

Severity-Conditioned Pixel-Wise Asynchronous Diffusion for Semi-Transparent Mural Restoration

Yu Wu

wuyu_0809@163.com

Lanzhou University of Technology

Erchao Li

Lanzhou University of Technology

Furui Jia

Lanzhou University of Arts and Science

Jinjiang Zhao

Lanzhou University of Technology

Lei Zhang

Gansu Academy of Sciences

Article

Keywords: mural restoration, diffusion model, degradation severity, pixel-wise asynchronous diffusion, residual diffusion

Posted Date: October 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11195229/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Severity-Conditioned Pixel-Wise Asynchronous Diffusion for Semi-Transparent Mural Restoration

Yu Wu^{1,2,*}, Erchao Li^{1,*}, Furui Jia³, Jinjiang Zhao^{1,2} and Lei Zhang^{2,*}

¹ School of Automation and Electrical Engineering, Lanzhou University of Technology, Lanzhou 730050, China;

² Institute of Automation, Gansu Academy of Sciences, Lanzhou 730000, China;

³ Lanzhou University of Arts and Science, Lanzhou 730000, China;

* Correspondence: wuyu_0809@163.com

Abstract

Murals are important carriers of cultural-heritage documentation and research, yet long-term environmental exposure and material aging can cause fading, discoloration, surface contamination, mold, cracking, and local detachment. For semi-transparent degradations that still preserve part of the original pigment, texture, and contour information, a binary damage mask specifies only the degraded support and cannot represent continuously varying degradation severity within the same affected region. Moreover, standard diffusion models assign a shared temporal progress to the entire image, making spatially differentiated restoration dynamics difficult to represent explicitly. We therefore propose Severity-Conditioned Pixel-Wise Asynchronous Diffusion (SC-PAD). SC-PAD first estimates a continuous terminal severity field from the degraded observation and a coarse binary support using bounded refinement and risk-oriented constraints to suppress large local errors and isolated high responses. It then predicts a bounded temporal-warping field conditioned on terminal severity and image evidence, reparameterizing a shared global progress into pixel-wise local restoration progress. A dynamic Alpha state satisfying endpoint consistency, temporal monotonicity, and the Alpha composition law subsequently controls the activation of the image-space residual. Experiments show that continuous severity representation and pixel-wise asynchronous progress are the two components most clearly supported by the current results, while the remaining gap between predicted and ground-truth severity indicates that severity estimation is still an important performance bottleneck. The proposed formulation provides a controllable diffusion framework for spatially non-uniform semi-transparent mural restoration.

Keywords: mural restoration; diffusion model; degradation severity; pixel-wise asynchronous diffusion; residual diffusion

1. Introduction

Murals preserve evidence of religious beliefs, social life, artistic styles, and material techniques from different historical periods, making them important objects of cultural-heritage conservation and research. Long-term exposure to fluctuations in temperature and humidity, salt migration, illumination, particulate deposition, biological activity, and human disturbance can produce fading, discoloration, contamination, mold, cracking, flaking, detachment, and local loss. Recent material analyses, microclimate studies, and digital-conservation investigations indicate that mural deterioration is jointly governed by material properties, environmental conditions, and long-term aging and is often spatially heterogeneous [1–3]. Physical conservation must account for minimal intervention, material compatibility, and reversibility, whereas digital image restoration does not directly act on the heritage object and can support restoration planning, digital presentation, and archival documentation [4–6].

Deep learning has substantially advanced digital mural restoration. Generative adversarial networks, reversible networks, and multi-stage generative models exploit contextual, structural, and textural information to reconstruct damaged regions and have been applied to mural datasets including Dunhuang murals [7–11]. Diffusion models have subsequently been introduced because of their progressive generation and flexible conditional modeling. GuidePaint uses image guidance to improve information preservation [12]; a diffusion-based heterogeneous network combines spatial enhancement and channel attention for large-scale damage

and detail recovery [13]; an improved DDPM integrates attention and multi-scale context for noise estimation [14]; and LRDiff incorporates non-local similarity and global low-rank priors into mural restoration [15]. Line-drawing guidance, structure-texture cooperation, and multimodal controlled diffusion further demonstrate the value of mural-specific structural, stylistic, and semantic conditions [16–18].

Despite these advances, many mural-restoration methods remain formulated primarily as missing-region completion. Automated defect segmentation and prompt-guided inpainting methods first localize damaged areas and then generate content inside the mask [19]. PGRDiff instead considers discoloration, fading, water stains, and surface contamination as semi-transparent degradations in which original color, texture, and contour information may remain visible, and models the degradation using image-space residual diffusion [20]. This setting exposes a limitation of binary masks: they identify the spatial support of degradation but do not encode continuously varying severity within that support. Recent DiffuMural, DMEA-Net, and unsupervised low-rank residual diffusion methods extend mural diffusion restoration through multi-scale cooperation, contextual attention, and learning from degraded data [21–23], but their primary modeling emphasis remains on structure, texture, semantic conditions, or residual representation.

Related progress in general image restoration provides additional diffusion formulations. StrDiffusion introduces structure guidance to reduce the semantic discrepancy between masked and unmasked regions during image inpainting [24], whereas RDDM decomposes denoising diffusion into directional residual diffusion and stochastic noise diffusion so that the restoration trajectory explicitly represents the residual transition between degraded and target images [25]. DTPM learns diffusion-based texture priors [26], iteratively preconditioned guidance improves observation consistency for inverse problems [27], and DeqIR reorganizes long-chain diffusion restoration through a deep-equilibrium formulation [28]. DiffBIR, SinSR, PASD, and StableSR further show that degradation conditions, generative priors, pixel-level structural constraints, and sampling paths strongly affect restoration fidelity and perceptual quality [29–32]. Fading, soot deposition, contamination, mold, and partial discoloration in murals are not equivalent to complete occlusion or pixel loss. Such degradations often modify only part of the original appearance, leaving varying amounts of texture, contour, and color information visible. Their intensity can also change continuously across space as a consequence of material, environmental, and deterioration processes [33–35]. A continuous severity field can therefore describe spatially non-uniform degradation more faithfully than a binary support. However, spatial modeling of degradation severity is not equivalent to temporal modeling of the restoration process.

In standard diffusion formulations, the timestep t or normalized reference progress m_t is shared across the image. A network can still produce spatially varying predictions from image content and conditioning information, but all pixels are defined at the same diffusion-time state. For semi-transparent degradation with spatially varying terminal severity, different pixels have different admissible state ranges; a shared temporal progress leaves these differences to be handled implicitly by network features. Even when continuous severity is supplied as a condition, it does not by itself change the temporal progression of the intermediate diffusion state. This observation motivates a pixel-dependent restoration progress within a unified global diffusion schedule. Such a local progress should represent spatial heterogeneity while preserving common endpoints and temporal ordering, so that temporal reparameterization does not alter the basic boundary conditions of the diffusion process. The local progress should also map to an intermediate state constrained by terminal severity, thereby linking continuous severity, local restoration dynamics, and image-space residual evolution in a single formulation.

