

Wavelet-Diffusion Collaborative Framework for Ancient Mural Restoration

Shengdong zhang

Shaoxing University

Jlaxin Lai

Nanfang College Guangzhou

Ting Hu

Shaoxing University

Zhe Yuan

yuanzhe@bfsu.edu.cn

Beijing International Studies University

Wenwen Sun

Shaoxing University

Article

Keywords: Mural Restoration, Diffusion Model, Frequency Domain, Wavelet Transform

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11216004/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Wavelet-Diffusion Collaborative Framework for Ancient Mural Restoration

Shengdong Zhang¹, Jiaxin Lai², Ting Hu¹, Zhe Yuan^{3*},
Wenwen Sun^{1*}

¹School of Intelligent Engineering, Shaoxing University, Chengnan Street, Shaoxing, 312000, Zhejiang, China.

²School of Engineering, Nanfang College, Wenquan Street, Guangzhou, 510970, Guangdong, China.

³The Country and Area Studies Academy, Beijing Foreign Studies University, West Third Ring Road North, Haidian, 100089, Beijing, China.

*Corresponding author(s). E-mail(s): yuanzhe@bfsu.edu.cn;
351947003@qq.com;

Contributing authors: shengdong1986@gmail.com;
laijx240038@stu.nfu.edu.cn; 25010812113@usx.edu.cn;

Abstract

Ancient mural paintings, as invaluable carriers of Chinese cultural heritage, suffer from various degradations including physical damage, stains, color fading, and texture aging. Traditional manual restoration is time-consuming, subjective, and prone to secondary damage. Existing deep learning approaches for ancient painting restoration suffer from two inherent limitations. Wavelet-based methods excel at frequency decomposition and noise suppression yet fundamentally lack the generative capacity to synthesize plausible content for large damaged regions. Conversely, diffusion models demonstrate powerful generative completion capabilities but incur prohibitive computational costs and substantial memory overhead when directly applied to high-resolution mural paintings. To address these challenges, this paper proposes a Wavelet-Diffusion Collaborative Restoration Framework (WDCRF). The proposed method first performs wavelet decomposition on the damaged painting to separate low-frequency structural components from high-frequency detail components. A conditional diffusion model is then applied to restore all four wavelet coefficients directly in the wavelet domain, where the damaged coefficients serve as the conditioning signal.

This design leverages the generative power of diffusion models while operating in the spatially downsampled wavelet representation to reduce computational cost. Finally, the restored coefficients are synthesized via inverse wavelet transform to obtain the complete restoration result. Extensive experiments on the Dunhuang Mural dataset demonstrate that our method achieves PSNR of 32.04 dB and SSIM of 0.922, outperforming existing approaches including EdgeConnect, WaveDiff, PGRDiff, and LRDiff on the DhMurals dataset, our method also achieves the best overall PSNR of 18.95 dB and SSIM of 0.7598. The results validate that operating diffusion models in the wavelet domain effectively balances structural coherence and texture fidelity for ancient painting restoration.

Keywords: Mural Restoration, Diffusion Model, Frequency Domain, Wavelet Transform

1 Introduction

Ancient mural paintings represent an essential component of our cultural heritage, documenting centuries of artistic evolution, social customs, and aesthetic sensibilities. From the bustling scenes of *Along the River During the Qingming Festival* to the ethereal landscapes of *Dwelling in the Fuchun Mountains*, these masterpieces carry irreplaceable historical and artistic value. However, their digital restoration has long been constrained by severe structural fragmentation and chromatic degradation caused by material aging, environmental erosion, and historical damage [1, 2]. The inherent fragility of traditional materials—rice paper, silk, and mineral pigments—combined with environmental factors such as humidity fluctuation, light exposure, microbial erosion, and air pollution, has led to widespread degradation. According to statistics from cultural heritage authorities, approximately 30% of the estimated two million ancient Chinese paintings and calligraphy works exhibit varying degrees of aging and damage, necessitating urgent restoration efforts.

Traditional manual restoration relies heavily on the expertise and experience of conservators, requiring decades of systematic training [1–4]. A single senior restorer can typically complete only a few works annually, making it impossible to address the massive backlog of damaged artifacts. Furthermore, manual intervention carries the inherent risk of irreversible secondary damage due to the delicate nature of the materials. Digital non-destructive restoration has consequently emerged as a promising alternative, offering non-contact operation, traceable processes, reproducible results, and high precision.

Inspired by the success of deep learning in natural image processing tasks such as image dehazing [5, 6], line detection [7, 8], and image inpainting [9, 10], a substantial body of research has been dedicated to developing deep learning-based mural restoration methods [11–13]. Several specialized approaches have been proposed to address the unique challenges of mural and ancient painting restoration. Among them, deformable convolution-based GAN approaches employ a two-stage coarse-to-fine strategy to handle large-area irregular damage, achieving notable improvements in

structural reconstruction [13]. Edge-guided methods integrate multi-scale spatial feature fusion with explicit edge enhancement mechanisms to mitigate the low contrast and blurred edges common in deteriorated murals [12]. Line drawing-guided progressive inpainting frameworks divide the restoration process into structure reconstruction and color correction stages, leveraging structural sketches as priors to ensure content authenticity and structural stability [11]. Despite their respective contributions, these methods share several inherent limitations. First, they universally rely on explicit structural priors—whether deformable convolutions, edge maps, or line drawings—to guide the restoration process. The extraction of such priors is itself a challenging task for severely damaged or stylistically complex artworks, and any inaccuracies in this initial step propagate through the subsequent stages, compromising the final output. Second, structure recovery and color synthesis are typically treated as sequential or loosely coupled processes, lacking a unified mechanism to simultaneously optimize both aspects. Consequently, these methods often struggle with the coupled problem of structural distortion and color deviation, particularly when restoring large missing regions where semantic reasoning and chromatic coherence must be jointly considered. Third, the computational overhead associated with these multi-stage architectures limits their practical applicability to high-resolution ancient paintings.

Recently, diffusion models have emerged as a powerful generative paradigm for image restoration, demonstrating remarkable capacity for modeling complex data distributions and generating fine-scale details while maintaining global structural coherence [14–16]. Several diffusion-based methods have been proposed for mural restoration, incorporating various guidance mechanisms to steer the generation process. LRDiff integrates low-rank structural priors into the diffusion process to capture non-local similarities within murals [17]. PGRDiff employs prompt-guided residual diffusion to handle diverse degradation types [18]. The edge-constrained diffusion method introduces an attention-guided edge prediction module to provide spatial supervision for subsequent generation, combined with a non-local color optimization module to ensure global color consistency [19]. DiffuMural adopts a multi-stage, condition-guided diffusion framework with multi-scale collaborative diffusion and frequency-domain refinement [20]. Despite these advances, existing diffusion-based methods still operate predominantly in the pixel domain, where the iterative denoising process must simultaneously handle global structural reasoning and local texture synthesis within a high-dimensional space. This entanglement of structure and texture not only incurs prohibitive computational costs but also leads to suboptimal performance, as the model often prioritizes one aspect at the expense of the other. Furthermore, most of these methods still rely on externally extracted priors that are themselves susceptible to error.

To overcome these limitations, we propose a fundamentally different paradigm: *performing the entire diffusion process in the wavelet domain*. Our approach is motivated by three key insights. First, the Discrete Wavelet Transform (DWT) [21] naturally decouples the image into low-frequency components that encode global structure and high-frequency components that capture local textures and edges. This explicit frequency decomposition resolves the structure-texture entanglement problem inherent in pixel-domain diffusion, allowing the model to handle global coherence and local fidelity

in a more targeted manner. Second, the spatial dimensions of each subband are reduced by a factor of two along each axis, so the diffusion model operates on a spatially down-sampled wavelet representation. This compression substantially lowers computational cost and memory footprint, making high-resolution restoration practical. Third, unlike latent-space compression methods, the wavelet transform is invertible and lossless, ensuring that no information is discarded during the decomposition-reconstruction cycle.

