Building Bridge Across the Time:
Disruption and Restoration of Murals In the Wild

Huiyang Shao[1,2]   Qianqian Xu[1*]   Peisong Wen[1]
Peifeng Gao[2]   Zhiyong Yang[3,4]   Qingming Huang[1,2,5,6*]

[1] Key Lab of Intell. Info. Process., Inst. of Comput. Tech., CAS
[2] School of Computer Science and Tech., University of Chinese Academy of Sciences
[3] State Key Lab of Info. Security, Inst. of Info. Engineering, CAS
[4] School of Cyber Security, University of Chinese Academy of Sciences
[5] BDKM, University of Chinese Academy of Sciences
[6] Peng Cheng Laboratory

shaohuiyang21@mails.ucas.ac.cn xuqianqian@ict.ac.cn
wenpeisong20z@ict.ac.cn gaopeifeng21@mails.ucas.ac.cn

yangzhiyong21@ucas.ac.cn qmhuang@ucas.ac.cn

Abstract

*In this paper, we focus on the mural-restoration task, which aims to detect damaged regions in the mural and repaint them automatically. Different from traditional image restoration tasks like in/out/blind-painting and image renovation, the corrupted mural suffers from more complicated degradation. However, existing mural-restoration methods and datasets still focus on simple degradation like masking. Such a significant gap prevents mural-restoration from being applied to real scenarios. To fill this gap, in this work, **we propose a systematic framework to simulate the physical process for damaged murals and provide a new benchmark dataset for mural-restoration**. Limited by the simplification of the data synthesis process, the previous mural-restoration methods suffer from poor performance in our proposed dataset. To handle this problem, **we propose the Attention Diffusion Framework (ADF) for this challenging task**. Within the framework, a damage attention map module is proposed to estimate the damage extent. Facing the diversity of defects, we propose a series of loss functions to choose repair strategies adaptively. Finally, experimental results support the effectiveness of the proposed framework in terms of both mural synthesis and restoration.*

1. Introduction

As an art form of recording ancient civilization's culture, murals describe people's life scenes and society's landscape, which have great scientific, historical, and artistic

---
*Corresponding authors.

Figure 1: **The disruption and restoration of murals.** We propose an simulation framework to generate a damaged mural image (top) and a restoration method to restore the corrupted image (bottom).

significance [54, 49]. However, with constant exposure to the terrible environment, ancient murals are inevitably destroyed by light, winds, bacteria, and human activities [30, 45]. According to the digital report of [30], manually restoring murals is still the primary restoration mode. This method is time-consuming due to large areas of missing content caused by wall peeling and color decay. Hence, it is necessary to design a mural restoration algorithm to accurately repair the structured and unstructured defects [45].

Prior to the deep learning era, researchers attempted to design geometry-based or patch-based methods to detect
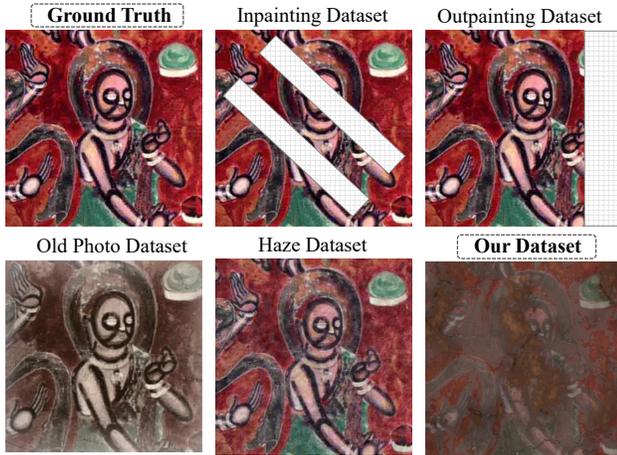
Figure 2: **Dataset comparison of image restoration tasks.**

the corrupted area automatically and fill them [2, 8, 3]. Limited by the scalability of traditional models, existing algorithms can only repair simple dust and spot which is not compatible with complex mural restoration scenarios. Therefore, these methods can only be used as auxiliary tools for manual repair. Recently, benefiting from the significant development of deep learning, many image inpainting techniques are proposed for repairing irregular damages and scratches of images [36, 59, 23, 33], some of which have attempted to restore corrupted mural images [7, 12, 48].

It is worth noting that some deep image restoration methods, *e.g.*, photo repair [46, 6, 47], noisy removal [33, 32, 25, 44] and image translation [60, 42, 21] have made great achievement in recent years. Unfortunately, these frameworks cannot be applied to mural image restoration. The main challenges of this task are three-fold: **(1)** The physical model of mural damage is more complicated. In previous works, the damaged images are generated by simple operations, *e.g.,* additive masking in in/out-painting and dehazing tasks; color degradation in old photo restoration [46, 47]. However, the mural damage process cannot be simply expressed by additive noise on the surface, but by complex unstructured and structured defects. See Fig. 2 for the demonstration of different tasks. **(2)** After a long period of ravage, all pixels in the mural are damaged less or more. Compared with other image restoration tasks where reliable information like the image structure of the original image is available, we have to detect the slightly damaged areas as a reliable reference to repair other areas. **(3)** The corrupted mural image is plagued by a series of different defects. This problem leads to different types and extents of missing information in damaged regions and requires adaptive restoration strategies.