Based on this formulation, we propose Severity-Conditioned Pixel-Wise Asynchronous Diffusion (SC-PAD) for digital restoration of spatially non-uniform semi-transparent mural degradation. SC-PAD first estimates a continuous terminal severity field from the degraded image and a coarse binary support. It then predicts a bounded pixel-wise temporal-warping exponent conditioned on the degraded image and terminal severity, converting the shared global reference progress into pixel-wise local restoration progress while preserving common endpoints and temporal order. Finally, a dynamic Alpha state is constructed from local progress and terminal severity and is used to control the activation of the image-space residual. The resulting

model explicitly separates degradation support, terminal severity, local progress, and the intermediate diffusion state.

The main contributions are summarized as follows.

1. We formulate continuous terminal severity as an explicit state variable for semi-transparent mural degradation. A mask-guided estimator with bounded residual-logit refinement and risk-oriented supervision reduces local severity mismatch, allowing severity to serve as a terminal-state constraint rather than merely an auxiliary conditioning channel.

2. We introduce severity-conditioned pixel-wise temporal reparameterization. A positive bounded temporal-warping exponent maps the shared global reference progress to pixel-wise local restoration progress while preserving common endpoints and temporal monotonicity, enabling spatially differentiated restoration trajectories within one diffusion schedule.

3. We construct a terminal-severity-constrained dynamic Alpha state and couple it with residual diffusion. The state satisfies endpoint consistency, monotonicity, and the Alpha composition law, and controls progressive activation of the image-space residual, establishing an explicit link among terminal severity, local time, and intermediate diffusion states.

2. Method

2.1 Overview

We consider semi-transparent mural restoration with spatially non-uniform degradation severity. The main notation is summarized in Table 1. Let $X_0 \in \mathbb{R}^{H \times W \times 3}$ denote the clean mural, $X_{deg} \in \mathbb{R}^{H \times W \times 3}$ the degradation appearance field, $Y \in \mathbb{R}^{H \times W \times 3}$ the degraded observation, and $B \in \{0,1\}^{H \times W}$ a coarse binary support mask. A continuous severity field $A \in [0,1]^{H \times W}$ specifies the contribution of the degradation appearance at each spatial location. The controlled semi-transparent degradation state is

$$Y = (1 - A) \odot X_0 + A \odot X_{deg} \quad (1)$$

where $\odot$ denotes element-wise multiplication. Eq.(1) is used as a controlled image-state representation for the semi-transparent degradation considered in this work; A is the terminal severity rather than a complete physical model of material deterioration or chemical reactions. We directly use the terminal image-space residual R^M in the subsequent diffusion process. Unlike a normalized degradation-layer representation that divides by A , R^M remains well defined as A approaches zero.

$$Y = X_0 + A \odot (X_{deg} - X_0) = X_0 + R^M, R^M = Y - X_0 \quad (2)$$

During supervised oracle evaluation, A is set to the ground-truth terminal severity A^{gt} ; during deployable inference, A is set to the estimated severity $\hat{A}$:

$$A = \begin{cases} A^{gt}, & \text{oracle evaluation,} \\ \hat{A}, & \text{deployable inference.} \end{cases} \quad (3)$$

SC-PAD comprises continuous severity estimation, pixel-wise local-time modeling, and dynamic-state-driven residual diffusion. Given Y and B , the severity estimator predicts $\hat{A}$. Conditioned on Y and A , a trajectory predictor estimates the temporal-warping field γ and transforms the shared global reference progress m_t into pixel-wise local progress

$$g_t(u) = m_t^\gamma(u) \quad (4)$$

where u indexes spatial location. This reparameterization retains a common global diffusion schedule while allowing different pixels to occupy different intermediate restoration progresses. The local progress is then mapped to a terminal-severity-constrained dynamic Alpha state α^t , from which the residual activation ratio ρ^t is obtained. The noisy intermediate state is

$$X_t = X_0 + \rho_t \odot R^M + \sigma_t \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I) \quad (5)$$

where σ_t denotes the noise schedule. Thus, terminal severity is not only used as a network condition but also participates directly in constructing the local progress and the diffusion state. Figure 1 illustrates the detailed workflow of the proposed SC-PAD algorithm.

Table 1 | Main notation used in SC-PAD.

Symbol	Definition
X_0, Y	Clean mural and degraded observation
B	Coarse binary support mask
X_{deg}	Degradation appearance field
$A^{\text{gt}} / \hat{A} / A$	Ground-truth / estimated / actually used terminal severity
m_t	Normalized global reference progress
$\gamma(u)$	Pixel-wise temporal-warping exponent
$g_t(u)$	Pixel-wise local restoration progress
$\alpha_t(u)$	Dynamic Alpha state
$\rho_t(u)$	Residual activation ratio
R^M	Terminal image-space residual, $Y - X_0$
X_t	Noisy intermediate image
σ_t	Noise scale at timestep t
ε	Standard Gaussian noise

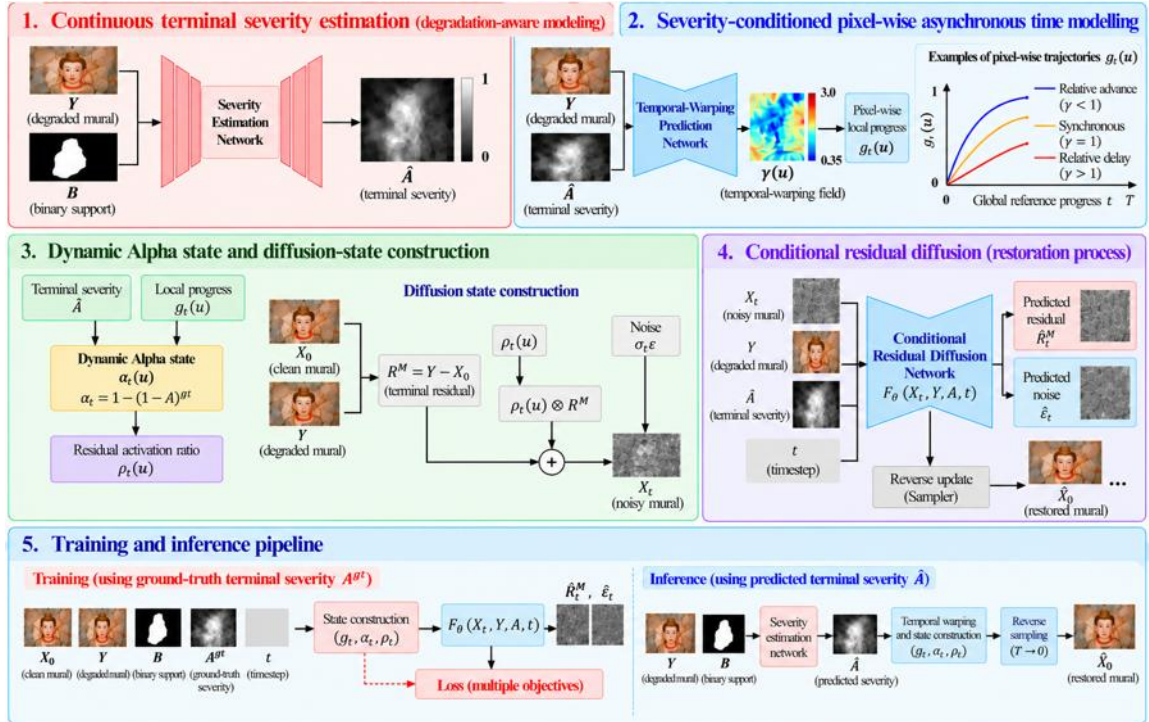


Figure 1 | Overall framework of the proposed SC-PAD.