Specifically, given a damaged ancient painting, we first apply a single-level 2D Discrete Wavelet Transform (2D-DWT) to decompose the input into one low-frequency subband and three high-frequency subbands. The four subbands are then jointly restored by a conditional diffusion model directly in the wavelet domain, where the damaged wavelet coefficients serve as the conditioning signal to guide the generative process. After the reverse diffusion process, the restored wavelet coefficients are synthesized back to the spatial domain via inverse DWT to obtain the final restored painting. This frequency-first design offers three principal advantages over existing approaches: (1) it explicitly separates global structure and local texture into different frequency subbands, allowing the diffusion model to reconstruct them in a unified and frequency-aware manner; (2) it performs the iterative denoising process on subbands whose spatial resolution is halved, thereby reducing the spatial dimension of each denoising step; and (3) it eliminates the need for externally extracted structural priors, relying instead on the self-conditioning mechanism provided by the damaged wavelet coefficients themselves.

Extensive experiments on the Dunhuang Mural dataset and the DhMurals dataset demonstrate that our method achieves superior performance over state-of-the-art approaches in terms of both quantitative metrics and visual quality, while maintaining a favorable balance between restoration fidelity and computational efficiency. The results validate that operating diffusion models in the wavelet domain provides a practical and effective solution for the challenging task of ancient painting restoration.

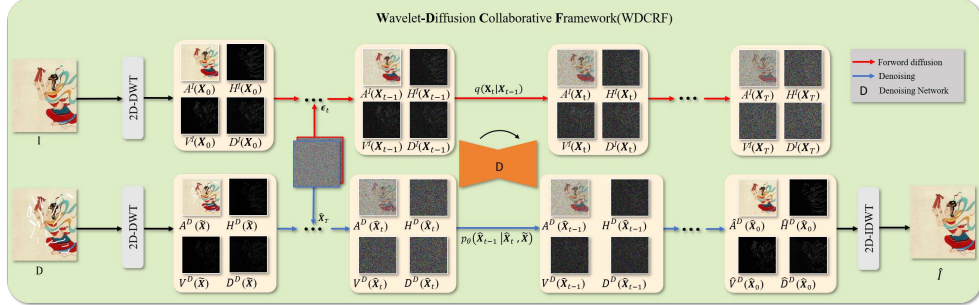


Fig. 1 Architecture of the proposed Wavelet-Diffusion Collaborative Framework (WDCRF) for Mural Restoration.

2 Methodology

The overall pipeline of our proposed WDCRF is illustrated in Fig. 1. It comprises three main stages: **Wavelet Decomposition**, **Wavelet-Domain Conditional Diffusion**, and **Wavelet Reconstruction**. In the **Wavelet Decomposition** stage, the damaged input painting I is first transformed into the wavelet domain using a single-level 2D Discrete Wavelet Transform (2D-DWT). This decomposition yields one low-frequency subband A and three high-frequency subbands, denoted as V , H , and D . In the subsequent **Wavelet-Domain Conditional Diffusion** stage, a conditional diffusion model is applied to restore the low-frequency wavelet coefficients $\mathbf{z}_0 = A$, which operates directly on the spatially downsampled wavelet representation. The damaged coefficients $\mathbf{c} = \{A_{\text{damaged}}, V_{\text{damaged}}, H_{\text{damaged}}, D_{\text{damaged}}\}$ serve as the conditioning signal, effectively guiding the restoration process. The high-frequency subbands are then refined by the HFEM. Finally, in the **Wavelet Reconstruction** stage, the restored coefficients $\hat{\mathbf{z}}_0 = \{\hat{A}, \hat{V}, \hat{H}, \hat{D}\}$ are synthesized back to the spatial domain via the inverse DWT (IDWT), yielding the final restored painting $\hat{I}$.

2.1 Motivation

Diffusion models have achieved remarkable success in image generation and restoration, yet their direct application to ancient painting restoration is constrained by three fundamental limitations.

Prohibitive Computational Cost. Standard diffusion models require hundreds to thousands of iterative denoising steps, each operating at full spatial resolution. For high-resolution ancient paintings, this incurs substantial GPU memory consumption and inference latency. While latent diffusion models [22] partially mitigate this issue by operating in a compressed latent space, the encoding-decoding process inevitably discards fine textures and subtle color gradations—information that is critical for preserving artistic authenticity.

Entangled Structure-Texture Learning. Pixel-domain diffusion models treat the image as a monolithic entity, compelling the network to simultaneously learn global semantic reasoning and local detail synthesis. This entanglement imposes a fundamental trade-off: the model either prioritizes structural coherence at the expense of texture fidelity, producing over-smoothed outputs, or focuses on local details while compromising structural integrity, resulting in semantic inconsistencies. For ancient paintings, where both holistic composition and intricate brushwork are equally vital, this compromise is particularly detrimental.

Style Deviation and Color Shift. As documented in recent studies [17, 18], diffusion models trained on general-purpose image datasets exhibit color shift and style deviation when applied to cultural heritage images. This phenomenon stems from two factors: the domain gap between modern photographic data and ancient artworks, and the model’s inherent tendency to favor statistically prevalent color distributions over the unique, aged palettes that characterize historical paintings.

To overcome these limitations, we propose a **Wavelet-Diffusion Collaborative Restoration Framework (WDCRF)**. Our approach is built upon a fundamental

paradigm shift: *perform restoration in the wavelet domain, where structural information is compact and the generative process is naturally decomposed*. The Discrete Wavelet Transform (DWT) offers three compelling advantages:

1. **Dimensionality Reduction:** DWT reduces the spatial dimensions of each sub-band by a factor of two along each axis, so the diffusion model operates on a spatially downsampled representation, lowering the spatial dimension of each denoising step.
2. **Frequency Decomposition:** DWT explicitly separates the image into low-frequency (structural/semantic) and high-frequency (textural/edge) components, allowing the diffusion model to handle these distinct aspects in a decoupled manner and resolving the structure-texture entanglement problem.
3. **Information Preservation:** Unlike latent-space compression methods, DWT is invertible and lossless—no information is discarded. The compact representation is a reorganization of the original signal, ensuring that all details remain accessible throughout the restoration process.

2.2 Wavelet Decomposition

Given a damaged ancient painting image $I \in \mathbb{R}^{H \times W \times 3}$, we first perform a single-level 2D Discrete Wavelet Transform (2D-DWT) to decompose the input into four frequency subbands:

$$A, V, H, D = \text{2D-DWT}(I), \quad (1)$$

where:

- $A \in \mathbb{R}^{H/2 \times W/2 \times 3}$ is the low-frequency approximation coefficient, encapsulating the global structure, semantic content, and smooth color gradients of the painting.
- $V \in \mathbb{R}^{H/2 \times W/2 \times 3}$ is the vertical high-frequency detail coefficient, capturing horizontal edges and textures.
- $H \in \mathbb{R}^{H/2 \times W/2 \times 3}$ is the horizontal high-frequency detail coefficient, capturing vertical edges and textures.
- $D \in \mathbb{R}^{H/2 \times W/2 \times 3}$ is the diagonal high-frequency detail coefficient, capturing diagonal edge information.

The spatial dimensions of all four subbands are reduced by a factor of 2 along each axis. The spatial resolution of each subband is halved along both axes, so the diffusion model operates on a spatially downsampled wavelet representation. This compression substantially decreases the computational burden for the diffusion model while preserving all information through the invertible nature of the wavelet transform.

