, the damage consists of a large crack. AOT and EdgeConnect do not exhibit blurring in this case, but noticeable artifacts remain in the restored regions. LaMa and our method produce better restoration results, with our method leading the vote with 47 votes, while LaMa follows with 32 votes. The fifth and sixth rows show common large-scale damage, where AOT and EdgeConnect fail to restore the image, resulting in severe blurring of edges and colors. LaMa's restoration is also unsatisfactory, with noticeable artifacts. In the voting, our method takes a clear lead, with 70 votes in the fifth row and 66 votes in the sixth.

In Table 2, since the variance of SSIM is on the order of  $10^{-5}$  to  $10^{-6}$ , and the variance of LPIPS is on the order of  $10^{-4}$  to  $10^{-5}$ , the stability of restoration outcomes using different random seeds is acceptable. Note, the metrics used to evaluate the outcomes indicate the restoration quality, but they are not directly related to the diversity. Figure 6 also demonstrates that, even with a low variance, the restoration results still exhibit diversity.

In Table 3, it can be seen that our method requires much more time consumption compared to the other three approaches, with our method requiring 162 s to restore an image, while the other three methods take less than one second. Nonetheless, in a task of cultural heritage preservation, restoration quality typically outweighs time constraints.

Overall, for simpler structures and smaller damaged areas, all models achieve satisfactory restoration, with LaMa and our method performing better than AOT and EdgeConnect. For more extensive and structurally complex damage, AOT and EdgeConnect struggle to restore the image, often failing entirely in highly detailed areas. LaMa performs better by

reconstructing basic shapes but falls short in fine details. Our method, however, stands out by effectively recovering intricate details and structures, leading to the best performance across these challenging cases. The experimental results also show that the stability of the outputs is acceptable. Although our method takes more time, quality outweighs time constraints in a task of cultural heritage preservation, and the longer time spent is acceptable.

### Data availability

The datasets used and/or analyzed in the current study are available from the corresponding author by reasonable request.

### Code availability

The code used and/or analyzed in the current study is available from the corresponding author by reasonable request.

Received: 11 December 2024; Accepted: 7 April 2025;

Published online: 22 April 2025

### References

1. Yang, X. & Wang, S. Dunhuang mural inpainting in intricate disrepairs region based on improvement of priority algorithm. *J. Comput. Aided Des. Comput. Graph.* **23**, 284–289 (2011).
2. Criminisi, A., Pérez, P. & Toyama, K. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Process.* **13**, 1200–1212 (2004).
3. Li, C., Wang, H., Wu, M. & Pan, S. Automatic recognition and virtual restoration of mud spot disease of tang dynasty tomb murals image. *Comput Eng. Appl.* **52**, 233–6 (2016).
4. Jiao, L., Wang, W., Li, B. & Zhao, Q. Wutai mountain mural inpainting based on improved block matching algorithm. *J. Comput. Aided Des. Comput. Graph* **31**, 119–25 (2019).
5. Wang, H., Li, Q. & Jia, S. A global and local feature weighted method for ancient murals inpainting. *Int. J. Mach. Learn. Cybern.* **11**, 1197–1216 (2020).
6. Cao, J., Li, Y., Zhang, Q. & Cui, H. Restoration of an ancient temple mural by a local search algorithm of an adaptive sample block. *Herit. Sci.* **7**, 39 (2019).
7. Jingni, S., Huiqin, W., Meng, W. & Wenzong, Y. Tang dynasty tomb murals inpainting algorithm of mca decomposition. *J. Front. Comput. Sci. Technol.* **11**, 1826 (2017).
8. Jaidilert, S. & Farooque, G. Crack detection and images inpainting method for Thai mural painting images. In: *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, pp. 143–148 (IEEE, 2018).
9. Shen, J. & Chan, T. F. Mathematical models for local nontexture inpaintings. *SIAM J. Appl. Math.* **62**, 1019–1043 (2002).