2.2 Continuous terminal severity estimation

Mask-guided coarse estimation and bounded refinement. The binary mask provides an approximate degradation support but cannot determine continuous severity within the region. A mask-guided estimator

M_ψ first predicts a coarse terminal severity field. A residual-logit refinement branch then fuses the degraded RGB image, binary support, coarse severity, and local first- and second-order image statistics:

$$A_0 = M_\psi(Y, B), \quad F_{\text{ref}} = \text{concat}(Y, B, A_0, \mu_r(Y), \sigma_r(Y)) \quad (6)$$

$$\hat{A} = \text{sigmoid}(\text{logit}(A_0) + f_\phi(F_{\text{ref}})) \quad (7)$$

The residual-logit parameterization has two useful properties. When the refinement branch outputs zero, the coarse estimate is recovered exactly; after the sigmoid mapping, the refined severity is constrained to $[0,1]$. The refinement branch therefore performs a bounded correction rather than independently predicting a second severity field. The local statistics $\mu_r(Y)$ and $\sigma_r(Y)$ provide neighborhood intensity context that helps distinguish isolated high-severity responses from spatially supported degradation.

Risk-oriented supervision. Let $e(u) = \hat{A}(u) - A^{\text{gt}}(u)$ denote the terminal-severity estimation error. With a conventional average regression loss, numerous easy pixels can dominate the objective and obscure a small number of large local errors. We therefore introduce three complementary terms: a top- q tail loss for the largest absolute errors, a one-sided penalty for positive severity errors, and a neighborhood spike penalty for isolated high-severity predictions:

$$L_{\text{tail}} = \frac{1}{|T_q|} \sum_{u \in T_q} |e(u)|, \quad L_{\text{over}} = \frac{1}{|\Omega|} \sum_{u \in \Omega} w(u)[e(u)]_+ \quad (8)$$

$$L_{\text{spike}} = \frac{1}{|\Omega|} \sum_{u \in \Omega} \left[\hat{A}(u) - \frac{1}{|N_r(u)|} \sum_{v \in N_r(u)} \hat{A}(v) - \tau_s \right]_+, \quad e(u) = \hat{A}(u) - A^{\text{gt}}(u) \quad (9)$$

Here, T_q contains the q proportion of pixels with the largest absolute severity errors, and $w(u)$ can emphasize visually or structurally sensitive regions. The spike term compares the current prediction with a local neighborhood reference and is activated only when the local excess exceeds tolerance τ_s , avoiding unnecessary penalties on genuinely extensive severe degradation.

$$L_A = L_{\text{base}} + \lambda_{\text{tail}} L_{\text{tail}} + \lambda_{\text{over}} L_{\text{over}} + \lambda_{\text{spike}} L_{\text{spike}} \quad (10)$$

The base objective combines pixel-wise regression, gradient consistency, multi-scale consistency, and correlation constraints:

$$L_{\text{base}} = \lambda_{\text{pix}} L_{\text{pix}} + \lambda_{\text{grad}} L_{\text{grad}} + \lambda_{\text{ms}} L_{\text{ms}} + \lambda_{\text{corr}} L_{\text{corr}} \quad (11)$$

2.3 Severity-conditioned pixel-wise asynchronous trajectory

Local-time reparameterization. Let $m_t \in [0,1]$ be a monotonic normalized global reference progress with $m_0=0$ and $m_1=1$. A synchronous process uses the same m_t for every pixel. The trajectory predictor H_η receives the degraded image Y and terminal severity A , outputs a spatial field z_η , and maps it through a bounded sigmoid to a positive temporal-warping exponent γ :

$$z_\eta(u) = H_\eta(Y, A)(u), \quad \gamma(u) = \gamma_{\min} + (\gamma_{\max} - \gamma_{\min}) \text{sigmoid}(z_\eta(u)), \quad g_t(u) = m_t^{\gamma(u)} \quad (12)$$

The parameterization explicitly constrains γ to a predefined interval and initializes the final-layer bias so that the initial prediction is close to the synchronous solution $\gamma=1$. When $\gamma(u)=1$, $g_t(u)=m_t$. For $0 < m_t < 1$, $\gamma(u) < 1$ gives $g_t(u) > m_t$, corresponding to relative advancement, whereas $\gamma(u) > 1$ delays the local progress. The direction of the temporal offset is therefore learned jointly from Y and A rather than imposed as a fixed monotonic function of severity.

Boundedness and temporal order. In our implementation $\gamma_{\min}=0.35$ and $\gamma_{\max}=3.0$. For every pixel and every interior reference progress,

$$\gamma_{\min} \leq \gamma(u) \leq \gamma_{\max}, \quad g_0(u) = 0, \quad g_T(u) = 1, \quad \frac{\partial g_t(u)}{\partial m_t} = \gamma(u) m_t^{\gamma(u)-1} > 0 \quad (13)$$

These relations ensure bounded temporal warping, common start and end points for all pixels, and a strictly increasing local clock with respect to the global reference progress. The asynchronous process therefore remains within a unified global temporal range.

Trajectory regularization. To suppress unnecessary temporal offsets, γ is set to 1 where terminal severity is close to zero. Within the degradation support, a synchronization prior on $|\log \gamma|$ discourages excessive deviations from $\gamma=1$. A terminal-severity-weighted total-variation term further suppresses isolated spatial oscillations:

$$S(u) = \mathbf{1}[A(u) > \varepsilon], \quad L_{\text{warp}} = \frac{\sum_u S(u) |\log \gamma(u)|}{\sum_u S(u) + \varepsilon} \quad (14)$$

$$L_{\text{time}} = \lambda_p L_{\text{warp}} + \lambda_{\text{tv}} \left[\text{mean}(w_x | \Delta_x \gamma) + \text{mean}(w_y | \Delta_y \gamma) \right] \quad (15)$$

The weights w_x and w_y are the maximum terminal severities of the corresponding neighboring pixels. This regularizer constrains the magnitude and spatial continuity of the local-time field without imposing a predetermined relationship between severity and temporal advancement or delay.

2.4 Dynamic Alpha state and residual diffusion

The pixel-wise local progress g_t determines the relative restoration progress at each spatial location, but does not directly specify the corresponding intermediate degradation state. We therefore map g_t to a terminal-severity-constrained dynamic Alpha state. For numerical stability, A is clipped to $[0, 1-\varepsilon_A]$, and the dynamic Alpha state is defined as

$$\alpha_t(u) = 1 - [1 - A(u)]^{g_t(u)}, \quad \alpha_0(u) = 0, \quad \alpha_T(u) = A(u) \quad (16)$$

This mapping satisfies $\alpha_0=0$, $\alpha_T=A$, and $0 \leq \alpha_t \leq A$ and is monotonic in g_t . The cumulative-opacity coordinate is motivated by the multiplicative structure of Alpha composition in the transmittance domain. For opacities a and b ,

$$a \oplus b = 1 - (1-a)(1-b), \quad \alpha(g_1 + g_2; A) = \alpha(g_1; A) \oplus \alpha(g_2; A) \quad (17)$$

Let $T(g) = 1 - \alpha(g; A)$. The semigroup relation in Eq. (17) becomes $T(g_1 + g_2) = T(g_1)T(g_2)$, where $g_1 + g_2 \leq 1$. Together with $T(0)=1$ and $T(1)=1-A$, this gives $T(g) = (1-A)^g$, yielding the dynamic Alpha in Eq. (16). Thus, the dynamic Alpha follows the Alpha composition law rather than an arbitrary nonlinear interpolation. To characterize how terminal severity and local time jointly affect state evolution, differentiating Eq. (16) with respect to the global reference progress m_t gives

$$\frac{\partial \alpha_t(u)}{\partial m_t} = -\log[1 - A(u)] [1 - \alpha_t(u)] \gamma(u) m_t^{\gamma(u)-1} \quad (18)$$

The state velocity is jointly determined by terminal severity, the remaining transmittance, and the learned temporal warping. Hence, A constrains the severity-dependent state trajectory, while $\gamma(u)$ controls its relative advancement or delay.