2.3 Wavelet-Domain Conditional Diffusion

Instead of performing diffusion in the original pixel space, we apply a conditional diffusion model to reconstruct all four wavelet coefficients directly in the wavelet domain. Let $\mathbf{z}_0 = A$ denote the clean low-frequency wavelet coefficients from the ground-truth image, and let $\mathbf{c} = \{A_{\text{damaged}}, V_{\text{damaged}}, H_{\text{damaged}}, D_{\text{damaged}}\}$ denote the corresponding damaged coefficients from the input image. The conditioning signal $\mathbf{c}$ provides the

diffusion model with crucial information about the original painting’s structure and texture, guiding the generation process toward content-consistent restoration.

2.3.1 Forward Diffusion Process

The forward process progressively adds Gaussian noise to the clean wavelet coefficients $\mathbf{z}_0$ over T steps:

$$q(\mathbf{z}_t | \mathbf{z}_{t-1}) = \mathcal{N}(\mathbf{z}_t; \sqrt{1 - \beta_t} \mathbf{z}_{t-1}, \beta_t \mathbf{I}), \quad (2)$$

where $\{\beta_t \in (0, 1)\}_{t=1}^T$ is a pre-defined variance schedule. Using the reparameterization trick, we can directly sample $\mathbf{z}_t$ at any timestep t without iterating through intermediate steps:

$$\mathbf{z}_t = \sqrt{\bar{\alpha}_t} \mathbf{z}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}), \quad (3)$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$. When T is sufficiently large, $\mathbf{z}_T$ approximates an isotropic Gaussian distribution, providing a tractable prior for the reverse process.

2.3.2 Reverse Denoising Process

The reverse process learns to reconstruct the clean coefficients $\mathbf{z}_0$ from the noisy input $\mathbf{z}_t$, conditioned on the damaged coefficients $\mathbf{c}$. The conditional reverse transition is formulated as:

$$p_\theta(\mathbf{z}_{t-1} | \mathbf{z}_t, \mathbf{c}) = \mathcal{N}(\mathbf{z}_{t-1}; \mu_\theta(\mathbf{z}_t, \mathbf{c}, t), \sigma_t^2 \mathbf{I}), \quad (4)$$

where the mean μ_θ is parameterized by a noise prediction network ϵ_θ :

$$\mu_\theta(\mathbf{z}_t, \mathbf{c}, t) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{z}_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(\mathbf{z}_t, \mathbf{c}, t) \right). \quad (5)$$

The noise prediction network ϵ_θ adopts a U-Net architecture that takes as input the concatenation of the noisy coefficients $\mathbf{z}_t$, the conditional coefficients $\mathbf{c}$, and the timestep embedding t . The network is trained to minimize the noise prediction error:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{\mathbf{z}_0, \epsilon \sim \mathcal{N}(0, \mathbf{I}), t} [\|\epsilon - \epsilon_\theta(\mathbf{z}_t, \mathbf{c}, t)\|^2]. \quad (6)$$

2.3.3 Conditioning Strategy

A key design choice in our framework is the conditioning mechanism. Unlike unconditional diffusion models that generate content from pure noise, our model is conditioned on the damaged wavelet coefficients $\mathbf{c}$. This conditioning serves two crucial purposes:

1. **Content Consistency:** The damaged coefficients provide a strong structural prior, ensuring that the generated content remains faithful to the original painting’s composition, even in regions that require extensive reconstruction.
2. **Frequency-Specific Guidance:** By conditioning separately on each frequency subband, the model learns to leverage the correlation between the low-frequency structure and the high-frequency details.

This dense conditioning is implemented through feature concatenation at multiple levels of the U-Net, enabling the model to incorporate the conditional information throughout the entire denoising process.

2.4 Wavelet Reconstruction

After obtaining the restored wavelet coefficients $\hat{\mathbf{z}}_0 = \{\hat{A}, \hat{V}, \hat{H}, \hat{D}\}$ from the diffusion model and the HFEM, we reconstruct the final restored image by applying the inverse Discrete Wavelet Transform (IDWT):

$$\hat{I} = \text{IDWT}(\hat{A}, \hat{V}, \hat{H}, \hat{D}). \quad (7)$$

The final output $\hat{I} \in \mathbb{R}^{H \times W \times 3}$ corresponds to the fully restored painting in the spatial domain. Since both DWT and IDWT are invertible and lossless transformations, no information is discarded during the decomposition-reconstruction cycle beyond the intentional modifications introduced by the diffusion model and the HFEM.

2.5 Training Objective with Frequency-Aware Supervision

In addition to the standard noise prediction loss, we introduce an auxiliary reconstruction loss that directly measures the fidelity of the restored low-frequency wavelet coefficients:

$$\mathcal{L}_{\text{recon}} = \|\hat{\mathbf{z}}_0 - \mathbf{z}_0\|_2^2, \quad (8)$$

where $\hat{\mathbf{z}}_0$ is the denoised output obtained by running the complete reverse trajectory during training. This loss operates in the wavelet domain, providing direct and precise supervision for the low-frequency subband.

The overall training objective is thus a weighted combination:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{diff}} + \lambda \mathcal{L}_{\text{recon}} + \mu \mathcal{L}_{\text{hfem}}, \quad (9)$$

where $\lambda > 0$ and $\mu > 0$ are balancing hyperparameters. This dual-objective design ensures that the model not only learns the generative distribution of clean wavelet coefficients but also maintains faithful reconstruction of the original content.

2.6 Accelerated Inference with DDIM

To overcome the slow sampling speed inherent in standard DDPMs, we adopt the Denoising Diffusion Implicit Model (DDIM) sampler [23] during inference. DDIM formulates the reverse process as a non-Markovian chain that can be sampled with substantially fewer steps:

$$\mathbf{z}_{t-1} = \sqrt{\bar{\alpha}_{t-1}} \hat{\mathbf{z}}_0 + \sqrt{1 - \bar{\alpha}_{t-1} - \sigma_t^2 \epsilon_\theta} + \sigma_t \epsilon, \quad (10)$$

where $\hat{\mathbf{z}}_0 = (\mathbf{z}_t - \sqrt{1 - \bar{\alpha}_t} \epsilon_\theta) / \sqrt{\bar{\alpha}_t}$ is the predicted clean coefficient, and σ_t controls the stochasticity of the sampling process. In practice, we train the diffusion model with

$T = 1000$ steps and reduce the number of sampling steps to $S = 100$ during inference, achieving a 10-fold speedup with negligible quality degradation.

2.7 Comparative Analysis

To situate our contribution within the existing literature, we compare our WDCRF against three representative diffusion-based mural restoration methods: PGRDiff [18], LRDiff [17], and DiffuMural [24].

Pixel-Domain vs. Frequency-Domain Generation.

PGRDiff, LRDiff, and DiffuMural all operate in the pixel domain, compelling the diffusion model to simultaneously handle global structure and local texture within a single high-dimensional space. Our method, in contrast, performs diffusion entirely in the wavelet domain, explicitly decoupling global semantics (low-frequency) from local details (high-frequency) via 2D-DWT. This frequency-aware decomposition allows the model to allocate its representational capacity more efficiently.

Self-Conditioning vs. External Priors.

Existing methods rely on external guidance mechanisms—prompts, low-rank structural priors, or multi-scale conditioning—that are susceptible to inaccuracies and may not generalize across diverse artistic styles. Our method employs the damaged wavelet coefficients themselves as the conditioning signal, constituting a simpler, self-contained mechanism that exploits cross-frequency correlations without requiring external priors or manual engineering.