10. Chan, T. F. & Shen, J. Nontexture inpainting by curvature-driven diffusions. *J. Vis. Commun. Image Represent.* **12**, 436–449 (2001).
11. Chen, Y. A lattice-boltzmann method for image inpainting. In: *2010 3rd International Congress on Image and Signal Processing*, vol. 3, pp. 1222–1225 (IEEE, 2010).
12. Zhang, J., Zhao, D. & Gao, W. Group-based sparse representation for image restoration. *IEEE Trans. Image Process.* **23**, 3336–3351 (2014).
13. Chen, Y., Ai, Y. & Guo, H. Inpainting algorithm for Dunhuang mural based on improved curvature-driven diffusion model. *J. Comput. Aided Des. Comput. Graph.* **32**, 787–796 (2020).
14. Yu, T. et al. End-to-end partial convolutions neural networks for Dunhuang grottoes wall-painting restoration. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 0–0 (2019).
15. Cao, J., Zhang, Z., Zhao, A., Cui, H. & Zhang, Q. Ancient mural restoration based on a modified generative adversarial network. *Herit. Sci.* **8**, 1–14 (2020).
16. Goodfellow, I. et al. Generative adversarial nets. *Advances in neural information processing systems* **27** (2014).
17. Wang, N., Wang, W., Hu, W., Fenster, A. & Li, S. Thanka mural inpainting based on multi-scale adaptive partial convolution and stroke-like mask. *IEEE Trans. Image Process.* **30**, 3720–3733 (2021).
18. Ciortan, I.-M., George, S. & Hardeberg, J. Y. Colour-balanced edge-guided digital inpainting: applications on artworks. *Sensors* **21**, 2091 (2021).
19. Li, L. et al. Line drawing guided progressive inpainting of mural damages. arXiv preprint arXiv:2211.06649 (2022).
20. Schmidt, A., Madhu, P., Maier, A., Christlein, V., Kost, R. Arin: adaptive resampling and instance normalization for robust blind inpainting of Dunhuang cave paintings. In: *2022 Eleventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 1–6 (IEEE, 2022).
21. Sun, W. & Chen, Z. Learned image downscaling for upscaling using content adaptive resampler. *IEEE Trans. Image Process.* **29**, 4027–4040 (2020).
22. Chen, L., Lu, X., Zhang, J., Chu, X., Chen, C. Hinet: Half instance normalization network for image restoration. In: *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 182–192 (2021).
23. 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).
24. Ren, H., Sun, K., Zhao, F. & Zhu, X. Dunhuang murals image restoration method based on generative adversarial network. *Herit. Sci.* **12**, 39 (2024).
25. Dhariwal, P. & Nichol, A. Diffusion models beat GANs on image synthesis. *Adv. Neural Inf. Process. Syst.* **34**, 8780–8794 (2021).
26. Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. *Adv. neural Inf. Process. Syst.* **33**, 6840–6851 (2020).
27. Liu, X. et al. More control for free! image synthesis with semantic diffusion guidance. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 289–299 (2023).
28. Zeng, Y., Fu, J., Chao, H. & Guo, B. Aggregated contextual transformations for high-resolution image inpainting. *IEEE Trans. Vis. Comput. Graph.* **29**, 3266–3280 (2022).
29. Nazeri, K., Ng, E., Joseph, T., Qureshi, F., Ebrahimi, M. Edgeconnect: Structure guided image inpainting using edge prediction. In: *Proc. of the IEEE/CVF International Conference on Computer Vision Workshops* pp. 0–0 (2019).
30. Suvorov, R. et al. Resolution-robust large mask inpainting with Fourier convolutions. In: *Proc. of the IEEE/CVF Winter Conference on Applications of Computer Vision*. pp. 2149–2159 (2022).
31. Liu, G. et al. Image inpainting for irregular holes using partial convolutions. In: *Proc. of the European Conference on Computer Vision (ECCV)*. pp. 85–100 (2018).
32. Yu, J. et al. Free-form image inpainting with gated convolution. In: *Proc. of the IEEE/CVF International Conference on Computer Vision*. pp. 4471–4480 (2019).

## Acknowledgements

This research was supported by the National Natural Science Foundation of China (Grant Nos. 62166048, 61263048), by the Applied Basic Research Project of Yunnan Province (Grant No. 2018FB102).

## Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Ying Yu.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025