For weak degradation, the dynamic Alpha admits the expansion

$$\alpha(g; A) = gA + \frac{g(1-g)}{2} A^2 + O(A^3), \quad A \rightarrow 0 \quad (19)$$

Therefore, the linear rule $\alpha=gA$ is the first-order approximation of the composition-consistent trajectory at low severity, while the nonlinear correction becomes more pronounced as A increases. This also provides the theoretical basis for comparing linear and composition-consistent Alpha parameterizations in the trajectory sensitivity analysis.

Image-space residual activation. To couple the dynamic Alpha state with image-space residual diffusion, the terminal residual $R^M = Y - X_0$ defined in Eq. (2) is progressively activated according to the current Alpha state. To avoid numerical instability when A approaches zero, the residual activation ratio is defined as

$$\rho_t(u) = \begin{cases} \frac{\alpha_t(u)}{A(u)}, & A(u) > \varepsilon_\rho, \\ [6pt] 0, & A(u) \leq \varepsilon_\rho \end{cases} \quad (20)$$

For nonzero degradation regions, Eq. (20) gives $\rho_0=0$ and $\rho_T=1$, whereas the residual drift is suppressed in healthy or near-healthy regions. The noisy intermediate state is then constructed using the image-space residual formulation introduced in Eq.(5). Under the controlled observation model and exact terminal severity, its noise-free component satisfies $X_t - \sigma_t \varepsilon = (1 - \alpha_t)X_0 + \alpha_t X_{deg}$. Thus, the deterministic intermediate states remain on the same semi-transparent trajectory defined by the terminal severity.

Conditional residual and noise prediction. A conditional U-Net F_θ jointly predicts the terminal image-space residual and Gaussian perturbation from X_t , Y , A , and timestep t . The prediction and corresponding clean-image estimate are written as

$$(\hat{R}_t^M, \hat{\varepsilon}_t) = F_\theta(X_t, Y, A, t), \quad \hat{X}_{0,t} = X_t - \rho_t \odot \hat{R}_t^M - \sigma_t \hat{\varepsilon}_t \quad (21)$$

Here, g_t , α_t and ρ_t are used to construct the diffusion state and reverse update rather than being concatenated as additional U-Net channels. Training jointly constrains residual and noise prediction, damaged-region reconstruction, re-degradation consistency, healthy-region preservation, color and low-frequency consistency, together with the trajectory regularization introduced in Severity-conditioned pixel-wise asynchronous trajectory.

Algorithm 1 | Training procedure of severity-conditioned asynchronous residual diffusion.

Input	Training sample (X_0, Y, A^{gt}) , reference schedules $\{m, \sigma\}$, trajectory predictor H_η , conditional U-Net F_θ , and numerical constants ε_A and ε_ρ .
1	During training set $A \leftarrow A^{gt}$; during deployment obtain $\hat{A}$ from the severity estimator and set $A \leftarrow \hat{A}$. Clip A to $[0, 1 - \varepsilon_A]$.
2	Predict $z_\eta \leftarrow H_\eta(Y, A)$, compute $\gamma \leftarrow \gamma_{min} + (\gamma_{max} - \gamma_{min}) \text{sigmoid}(z_\eta)$, and set $\gamma=1$ where $A \approx 0$.
3	Sample timestep t and Gaussian noise ε , and compute local progress $g_t \leftarrow m^\gamma$.
4	Compute dynamic Alpha $\alpha_t \leftarrow 1 - (1 - A)^{g_t}$; obtain residual activation ρ_t from the safe piecewise ratio α_t / A defined in Eq. (20).
5	Construct terminal image-space residual $R^M \leftarrow Y - X_0$ and synthesize $X_t \leftarrow X_0 + \rho_t \odot R^M + \sigma_t \varepsilon$.
6	Feed X_t , Y , A and timestep t to the conditional U-Net to predict $\hat{R}_t^M$ and $\hat{\varepsilon}_t$.
7	Reconstruct $\hat{X}_0$ from $\hat{R}_t^M$, $\hat{\varepsilon}_t$, and ρ_t , and compute residual, noise, damaged-region reconstruction, and re-degradation consistency losses.
8	Add healthy-region preservation, color and low-frequency consistency, trajectory-prior, and total-variation regularization; update H_η and F_θ .
9	Repeat Steps 2–8 until convergence.
Output	Trained trajectory predictor H_η and conditional residual/noise U-Net F_θ ; the severity estimator provides $\hat{A}$ during deployment.

3. Experiments

3.1 Experimental setup

3.1.1 Dataset

Experiments are conducted on the DUNHUANG mural dataset using a fixed split of 400 training images and 100 test images. To synthesize degradation with varying transparency, real degradation patches and their corresponding transparency maps are blended with intact mural images. The degradation set covers ten representative categories: salt crystallization, discoloration, cracking, fading, detachment, insect damage, mold, scratches, soot deposition, and water stains. Each category contains multiple spatial patterns and transparency levels. The fixed split is used for the main training and deployment evaluation to avoid changes in the test set confounding metric comparisons.

3.1.2 Comparison methods

We compare SC-PAD with representative diffusion-based restoration methods spanning mean-reverting stochastic differential equations, structure-guided diffusion, residual diffusion, and mural-specific diffusion

restoration. IR-SDE [36] models restoration using a mean-reverting stochastic differential equation and serves as a reference for continuous stochastic restoration dynamics. StrDiffusion [24] introduces time-dependent structure guidance to reduce the semantic discrepancy between masked and unmasked regions in image inpainting. RDDM [25] decomposes conventional denoising diffusion into directional residual diffusion and stochastic noise diffusion. LRDiff [15] is designed for mural restoration and incorporates global low-rank priors and non-local similarity through Tucker-based low-rank modeling. PGRDiff [20] models semi-transparent mural degradation with prompt-guided image-space residual diffusion and is the closest baseline to our degradation setting. Table 2 reports only methods for which quantitative records are available in the current experimental set.

3.1.3 Implementation details

All experiments are implemented in PyTorch with an input resolution of 256x256. The restoration network adopts a four-scale U-Net with 64 base channels and channel multipliers of (1,2,4,8). The diffusion process consists of 100 timesteps, with 10 sampling steps used during inference. The model is trained for 10,000 optimization steps with a batch size of 1 and an initial learning rate of 1×10^{-4} . The temporal-warping exponent is constrained to [0.35,3.0]. Unless otherwise specified for oracle evaluation, the predicted soft Alpha is used as the terminal severity input during deployment.