Compressed-Domain Efficiency.

DiffuMural’s multi-scale approach increases computational cost, LRDiff’s low-rank decomposition offers only limited efficiency gains, and PGRDiff requires full-resolution iterative refinement. Our wavelet-domain operation halves the spatial resolution of each denoising step, accelerating both training and inference while substantially reducing memory consumption. Critically, this efficiency gain does not compromise fidelity, as the wavelet transform is lossless.

Frequency-Aware Loss.

Existing methods rely primarily on pixel-space perceptual losses (such as LPIPS and style loss) that only indirectly influence the generation process. Our auxiliary reconstruction loss $\mathcal{L}_{\text{recon}}$ operates directly in the wavelet domain, providing explicit, per-subband supervision that enables superior reconstruction accuracy while preserving artistic style.

In summary, our wavelet-domain diffusion framework distinguishes itself through four key innovations: (1) frequency decomposition that decouples structure from texture, (2) self-conditioning that eliminates external priors, (3) compressed-domain operation that drastically reduces computational cost, and (4) frequency-aware loss that provides direct supervision. This frequency-first paradigm achieves state-of-the-art restoration performance while offering a computationally practical solution for cultural heritage preservation.

3 Results

This section presents a series of experiments evaluating the performance of mural restoration methods. The dataset and evaluation metrics are first introduced. Next, the implementation of the proposed method is described. Then, quantitative and qualitative comparisons with other methods are provided. Finally, ablation studies are conducted to validate the contribution of each module. In these experiments, we compare with a number of state-of-the-art mural restoration methods, such as LRDiff [17], URDiff [25], PGRDiff [18], EdgeConnect [26], HD-Painter [27], CMAMRNet [28], WaveDiff [16], RDDM [29], and SPA-Former [30].

3.1 Dataset settings and evaluation metrics

Dunhuang Mural dataset [31] has been widely used to train and evaluate the performance of mural restoration methods. Dunhuang Mural has 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 features. These frescoes predominantly illustrate Buddhist scriptures, Jataka tales, and celestial scenes, featuring iconic flying apsaras (Apsarases), musicians, and dancers that vividly reflect the fusion of Indian, Central Asian, and Chinese artistic traditions. The Dunhuang mural dataset is constructed by selecting 600 representative mural images, of which 500 are allocated to training and 100 to testing. All training and testing images are resized to the size of 512×512 pixels as input and the batch size is set to 12.

To further verify the superiority of the proposed method, we conduct another experiment on the Dunhuang Mural Datasets (DhMurals dataset) [11, 32]. This dataset integrates four public sub-datasets, namely MuralDH, Dunhuang-Grottoes-Painting, Dunhuang-Faces, and DhMurals1714, and is specifically designed for mural inpainting research. The dataset is organised into three folders: images, lines, and masks. The images folder contains 512×512 versions of the original images, obtained by resizing or cropping the images from the four sub-datasets. The lines folder stores edge maps extracted from the images using the XDoG (eXtended Difference of Gaussians) operator. The masks folder contains binary masks that mark the damaged regions. The dataset is split into training, validation, and test sets, with 1,564, 100, and 50 images, respectively.

Four metrics—PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Index Measure), LPIPS (Learned Perceptual Image Patch Similarity), and FID (Fréchet Inception Distance)—are employed to assess the restoration performance. PSNR measures pixel-wise error between two images, with higher values indicating less distortion. SSIM evaluates luminance, contrast, and structural similarity, with values closer to 1 meaning better perceptual quality. LPIPS uses deep CNN features to quantify perceptual distance, where lower scores mean images appear more similar to humans. FID compares the feature distributions of real and generated images via Inception-v3, with lower values indicating higher fidelity and diversity.

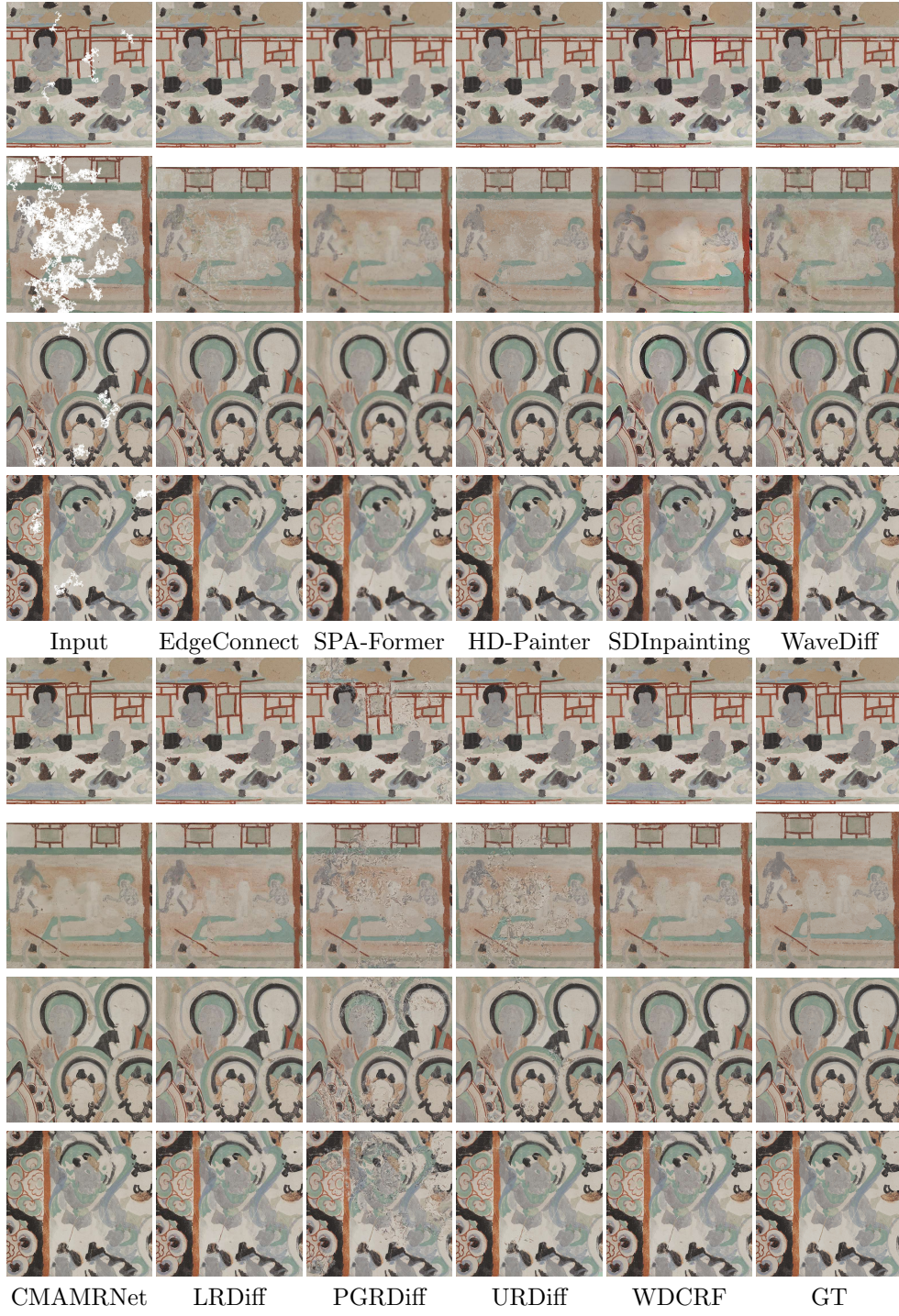


Fig. 2 The restoration results from different methods on the Dunhuang Mural dataset.

Table 1 Comparison of PSNR (dB) under different degradation ratios on the Dunhuang Mural dataset. The best result in each column is highlighted in bold.

Method	0–10%	10–20%	20–30%	Total
WDCRF	34.87	27.96	26.35	32.04
PGRDiff	26.94	25.87	24.08	26.44
URDiff	28.23	23.53	21.86	26.29
CMAMRNet	31.40	29.30	28.77	30.54
WaveDiff	29.88	25.26	23.88	27.98
SDInpainting	23.65	20.59	19.28	22.37
SPA-Former	28.39	26.55	26.04	27.63
LRDiff	30.11	26.15	24.89	28.47

Table 2 Comparison of SSIM under different degradation ratios on the Dunhuang Mural dataset. The best result in each column is highlighted in bold.

Method	0–10%	10–20%	20–30%	Total
WDCRF	0.960	0.872	0.816	0.922
PGRDiff	0.775	0.747	0.718	0.763
URDiff	0.817	0.691	0.607	0.763
CMAMRNet	0.858	0.823	0.809	0.844
WaveDiff	0.860	0.746	0.676	0.811
SDInpainting	0.604	0.507	0.490	0.565
SPA-Former	0.786	0.745	0.745	0.770
LRDiff	0.826	0.743	0.693	0.791

3.2 Implementation details

The model is implemented using the PyTorch deep learning framework. It was trained on a machine equipped with 8 TITAN V GPUs, and is also capable of running effectively on a single GPU. The Adam optimizer was employed for optimization. Following [17], we set the initial learning rate to 0.0001 and applied a Cosine annealing scheduler to gradually decay it to 0.000001, with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The model was trained for 600 epochs. Data augmentation included random horizontal and vertical flipping as well as 90-degree rotations. We set the diffusion timestep T to 1000 during training and use $S = 100$ DDIM sampling steps during inference. To preserve the lightweight nature of the proposed model, we omit the group normalization and self-attention layers from the U-Net architecture [14].

3.3 Quantitative and qualitative results on Dunhuang Mural dataset

Based on the comprehensive quantitative comparisons reported in Tables 1, 2, 3, and 4, we analyze the performance of our proposed WDCRF against several state-of-the-art restoration methods, including PGRDiff, URDiff, CMAMRNet, WaveDiff, Stable Diffusion Inpainting, SPA-Former, and LRDiff. The evaluation is conducted under three degradation levels defined by the proportion of missing pixels in the masked

Table 3 Comparison of LPIPS (lower is better) under different degradation ratios on the Dunhuang Mural dataset. The best result in each column is highlighted in bold.

Method	0–10%	10–20%	20–30%	Total
WDCRF	0.017	0.061	0.084	0.036
PGRDiff	0.135	0.155	0.176	0.144
URDiff	0.099	0.201	0.269	0.143
CMAMRNet	0.049	0.081	0.093	0.062
WaveDiff	0.059	0.155	0.206	0.100
SDInpainting	0.116	0.200	0.231	0.151
SPA-Former	0.241	0.250	0.230	0.243
LRDiff	0.040	0.086	0.106	0.059

Table 4 FID scores on the Dunhuang Mural dataset. Lower is better.

Method	FID
WDCRF	80.221
PGRDiff	836.346
URDiff	807.935
CMAMRNet	147.683
WaveDiff	282.598
SDInpainting	198.245
SPA-Former	5345.183
LRDiff	145.672

input images: 0%–10%, 10%–20%, and 20%–30%. We also report the overall average across all test samples for a comprehensive assessment.

PSNR Analysis. As shown in Table 1, our method achieves the highest total PSNR of **32.04 dB**, outperforming the second-best competitor CMAMRNet (30.54 dB) by a clear margin. More importantly, WDCRF delivers superior performance on the lowest degradation level (0%–10% mask ratio), reaching **34.87 dB**, where it surpasses all baselines by a significant margin—for instance, outperforming CMAMRNet by 3.47 dB and LRDiff by 4.76 dB. Under more challenging degradation levels (10%–20% and 20%–30% mask ratios), CMAMRNet exhibits slightly better results (29.30 and 28.77 dB), while our method still achieves competitive scores of 27.96 and 26.35 dB, ranking among the top three. In contrast, Stable Diffusion Inpainting and WaveDiff show substantial performance drops, particularly at higher degradation ratios, indicating their limited robustness to severe structural damage.

SSIM Analysis. Table 2 reveals that WDCRF achieves the best overall SSIM of **0.922**, clearly surpassing CMAMRNet (0.844) and LRDiff (0.791). Our model excels particularly on mask ratios of 0%–10% and 10%–20%, achieving **0.960** and **0.872** respectively—both being the highest among all compared methods. On the most challenging degradation level (20%–30% mask ratio), WDCRF yields the top score (0.816), while CMAMRNet (0.809) remains competitive. LRDiff and WaveDiff obtain lower

scores of 0.693 and 0.676, respectively. The substantial margin in SSIM underscores the effectiveness of our wavelet-guided diffusion framework in preserving structural coherence and global composition, even when the input is only mildly degraded.

LPIPS Analysis. Since LPIPS measures perceptual dissimilarity, lower values indicate better perceptual quality. Table 3 shows the LPIPS results of the state-of-the-art methods and the proposed WDCRF. The proposed WDCRF stands out with the lowest total LPIPS of **0.036**, marginally outperforming LRDiff (0.059) and CMAMRNet (0.062). Notably, our method achieves the best scores across all individual degradation levels: **0.017** (0%–10% mask ratio), **0.061** (10%–20% mask ratio), and **0.084** (20%–30% mask ratio). This consistent superiority indicates that the generated restorations are not only pixel-wise accurate but also perceptually closer to the ground truth, preserving fine textures and artistic brushstrokes that are critical for ancient mural authenticity. In contrast, SPA-Former and Stable Diffusion Inpainting yield substantially higher LPIPS values (0.243 and 0.151, respectively), reflecting their inability to handle the domain-specific characteristics of mural images.

FID Analysis. As reported in Table 4, our method achieves the lowest FID score of **80.221** among all evaluated models, significantly outperforming CMAMRNet (147.683), LRDiff (145.672), and URDiff (807.935). This demonstrates that WDCRF generates restored images whose feature distribution is substantially closer to that of real pristine murals, validating the generative fidelity of our dual-path restoration strategy. The extremely high FID of SPA-Former (5345.183) further confirms its inadequacy for this task, likely due to its inability to capture the complex statistical patterns inherent in ancient mural imagery.

Overall, the quantitative results consistently demonstrate that our WDCRF achieves state-of-the-art or highly competitive performance across all metrics and degradation scenarios. The method excels in perceptual quality (LPIPS) and distributional fidelity (FID), while maintaining strong structural integrity (SSIM) and competitive reconstruction accuracy (PSNR). These findings substantiate the efficacy of our frequency-decomposed conditional diffusion design, which effectively balances global semantic recovery and local detail refinement, particularly for the challenging task of ancient mural restoration.

Visual Quality Analysis. Beyond quantitative metrics, we conduct a qualitative comparison of the restored images produced by different methods, as illustrated in Fig. 2. The figure presents representative examples from the Dunhuang Mural test set under varying degradation levels, showcasing typical regions with facial features, clothing folds, and background decorative patterns. Our WDCRF consistently generates visually appealing restorations that are semantically coherent and stylistically faithful to the original artwork. In regions with mild degradation (0%–10% mask ratio), our method preserves the sharpness of facial contours and the subtle gradations of pigment colors, closely resembling the ground truth. In contrast, CMAMRNet and LRDiff produce slightly over-smoothed textures, losing the characteristic granularity of ancient mural surfaces, while WaveDiff and Stable Diffusion Inpainting tend to introduce blurring artifacts around the repaired boundaries. For more challenging cases with larger missing areas (20%–30% mask ratio), the advantages of our frequency-decomposed design become even more pronounced. The low-frequency diffusion model successfully

Table 5 Comparison of PSNR (dB) under different degradation ratios on the DhMurals dataset. The best result in each column is highlighted in bold.