3.1.4 Evaluation metrics

We use peak signal-to-noise ratio (PSNR) [37], structural similarity index measure (SSIM) [38], and learned perceptual image patch similarity (LPIPS) [39]. Higher PSNR and SSIM indicate better pixel fidelity and structural consistency, respectively, whereas lower LPIPS indicates smaller perceptual feature distance. FID is not included because it is not reported consistently across the current main result tables.

3.2 Overall comparison with existing methods

The first experiment evaluates the complete deployable configuration of SC-PAD on the fixed 100-image test set. During inference, SC-PAD uses the estimated terminal severity $\hat{A}$ rather than the ground-truth severity. Quantitative results are summarized in Table 2.

Table 2 | Quantitative comparison on the DUNHUANG dataset.

Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
LRDiff [15]	27.6	0.884	0.112
StrDiffusion [24]	28.3	0.899	0.098
RDDM [25]	28.8	0.908	0.087
PGRDiff [20]	29.7	0.918	0.063
SC-PAD (Ours)	30.6122	0.9253	0.0712

Quantitative comparison: As shown in Table 2, SC-PAD achieves a PSNR of 30.6122 dB and an SSIM of 0.9253, yielding the highest values among the compared methods. Compared with LRDiff, StrDiffusion, and RDDM, SC-PAD improves PSNR by 3.0122, 2.3122, and 1.8122 dB, respectively, and SSIM by 0.0413, 0.0263, and 0.0173. Compared with PGRDiff, SC-PAD achieves gains of 0.9122 dB in PSNR and 0.0073 in SSIM, indicating improved pixel-level reconstruction fidelity and structural preservation.

These results are consistent with the design of SC-PAD. The method represents spatially non-uniform degradation using the continuous terminal severity A , while the local progress g_t , dynamic Alpha state a_t , and residual activation ρ_t regulate state evolution across spatial locations and diffusion stages. This formulation enables different degraded regions to follow spatially varying intermediate restoration states. The overall comparison suggests that such modeling contributes to improved restoration accuracy, while the individual effects of these components are further examined in subsequent experiments.

For LPIPS, SC-PAD achieves 0.0712, outperforming LRDiff (0.112), StrDiffusion (0.098), and RDDM (0.087), but remaining higher than PGRDiff (0.063). Thus, the advantage of SC-PAD is mainly reflected in

pixel fidelity and structural consistency rather than uniform superiority across all perceptual metrics. Since the experimental protocols of the baseline methods are not fully unified, Table 2 is used primarily for overall performance comparison rather than for claims of statistical significance.

Qualitative comparison: Figures 2 and 3 present representative restoration results for contamination, fading, cracks, and block-like semi-transparent degradation. The qualitative comparison focuses on structural preservation, continuity across degradation boundaries, and consistency of restored color and texture with the surrounding regions.

Overall, SC-PAD effectively attenuates degradation while preserving the principal composition and major contours of the murals. Since semi-transparent degradation often retains useful underlying image information, the restoration process should avoid excessive modification of the original content. The results show that SC-PAD generally preserves the positions and shapes of major contours, consistent with its higher PSNR and SSIM values.

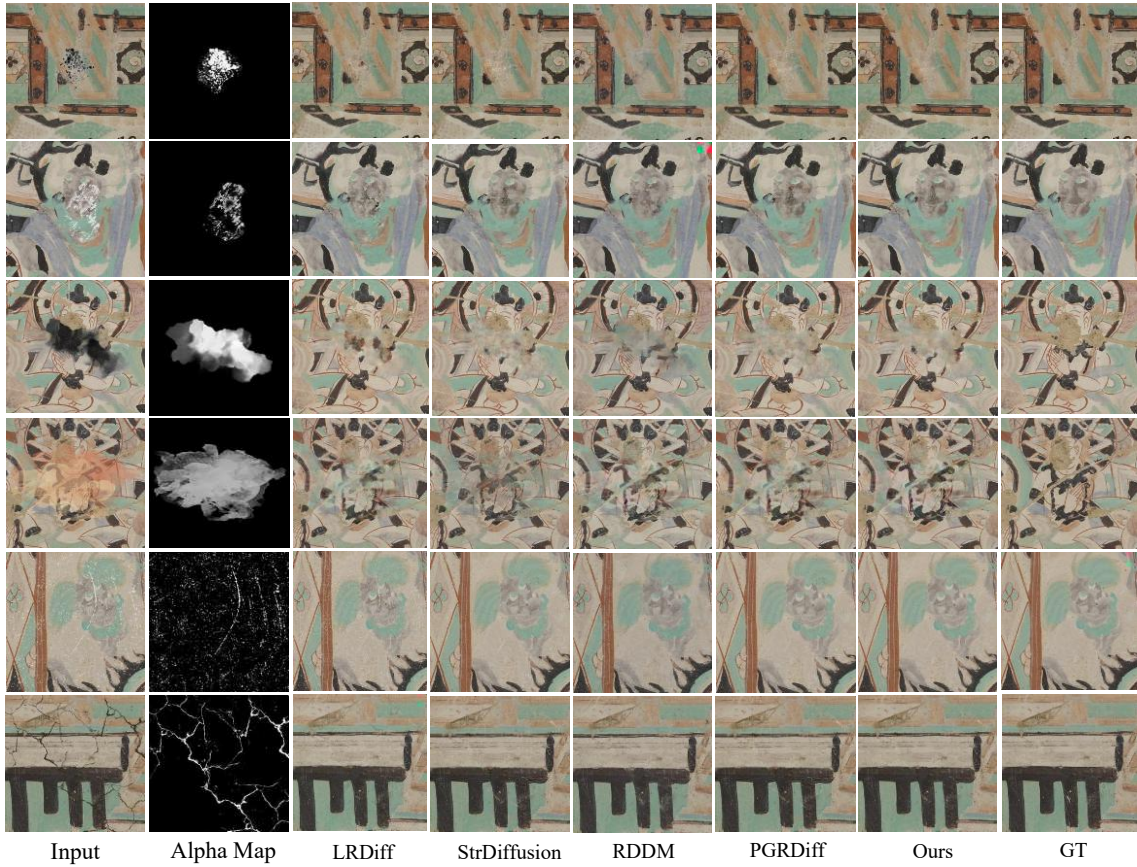


Figure 2 | Visual comparison of representative mural restoration results. Input denotes the degraded mural, Alpha Map the degradation severity map, and GT the reference image.

At degradation boundaries, continuous severity modeling captures gradual spatial variations in degradation intensity and provides finer-grained state control than a binary mask. Consequently, smoother transitions are maintained between restored and relatively healthy regions, particularly for semi-transparent degradation with gradual intensity variations, such as fading and contamination. SC-PAD also preserves the dominant contour directions and spatial organization in regions containing edges, line patterns, and local structures, demonstrating favorable structural consistency. Nevertheless, fine lines, dense textures, and severely degraded regions remain challenging, with occasional reductions in edge sharpness, texture displacement, or local contrast differences. The higher LPIPS of SC-PAD relative to PGRDiff also indicates that its improvements are not uniformly reflected in perceptual similarity. Overall, the quantitative and qualitative results demonstrate that SC-PAD provides favorable pixel fidelity, structural preservation, and

boundary continuity, while fine-detail reconstruction and severe degradation remain areas for further improvement.