Method	0–20%	20–40%	$\geq 40\%$	Total
WDCRF	20.77	18.38	16.67	18.95
EdgeConnect	18.26	11.60	11.35	11.84
SPA-Former	20.68	18.36	16.86	18.31
WaveDiff	20.52	18.40	17.24	18.19
LRDiff	17.22	13.24	12.75	13.35
SDInpainting	17.80	17.63	16.19	17.49
CMAMRNet	20.63	19.08	17.31	18.76
PGRDiff	20.74	18.20	17.23	18.20
RDDM	18.31	13.75	13.93	13.95

reconstructs the global composition and color harmony, preventing the structural distortions that plague SPA-Former and URDiff—the former often generates unnatural facial expressions, while the latter produces discontinuous clothing lines. PGRDiff, although capable of producing plausible global structures, frequently introduces repetitive texture patterns that deviate from the authentic artistic style, indicating its limited ability to capture the unique stylistic prior of Dunhuang murals. Furthermore, our method exhibits superior color fidelity. The conditional diffusion process in the wavelet domain effectively constrains the generative process to adhere to the original color palette, avoiding the color shifts or oversaturation observed in CMAMRNet and WaveDiff. The auxiliary reconstruction loss further anchors the output to the ground-truth distribution, preventing the stochastic sampling from producing anachronistic or visually inconsistent color tones.

In summary, the visual results corroborate the quantitative findings: WDCRF achieves a favorable balance between global structural integrity and local detail fidelity. The synergistic combination of wavelet-domain diffusion and high-frequency refinement enables our model to produce restorations that are not only numerically accurate but also perceptually authentic, meeting the stringent requirements of cultural heritage preservation.

3.4 Quantitative and qualitative results on DhMurals dataset

Based on the comprehensive quantitative comparisons reported in Tables 5–8, we analyze the performance of the proposed WDCRF against several state-of-the-art restoration methods on the DhMurals dataset, including EdgeConnect, SPA-Former, WaveDiff, LRDiff, SDInpainting, CMAMRNet, PGRDiff, and RDDM. The evaluation is conducted under three degradation levels defined by the proportion of missing pixels in the masked input images: 0% – 20%, 20% – 40%, and $\geq 40\%$. We also report the overall average across all test samples for a comprehensive assessment.

PSNR Analysis.

As shown in Table 5, WDCRF achieves the highest overall PSNR of 18.95 dB, outperforming CMAMRNet (18.76 dB), SPA-Former (18.31 dB), PGRDiff (18.20 dB),



Fig. 3 Qualitative comparison of different mural restoration methods on the DhMurals dataset. Representative restoration results under different damage conditions are presented. From left to right, the columns show the damaged input image, existing restoration methods, the proposed WDCRF, and the ground truth (GT). GT denotes the original undamaged mural image used as the reference target.

Table 6 Comparison of SSIM under different degradation ratios on the DhMurals dataset. The best result in each column is highlighted in bold.

Method	0–20%	20–40%	$\geq 40\%$	Total
WDCRF	0.8203	0.7419	0.6860	0.7598
EdgeConnect	0.7370	0.6096	0.6143	0.6152
SPA-Former	0.8606	0.7410	0.6940	0.7411
WaveDiff	0.8488	0.7480	0.7332	0.7506
LRDiff	0.6844	0.4236	0.4762	0.4393
SDInpainting	0.5770	0.5096	0.5816	0.5195
CMAMRNet	0.8016	0.7074	0.6819	0.7474
PGRDiff	0.8133	0.6911	0.6481	0.6917
RDDM	0.6947	0.5474	0.5827	0.5568

Table 7 Comparison of LPIPS (lower is better) under different degradation ratios on the DhMurals dataset. The best result in each column is highlighted in bold.

Method	0–20%	20–40%	$\geq 40\%$	Total
WDCRF	0.1197	0.2026	0.2871	0.2078
EdgeConnect	0.1840	0.3801	0.4654	0.3808
SPA-Former	0.1418	0.2714	0.3919	0.2783
WaveDiff	0.1256	0.2352	0.2953	0.2369
LRDiff	0.1695	0.4646	0.5312	0.4595
SDInpainting	0.1695	0.2444	0.2693	0.2439
CMAMRNet	0.1437	0.2287	0.2872	0.2312
PGRDiff	0.1363	0.2776	0.3485	0.2791
RDDM	0.2213	0.4269	0.4626	0.4222

WaveDiff (18.19 dB), SDInpainting (17.49 dB), RDDM (13.95 dB), LRDiff (13.35 dB), and EdgeConnect (11.84 dB). In the mild degradation setting (0% – 20%), WDCRF obtains the best result of 20.77 dB, followed by PGRDiff (20.74 dB), SPA-Former (20.68 dB), CMAMRNet (20.63 dB), and WaveDiff (20.52 dB). Under the moderate degradation setting (20% – 40%), CMAMRNet achieves the highest PSNR of 19.08 dB, while WaveDiff (18.40 dB), WDCRF (18.38 dB), SPA-Former (18.36 dB), and PGRDiff (18.20 dB) remain competitive. In the severe degradation setting ($\geq 40\%$), CMAMRNet again obtains the best score of 17.31 dB, followed by WaveDiff (17.24 dB), PGRDiff (17.23 dB), SPA-Former (16.86 dB), and WDCRF (16.67 dB). Although CMAMRNet performs slightly better under larger missing regions, WDCRF achieves the best overall PSNR, demonstrating stronger average reconstruction accuracy across different degradation levels.

SSIM Analysis.

Table 6 reports the SSIM results on the DhMurals dataset. WDCRF achieves the best overall SSIM of 0.7598, surpassing WaveDiff (0.7506), CMAMRNet (0.7474), SPA-Former (0.7411), PGRDiff (0.6917), EdgeConnect (0.6152), RDDM (0.5568),

Table 8 FID scores on the DhMurals dataset. Lower is better.

Method	FID
WDCRF	85.360
EdgeConnect	260.232
SPA-Former	158.863
WaveDiff	149.538
LRDiff	92.817
SDInpainting	86.653
CMAMRNet	180.580
PGRDiff	141.494
RDDM	212.671

SDInpainting (0.5195), and LRDiff (0.4393). In the 0% – 20% degradation range, SPA-Former obtains the highest SSIM of 0.8606, followed by WaveDiff (0.8488), WDCRF (0.8203), PGRDiff (0.8133), and CMAMRNet (0.8016). For the 20% – 40% range, WaveDiff achieves the best score of 0.7480, while WDCRF obtains 0.7419, which is close to SPA-Former (0.7410) and CMAMRNet (0.7074). Under the most challenging condition ($\geq 40\%$), WaveDiff achieves the highest SSIM of 0.7332, followed by SPA-Former (0.6940), WDCRF (0.6860), and CMAMRNet (0.6819). Overall, WDCRF delivers the best total SSIM, indicating its advantage in maintaining global structural coherence and balanced restoration quality.

LPIPS Analysis.