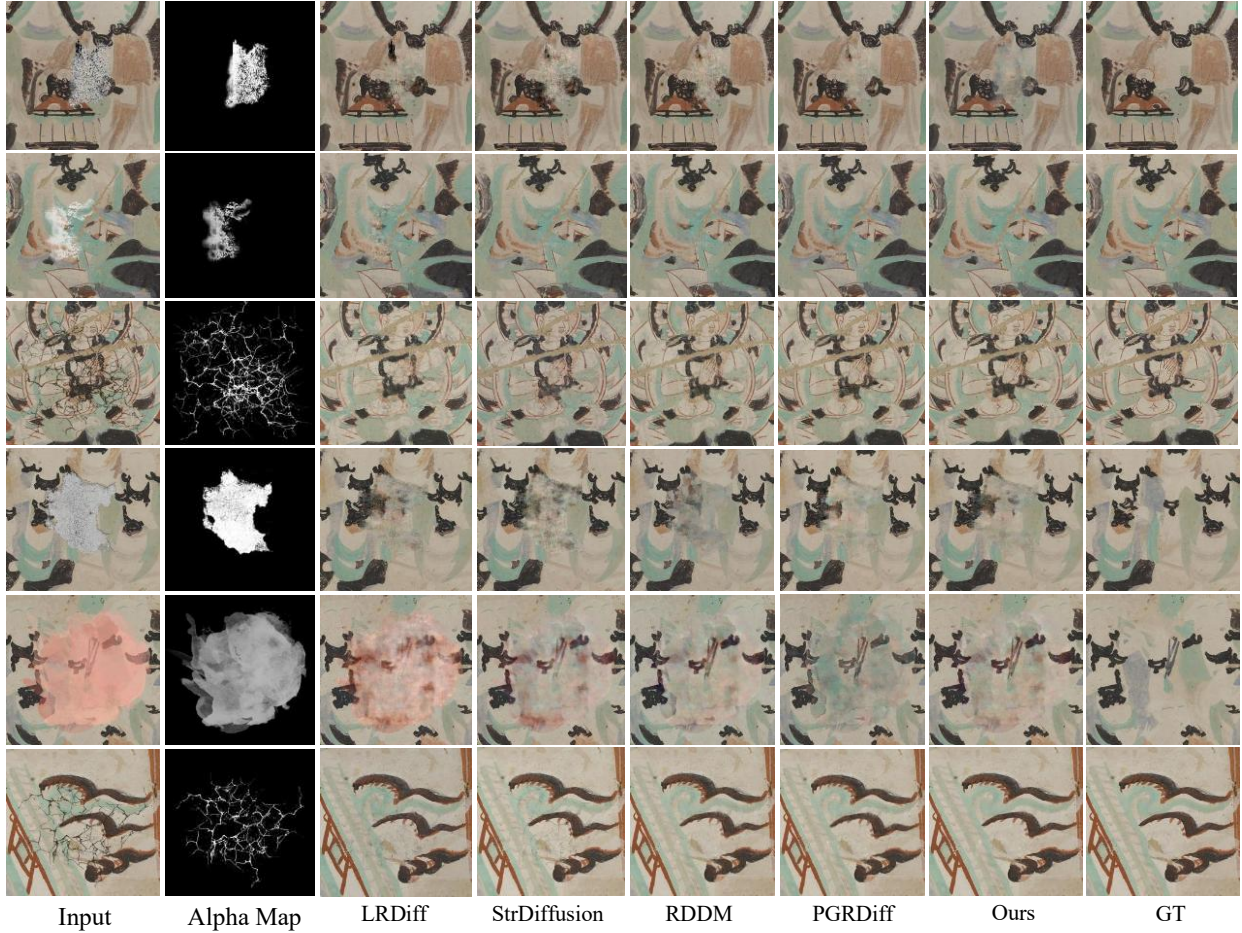


Figure 3 | Visual comparison of representative mural restoration results. Input denotes the degraded mural, Alpha Map the degradation severity map, and GT the reference image.

3.3 Effectiveness of continuous soft Alpha

The second experiment evaluates the effect of terminal-severity representation while keeping the asynchronous restoration structure unchanged. Table 3 compares the restoration performance obtained using three different severity representations: the binary mask, predicted soft Alpha, and ground-truth Alpha. The binary mask B provides only degradation support, the predicted soft Alpha $\hat{A}$ represents the estimated continuous severity used in deployable inference, and the ground-truth Alpha A^{gt} provides an oracle condition for assessing the potential performance under accurate severity information.

Table 3 | Effect of terminal-severity representation on restoration performance.

Severity input	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
Binary mask B	27.3421	0.8740	0.1151
Predicted soft Alpha $\hat{A}$	30.6122	0.9253	0.0712
Ground-truth Alpha A^{gt}	33.8661	0.9482	0.0412

Compared with the binary mask, predicted soft Alpha improves PSNR from 27.3421 to 30.6122 dB, corresponding to a gain of 3.2701 dB. SSIM increases from 0.8740 to 0.9253, while LPIPS decreases from 0.1151 to 0.0712. The consistent improvement across all three metrics indicates that continuous severity provides restoration-relevant information beyond binary degradation support. While the binary mask

identifies where degradation occurs, soft Alpha further describes how severe the degradation is at each location, enabling the restoration process to respond to spatial variations within the degraded region.

This advantage is closely related to the state construction of SC-PAD. The terminal severity A participates in the generation of the local restoration trajectory and determines the dynamic Alpha state and residual activation ratio. Replacing continuous severity with a binary mask removes intra-region severity variation and reduces the ability to distinguish pixels with different degradation levels. In contrast, predicted soft Alpha preserves gradual severity changes and therefore provides finer spatial control over the intermediate restoration states. The substantial improvement over the binary setting supports the effectiveness of continuous severity modeling in semi-transparent degradation restoration.

When ground-truth Alpha is used, PSNR further increases to 33.8661 dB, 3.2539 dB higher than that obtained with predicted soft Alpha. Meanwhile, SSIM increases from 0.9253 to 0.9482 and LPIPS decreases from 0.0712 to 0.0412. Since the restoration structure remains unchanged, this oracle result indicates that more accurate terminal-severity information can provide substantial additional performance gains. The gap between predicted and ground-truth Alpha therefore identifies severity estimation accuracy as an important factor limiting the current restoration performance.

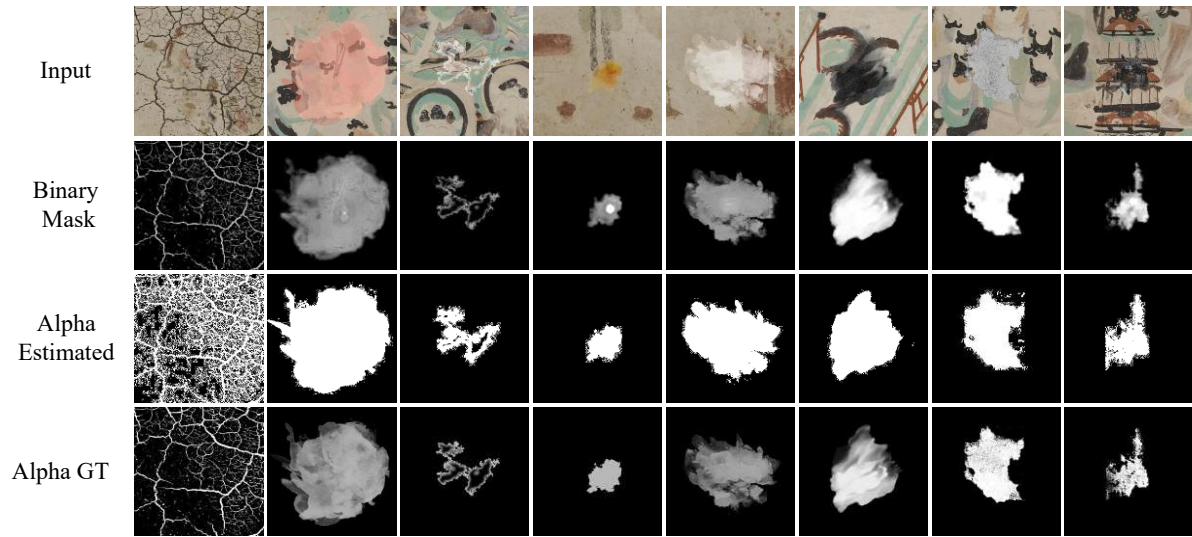


Figure 4 | Spatial comparison of the binary mask, predicted soft Alpha, and ground-truth Alpha.