Since LPIPS measures perceptual dissimilarity, lower values indicate better perceptual quality. As shown in Table 7, WDCRF achieves the lowest overall LPIPS of 0.2078, outperforming CMAMRNet (0.2312), WaveDiff (0.2369), SDInpainting (0.2439), SPA-Former (0.2783), PGRDiff (0.2791), EdgeConnect (0.3808), RDDM (0.4222), and LRDiff (0.4595). In the 0% – 20% range, WDCRF obtains the best LPIPS of 0.1197, followed by WaveDiff (0.1256), PGRDiff (0.1363), SPA-Former (0.1418), and CMAMRNet (0.1437). For the 20% – 40% range, WDCRF again achieves the best result of 0.2026, better than CMAMRNet (0.2287), WaveDiff (0.2352), and SDInpainting (0.2444). Under the $\geq 40\%$ condition, SDInpainting achieves the lowest LPIPS of 0.2693, while WDCRF obtains 0.2871, slightly better than CMAMRNet (0.2872) and clearly better than WaveDiff (0.2953), PGRDiff (0.3485), SPA-Former (0.3919), RDDM (0.4626), EdgeConnect (0.4654), and LRDiff (0.5312). These results indicate that WDCRF produces restorations that are perceptually closer to the ground truth, preserving fine textures and artistic brushstrokes that are critical for ancient mural authenticity.

FID Analysis.

As reported in Table 8, WDCRF achieves the lowest FID score of 85.360, demonstrating that its restored images have a feature distribution closest to that of real pristine murals. The second-best method is SDInpainting with an FID of 86.653, followed by LRDiff (92.817), PGRDiff (141.494), WaveDiff (149.538), SPA-Former (158.863),

CMAMRNet (180.580), RDDM (212.671), and EdgeConnect (260.232). The substantial advantage of WDCRF over most competing methods validates the generative fidelity of the proposed wavelet-domain conditional diffusion framework. In particular, the low FID indicates that WDCRF not only reconstructs missing content accurately but also maintains the overall statistical and stylistic consistency of ancient mural imagery.

Visual Quality Analysis.

Beyond quantitative metrics, we conduct a qualitative comparison of the restored images produced by different methods, as illustrated in Fig. 3. The figure presents representative examples from the DhMurals test set under varying degradation levels, including facial features, clothing folds, and background decorative patterns. WDCRF consistently generates visually coherent restorations that are semantically faithful and stylistically consistent with the original artwork. In regions with mild degradation, our method preserves the sharpness of facial contours and the subtle gradations of pigment colors, closely resembling the ground truth. In contrast, CMAMRNet and LRDiff tend to produce over-smoothed textures, losing the characteristic granularity of ancient mural surfaces, while WaveDiff and SDInpainting often introduce blurring artifacts around the repaired boundaries.

For more challenging cases with larger missing areas, the advantages of the frequency-decomposed design become more pronounced. The low-frequency diffusion process successfully reconstructs the global composition and color harmony, preventing the structural distortions that affect SPA-Former and RDDM. PGRDiff, although capable of producing plausible global structures, frequently introduces repetitive texture patterns that deviate from the authentic artistic style. Furthermore, WDCRF exhibits superior color fidelity. The conditional diffusion process in the wavelet domain effectively constrains the generative process to adhere to the original color palette, reducing the color shifts or oversaturation observed in CMAMRNet and WaveDiff. The auxiliary reconstruction loss further anchors the output to the ground-truth distribution, preventing stochastic sampling from producing visually inconsistent color tones.

In summary, the visual results corroborate the quantitative findings: WDCRF achieves a favorable balance between global structural integrity and local detail fidelity. The synergistic combination of wavelet-domain diffusion and high-frequency refinement enables our model to produce restorations that are not only numerically accurate but also perceptually authentic, meeting the stringent requirements of cultural heritage preservation.

3.5 Ablation study

In this section, we conduct a series of ablation experiments to systematically validate the effectiveness of each core component in our proposed WDCRF framework. Specifically, we aim to answer one key question: Does performing diffusion in the frequency domain offer advantages over standard spatial-domain diffusion? To this end, we design two model variants with progressively increasing complexity, enabling a clear attribution of performance gains to each design choice.



Fig. 4 Ablation comparison of different variants on the Dunhuang Mural dataset.

Table 9 Ablation study results on the Dunhuang Mural test set. The best results are highlighted in **bold**.

Variant	PSNR $\uparrow$	SSIM $\uparrow$	Time (s) $\downarrow$
Variant I: Spatial DDPM	29.20	0.898	35.31
Variant II: Full WDCRF	32.04	0.922	9.56

Variant I: Standard Diffusion Model.

We first establish a baseline by employing a vanilla denoising diffusion probabilistic model (DDPM) [33] directly in the pixel space, without any wavelet decomposition or frequency-based conditioning. This variant follows the standard diffusion framework commonly used for image inpainting tasks [15]. The model is trained with the same U-Net backbone and training settings as our full model, but operates on the original image resolution, resulting in higher computational cost. This baseline serves as the lower bound, allowing us to assess the performance improvement brought by our frequency-domain design.

Variant II: Full Model that performs restoration in frequency domain.

Instead of operating in pixel space, this variant performs restoration entirely in the wavelet domain. The damaged input is decomposed via 2D-DWT into one low-frequency and three high-frequency subbands. A conditional diffusion model restores all four subbands jointly, using the damaged coefficients as the conditioning signal. The restored coefficients are then reconstructed via inverse DWT. Sharing the same U-Net backbone and training settings as the full model, this variant operates on the spatially downsampled wavelet representation, reducing the spatial dimension of each denoising step while decoupling structural recovery from texture synthesis.

Quantitative Ablation Analysis.

Table 9 compares the spatial-domain baseline (Variant I) with the full WDCRF (Variant II). The full model improves PSNR from 29.20 dB to 32.04 dB and SSIM from 0.898 to 0.922, while reducing inference time from 35.31 s to 9.56 s. These results confirm that wavelet-domain compression benefits both structural reconstruction and computational efficiency, while the HFEM is necessary for texture fidelity.

Visual Comparison of Ablation Variants.

Fig. 4 provides a qualitative comparison. Variant I reconstructs the global layout but produces over-smoothed textures and blurred edges. In contrast, the full model (Variant II) generates coherent structures and sharp, authentic textures, with the HFEM restoring broken lines, crack patterns, and brushstroke details. The coarse-to-fine reconstruction also ensures seamless integration between low- and high-frequency components. Overall, the visual results corroborate the quantitative findings: wavelet-domain diffusion supplies a reliable structural foundation, while the HFEM is indispensable for perceptually faithful mural restoration.

4 Conclusion

In this paper, we proposed a Wavelet-Diffusion Collaborative Restoration Framework (WDCRF) for ancient mural restoration, addressing the dual challenges of computational efficiency and detail fidelity. The framework follows a frequency-aware “decompose and conquer” strategy that separates global semantic reconstruction from local detail synthesis. Using the Discrete Wavelet Transform, the damaged input is decomposed into low-frequency and high-frequency components. A conditional diffusion model then reconstructs all four wavelet subbands in the wavelet domain, using the damaged coefficients as the conditioning signal. The inverse wavelet transform integrates both pathways in a coarse-to-fine manner, ensuring global structural coherence and local texture fidelity.

Extensive experiments on the Dunhuang Mural dataset demonstrate that WDCRF achieves state-of-the-art performance, with PSNR of 32.04 dB, SSIM of 0.922, LPIPS of 0.036, and FID of 80.221. On the DhMurals dataset, our method also achieves the best overall PSNR of 18.95 dB and SSIM of 0.7598, along with the lowest LPIPS of 0.2078 and FID of 85.360. Qualitative results further confirm that our method produces visually authentic restorations that preserve the stylistic and artistic integrity of the original murals. Ablation studies validate that wavelet-domain diffusion is essential for optimal performance.

Acknowledgements. This work is supported by the Natural Science Foundation of Zhejiang Province, China under Grant Number LTGG24F020005 and Philosophical Social Science Foundation of Zhejiang Province under Grant Number 26JCXK010YB.

References

- [1] Wang, Y., Wu, X.: Current progress on murals: distribution, conservation and utilization. *Heritage Science* **11**(1), 61 (2023)
- [2] Li, X., Lu, D., Pan, Y.: Virtual dunhuang mural restoration system in collaborative network environment. In: *Computer Graphics Forum*, vol. 19, pp. 331–340 (2000). Wiley Online Library

- [3] Mol, V.R., Maheswari, P.U.: The digital reconstruction of degraded ancient temple murals using dynamic mask generation and an extended exemplar-based region-filling algorithm. *Heritage Science* **9**(1), 1–18 (2021)
- [4] Purkait, P., Chanda, B.: Digital restoration of damaged mural images. In: *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing*, pp. 1–8 (2012)
- [5] Zhang, S., He, F., Ren, W., Yao, J.: Joint learning of image detail and transmission map for single image dehazing. *Visual Computer* **36**(2), 305 (2020)
- [6] Zhang, S., He, F., Ren, W.: Nldn: Non-local dehazing network for dense haze removal. *Neurocomputing* **410**, 363–373 (2020)
- [7] Zhang, S., Zhang, X., Wan, S., Jiang, Y.M., Zhou, W., Shen, L., Ren, W.: Hierarchical multi-modal enhancement for robust transmission line detection. *IEEE Transactions on Automation Science and Engineering* (2026)
- [8] Zhang, S., Zhang, X., Ren, W., Shen, L., Zhang, J.: Shape-aware and feature fused power line detection network. *Engineering Applications of Artificial Intelligence* **170**, 113981 (2026)
- [9] Xiang, H., Zou, Q., Nawaz, M.A., Huang, X., Zhang, F., Yu, H.: Deep learning for image inpainting: A survey. *Pattern Recognition* **134**, 109046 (2023)
- [10] Nazeri, K., Ng, E., Joseph, T., Qureshi, F., Ebrahimi, M.: EdgeConnect: Structure Guided Image Inpainting using Edge Prediction. In: *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 3265–3274. IEEE, Seoul, Korea (South) (2019). <https://doi.org/10.1109/ICCVW.2019.00408>. TLDR: This work proposes a two-stage model that separates the inpainting problem into structure prediction and image completion, similar to sketch art, and demonstrates that this approach outperforms current state-of-the-art techniques quantitatively and qualitatively. <https://ieeexplore.ieee.org/document/9022543/> Accessed 2025-10-06
- [11] Li, L., Zou, Q., Zhang, F., Yu, H., Chen, L., Song, C., Huang, X., Wang, X., Li, Q.: Line Drawing-Guided Progressive Inpainting for Mural Damage. *Journal on Computing and Cultural Heritage* **18**(3), 1–20 (2025) <https://doi.org/10.1145/3737703>. TLDR: Qualitative and quantitative results demonstrate the superiority of the proposed line drawing guided progressive mural inpainting method against the current state-of-the-art image inpainting methods. Accessed 2025-10-04
- [12] Liu, Z., Liu, Y., Hu, W.: Mural restoration via the fusion of edge-guided and multi-scale spatial features. *npj Heritage Science* **13**(1), 491 (2025) <https://doi.org/10.1038/s40494-025-02073-3>. Accessed 2025-10-08
- [13] Ren, H., Sun, K., Zhao, F., Zhu, X.: Dunhuang murals image restoration method

- based on generative adversarial network. *Heritage Science* **12**(1), 39 (2024) <https://doi.org/10.1186/s40494-024-01159-8> . Accessed 2025-10-06
- [14] Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. *Advances in neural information processing systems* **33**, 6840–6851 (2020)
 - [15] Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., Van Gool, L.: Repaint: Inpainting using denoising diffusion probabilistic models. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11461–11471 (2022)
 - [16] Jiang, H., Luo, A., Fan, H., Han, S., Liu, S.: Low-light image enhancement with wavelet-based diffusion models. *ACM Transactions on Graphics (ToG)* **42**(6), 1–14 (2023)
 - [17] Lei, T., Li, Y., Cheng, Z., Zhou, W.: Low-rank structure guided diffusion for shaanxi temple mural restoration. *npj Heritage Science* **13**(1), 295 (2025)
 - [18] Jiang, C., Ren, T., Cheng, Z.: All-in-one mural restoration with prompt-guided residual diffusion. *npj Heritage Science* **13**(1), 667 (2025)
 - [19] Zhang, L., Wu, Y., Kong, L., Zhao, J., Jia, F.: Mural inpainting for structural distortion and color deviation: edge-constrained diffusion and adversarial optimization strategy. *npj Heritage Science* **14**(1), 363 (2026)
 - [20] Ding, Y., Wu, K., Tian, B.: Frequency-domain information guidance: Diffusion models for the Inpainting of Dunhuang murals. *Knowledge-Based Systems* **314**, 113188 (2025) <https://doi.org/10.1016/j.knosys.2025.113188> . Accessed 2025-10-06
 - [21] Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence* **11**(7), 674–693 (1989)
 - [22] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695 (2022)
 - [23] Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: *International Conference on Learning Representations* (2021). <https://openreview.net/forum?id=StlgiaRCHLP>
 - [24] Han, P., Pan, Y., Pan, E., Jin, Q., Jiang, J., Zhang, Z., Liu, Z., Kang, J., Gong, L.: Diffumural: a diffusion model for dunhuang murals restoration via multi-scale convergence and cooperative guidance. *npj Heritage Science* **14**(1), 44 (2026)

- [25] Yan, Y., Lv, Z., Jiang, C., Cheng, Z.: Learning mural restoration from degraded data via unsupervised low-rank residual diffusion (2026)
- [26] Nazeri, K., Ng, E., Joseph, T., Qureshi, F., Ebrahimi, M.: Edgeconnect: Structure guided image inpainting using edge prediction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops (2019)
- [27] Manukyan, H., Sargsyan, A., Atanyan, B., Wang, Z., Navasardyan, S., Shi, H.: Hd-painter: high-resolution and prompt-faithful text-guided image inpainting with diffusion models. In: International Conference on Learning Representations, vol. 2025, pp. 96301–96330 (2025)
- [28] Lei, Y., Yi, F., Dong, Y., Liu, W., Zhang, X., Li, Z., Pun, C.-M., Chen, X.: Cmam-rnet: A contextual mask-aware network enhancing mural restoration through comprehensive mask guidance. arXiv preprint arXiv:2508.07140 (2025)
- [29] Liu, J., Wang, Q., Fan, H., Wang, Y., Tang, Y., Qu, L.: Residual denoising diffusion models. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2773–2783 (2024). IEEE
- [30] Huang, W., Deng, Y., Hui, S., Wu, Y., Zhou, S., Wang, J.: Sparse self-attention transformer for image inpainting. Pattern Recognition **145**, 109897 (2024)
- [31] Yu, T., Zhang, S., Lin, C., You, S., Wu, J., Zhang, J., Ding, X., An, H.: Dunhuang grottoes painting dataset and benchmark. arXiv preprint arXiv:1907.04589 (2019)
- [32] QIAN, J.: DUNGHUANG MURAL DATASETS. <https://doi.org/10.21227/sad5-pa49> . <https://dx.doi.org/10.21227/sad5-pa49>
- [33] Huang, Z., Wei, Y., Chen, Y.: Denoising diffusion probabilistic models are optimally adaptive to unknown low dimensionality. Mathematics of Operations Research (2026)