The spatial comparison in Fig. 4 further illustrates these differences. The binary mask provides only discrete degradation boundaries and cannot represent severity variations within the affected region. Predicted soft Alpha recovers the principal degraded regions and their overall severity gradients, but deviations remain near irregular boundaries, high-severity areas, and isolated local responses. In comparison, ground-truth Alpha provides more accurate spatial severity information, which is consistent with its superior quantitative performance.

Overall, the results establish a clear progression from binary support to predicted continuous severity and then to oracle severity. Continuous soft Alpha substantially improves restoration performance over binary representation, while the remaining gap to ground-truth Alpha suggests that further improving severity estimation, particularly around complex boundaries and severe degradation regions, is a promising direction for enhancing SC-PAD.

3.4 Trajectory sensitivity to pixel-wise asynchronous time and dynamic Alpha

The third experiment fixes the continuous severity input and evaluates a 2x2 combination of synchronous/asynchronous local time and linear/composition-consistent Alpha states. The synchronous setting follows the shared global progress, whereas the asynchronous setting uses learned pixel-wise progress. The linear state is defined as $\alpha_t = g_t A$, while the composition-consistent state adopts $\alpha_t = 1 - (1 - A)^{g_t}$.

This experiment modifies the trajectory formulation during inference rather than training separate models for each configuration. Accordingly, Table 4 characterizes the sensitivity of restoration performance to temporal and state parameterizations and is not treated as a strict component ablation.

Table 4 | Sensitivity of restoration performance to local-time and Alpha-state formulations.

Setting	Local time	Alpha state	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
Sync + Linear	Synchronous	Linear	30.6401	0.9385	0.0517
Sync + Proposed	Synchronous	Composition-consistent	29.1956	0.9331	0.0585
Async + Linear	Asynchronous	Linear	31.1265	0.9430	0.0486
Async + Proposed (Full)	Asynchronous	Composition-consistent	31.5152	0.9445	0.0465

Pixel-wise asynchronous time yields consistent improvements under both Alpha-state formulations. For the linear formulation, PSNR increases from 30.6401 to 31.1265 dB and SSIM from 0.9385 to 0.9430, while LPIPS decreases from 0.0517 to 0.0486. For the composition-consistent formulation, PSNR increases from 29.1956 to 31.5152 dB, SSIM from 0.9331 to 0.9445, and LPIPS decreases from 0.0585 to 0.0465. The consistent changes across the three metrics indicate that spatially varying local progress is beneficial under both state formulations.

Synchronous time imposes a shared restoration progress across spatial locations, whereas asynchronous time permits spatially varying intermediate progress within the same global diffusion schedule. This formulation is better suited to spatially non-uniform degradation, in which different regions may require different restoration trajectories. The improvements obtained under both Alpha formulations are consistent with the intended role of pixel-wise temporal reparameterization. The influence of Alpha-state construction varies with the temporal setting. Under synchronous time, the linear formulation achieves better values on all three metrics. Under asynchronous time, the composition-consistent formulation achieves 31.5152 dB PSNR, 0.9445 SSIM, and 0.0465 LPIPS, compared with 31.1265 dB, 0.9430, and 0.0486 for the linear formulation. Thus, the effect of Alpha-state construction is associated with the temporal parameterization and does not exhibit a consistent independent advantage across the evaluated settings.

Overall, the sensitivity analysis consistently supports pixel-wise asynchronous time, as improvements in PSNR, SSIM, and LPIPS are observed under both Alpha-state formulations. In contrast, the composition-consistent Alpha state does not yield uniform metric improvements across temporal settings. Its primary role is to define an intermediate state satisfying endpoint consistency, monotonic evolution, and Alpha-composition properties. Therefore, the current experiment supports the empirical contribution of asynchronous local progress, while the independent performance contribution of the composition-consistent Alpha formulation requires further evaluation under a unified training protocol.

4. Conclusion

We proposed Severity-Conditioned Pixel-Wise Asynchronous Diffusion (SC-PAD) for restoration of semi-transparent mural degradation. The method separates continuous terminal severity, pixel-wise local restoration progress, and the dynamic Alpha state, allowing spatially non-uniform degradation to influence the evolution of the diffusion state rather than acting only as an auxiliary condition. Continuous severity is estimated from the degraded image and a coarse binary support, a bounded temporal reparameterization assigns different local progress to different pixels under a shared global schedule, and a terminal-severity-constrained dynamic Alpha controls progressive activation of the image-space residual. Experiments show that SC-PAD improves pixel fidelity and structural consistency in the overall comparison, that continuous soft Alpha clearly outperforms a binary mask across PSNR, SSIM, and LPIPS, and that pixel-wise asynchronous time consistently improves both Alpha-state parameterizations in the current trajectory sensitivity analysis. The oracle severity results further show that severity estimation remains an important performance bottleneck. Future work will expand evaluation across sites, materials, and degradation types, investigate domain generalization and uncertainty in severity estimation under complex real degradation, and

incorporate evidence such as multispectral imaging, material analysis, and conservation records to better connect image-state modeling with observable deterioration characteristics.

5. Patents

Author Contributions: Conceptualization, Y.W. and E.L.; methodology, Y.W. and E.L.; software, Y.W.; validation, Y.W., J.Z. and F.J.; formal analysis, Y.W. and L.Z.; investigation, Y.W. and F.J.; resources, E.L., L.Z. and F.J.; data curation, Y.W.; writing—original draft preparation, Y.W.; writing—review and editing, E.L., L.Z., J.Z. and F.J.; visualization, Y.W.; supervision, E.L.; project administration, E.L. and L.Z.; funding acquisition, E.L. and L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key R&D Program of Gansu Province under Grant 24YFFA029, in part by the Key R&D Project of Gansu Provincial Academy of Sciences under Grant 2025PTXM-03, in part by the Youth Fund Project of Gansu Academy of Sciences under Grant 2025QN-01, in part by the Gansu Provincial Department of Education University Teacher Innovation Fund Project 2026B-252, and in part by the Achievement Transformation and Technical Application Service Project of Gansu Academy of Sciences 2026CY-02.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- [1] Zeng, Z. et al. Virtual restoration of ancient tomb murals based on hyperspectral imaging. *Heritage Science* 12, 410 (2024).
- [2] Xu, Z. & Geng, C. Color restoration of mural images based on a reversible neural network: leveraging reversible residual networks for structure and texture preservation. *Heritage Science* 12, 351 (2024).
- [3] Jiménez-Desmond, D., Pozo-Antonio, J. S. & Arizzi, A. Present and future of chromatic reintegrations of wall paintings. *Journal of Cultural Heritage* 67, 237–247 (2024).
- [4] Ren, H., Sun, K., Zhao, F. et al. Dunhuang murals image restoration method based on generative adversarial network. *Heritage Science* 12, 39 (2024).
- [5] Zhao, F. et al. GAN-based heterogeneous network for ancient mural restoration. *Heritage Science* 12, 418 (2024).
- [6] Sun, Z., Lei, Y. & Wu, X. Ancient paintings inpainting based on dual encoders and contextual information. *Heritage Science* 12, 266 (2024).
- [7] Wei, X., Fan, B., Wang, Y. et al. Progressive enhancement and restoration for mural images under low-light and defective conditions based on multi-receptive field strategy. *npj Heritage Science* 13, 63 (2025).
- [8] Wu, X., Yu, Y. & Li, Y. A damage detection network for ancient murals via multi-scale boundary and region feature fusion. *npj Heritage Science* 13, 82 (2025).
- [9] Lyu, Q., Zhao, N., Song, J. et al. Mural inpainting via two-stage generative adversarial network. *npj Heritage Science* 13, 188 (2025).
- [10] Zhang, J., Bai, S., Zeng, X. et al. Supporting historic mural image inpainting by using coordinate attention aggregated transformations with U-Net-based discriminator. *npj Heritage Science* 13, 305 (2025).
- [11] Guan, J., Li, H., Cai, X. et al. Progressive generative mural image restoration based on adversarial structure learning. *npj Heritage Science* 13, 309 (2025).
- [12] Hu, J., Yu, Y. & Zhou, Q. GuidePaint: lossless image-guided diffusion model for ancient mural image restoration. *npj Heritage Science* 13, 118 (2025).
- [13] Zhao, F., Ren, H., Su, Z. et al. Diffusion-based heterogeneous network for ancient mural restoration. *npj Heritage Science* 13, 206 (2025).
- [14] Li, Y., Zhang, C., Li, Y., Sui, D. & Guo, M. An improved mural image restoration method based on diffusion model. *npj Heritage Science* 13, 347 (2025).
- [15] Lei, T., Li, Y., Cheng, Z. et al. Low-rank structure guided diffusion for Shaanxi temple mural restoration. *npj Heritage Science* 13, 295 (2025).
- [16] Tang, X., Sui, Y., Sun, K. & Xiang, L. DiffInpaint: line drawing guided murals restoration with diffusion model. *Measurement* 259, 119223 (2026).
- [17] Jia, R. et al. St-diffnet: Diffusion-based inpainting of Dunhuang murals with structural and textural guidance. *Multimedia Systems* 31, 467 (2025).
- [18] Lv, G., Wang, H., Wang, K. et al. Virtual restoration method of Kizil Grotto murals based on multimodal controlled diffusion models. *npj Heritage Science* 13, 554 (2025).
- [19] Xu, Z. et al. Automated mural restoration via semi supervised segmentation and prompt guided diffusion inpainting. *npj Heritage Science* 13, 504 (2025).

- [20] Jiang, C., Ren, T. & Cheng, Z. All-in-one mural restoration with prompt-guided residual diffusion. *npj Heritage Science* 13, 667 (2025).
- [21] Han, P. et al. DiffuMural: a diffusion model for Dunhuang murals restoration via multi-scale convergence and cooperative guidance. *npj Heritage Science* 14, 44 (2026).
- [22] Li, Y., Zhang, C., Li, Y. et al. DMEA-Net: a diffusion-based network for mural image restoration using enhanced multi-head contextual attention. *npj Heritage Science* 14, 348 (2026).
- [23] Yan, Y., Lv, Z., Jiang, C. et al. Learning mural restoration from degraded data via unsupervised low-rank residual diffusion. *npj Heritage Science* 14, 388 (2026).
- [24] Liu, H., Wang, Y., Qian, B., Wang, M. & Rui, Y. Structure Matters: Tackling the Semantic Discrepancy in Diffusion Models for Image Inpainting. *Proc. IEEE/CVF CVPR*, 8038–8047 (2024).
- [25] Liu, J., Wang, Q., Fan, H., Wang, Y., Tang, Y. & Qu, L. Residual Denoising Diffusion Models. *Proc. IEEE/CVF CVPR*, 2773–2783 (2024).
- [26] Ye, T., Chen, S., Chai, W. et al. Learning Diffusion Texture Priors for Image Restoration. *Proc. IEEE/CVF CVPR*, 2524–2534 (2024).
- [27] Garber, T. & Tirer, T. Image Restoration by Denoising Diffusion Models with Iteratively Preconditioned Guidance. *Proc. IEEE/CVF CVPR*, 25245–25254 (2024).
- [28] Cao, J., Shi, Y., Zhang, K., Zhang, Y., Timofte, R. & Van Gool, L. Deep Equilibrium Diffusion Restoration with Parallel Sampling. *Proc. IEEE/CVF CVPR*, 2824–2834 (2024).
- [29] Lin, X., He, J., Chen, Z. et al. DiffBIR: Toward Blind Image Restoration with Generative Diffusion Prior. *Proc. ECCV* (2024).
- [30] Wang, Y., Yang, W., Chen, X. et al. SinSR: Diffusion-Based Image Super-Resolution in a Single Step. *Proc. IEEE/CVF CVPR*, 25796–25805 (2024).
- [31] Yang, T., Wu, R., Ren, P., Xie, X. & Zhang, L. Pixel-Aware Stable Diffusion for Realistic Image Super-Resolution and Personalized Stylization. *Proc. ECCV*, 74–91 (2024).
- [32] Wang, J., Yue, Z., Zhou, S., Chan, K. C. K. & Loy, C. C. Exploiting Diffusion Prior for Real-World Image Super-Resolution. *International Journal of Computer Vision* 132, 5929–5949 (2024).
- [33] Li, Y. et al. The relationship between the microclimate and efflorescence of revealed mural paintings and the later protection strategy. *Science of the Total Environment* 922, 171337 (2024).
- [34] Guarnieri, N. et al. Rapid chromatic alteration of street art: Mechanisms of deterioration of the painting materials of the 20 years of freedom and democracy mural. *Dyes and Pigments* 239, 112733 (2025).
- [35] Guarnieri, N. et al. Preserving the contemporary mural “Musica Popolare” by Orticanoodles in Milan, Italy: Deterioration processes and protection performance of commercial coatings. *Journal of Cultural Heritage* 75, 326–332 (2025).
- [36] Luo, Z., Gustafsson, F. K., Zhao, Z., Sjölund, J. & Schön, T. B. Image restoration with mean-reverting stochastic differential equations. *Proceedings of the 40th International Conference on Machine Learning (ICML)*, PMLR 202, 23045–23066 (2023).
- [37] Huynh-Thu, Q. & Ghanbari, M. Scope of validity of PSNR in image/video quality assessment. *Electronics Letters* 44, 800–801 (2008).
- [38] Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 600–612 (2004).
- [39] Zhang, R., Isola, P., Efros, A. A., Shechtman, E. & Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. *Proc. IEEE/CVF CVPR*, 586–595 (2018).