For challenge **(1)**, we model the damage process of murals with 8 types of basic physics processes and roughly divide them into 2 categories: unstructured (color, scratch, blunt, grunge) and structured defects (cracking, breaking,

peeling, wearing). On top of this, we propose the first framework to generate realistic damaged murals with 3D open-source software `blender` [1]. The proposed framework can not only simulate the random noisy on the mural surface but also the destruction process of its carrier through the transformation of the mural 3D model.

For challenge **(2)**, limited by existing mural datasets, current algorithms may be biased toward ideal conditions and cannot handle real-world images. Therefore, in this work, we propose a diffusion-based model as a baseline for our dataset, namely *Attention Diffusion Framework (ADF)* to restore the damaged mural image. Moreover, we propose a lightweight module called *Damage Attention Map (DAM)* to predict the damage extent, the core of which is an attention block to adaptive adjust the receptive field.

For challenge **(3)**, a series of loss functions are proposed to form a comprehensive training target. Combined with the damage attention map, these losses can repair the missing information while preserving the basic structure and style of the input image.

In a nutshell, the main contributions are as follows:

- We propose a novel pipeline to generate damaged murals for mural-restoration, which is more consistent with the real physical process. On top of this, we provide a new mural-restoration benchmark dataset.

- According to the characteristic of mural-restoration, we propose a restoration framework based on a generative diffusion model, whose core components include a damage attention map module and a series of losses.

- Extensive experiments on synthetic and real-world datasets demonstrate the effectiveness of the proposed framework.

2. Related Work

Unstructured Degradation Image Restoration. Over the past two decades, restoring unstructured degradation has attracted attention in image-to-image translation tasks. Benefiting from the development of deep learning, end-to-end methods have outperformed conventional restoration approaches [53, 1, 31, 15, 4, 35] in recent years. [17] firstly presents an end-to-end image super-resolution method SR-CNN and [5] proposes DehazeNet for the haze removal task. Following these studies, numerous models and techniques have been proposed for image reconstruction[32, 52], super-resolution [44, 58], non-uniform particle removal [38, 55, 25], and photo renovation [46, 6, 47]. Most of these studies focus on completing specific degradation removal tasks. However, the degradation of mural data is rather complex, thus even the well-performed models in specific degradation removal tasks cannot be applied directly in practice.

---

[1]https://www.blender.org/
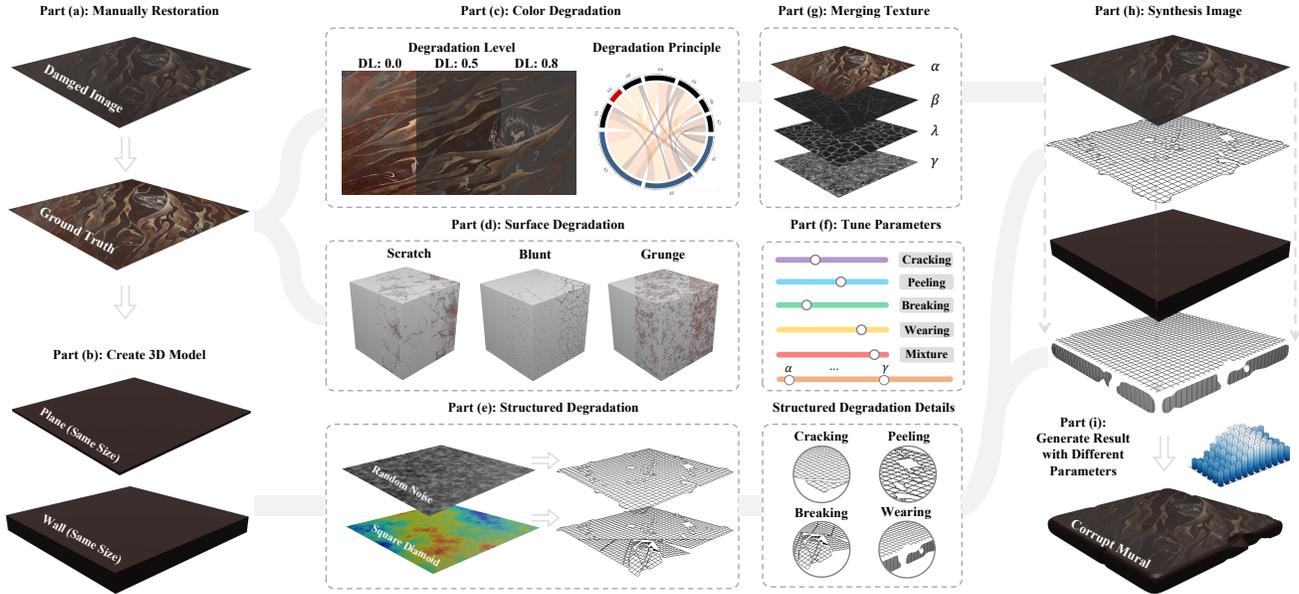
Figure 3: **An overview of simulation framework.** All the processes are handled in 3D model software `blender`. With the help of the texture engine, 4 types of unstructured defects are simulated including color degradation, scratch, blunt, and grunge. Benefiting from the real physics engine, 4 types of mural wall wear processes are simulated including cracking, peeling, breaking, and wearing. A video of the simulation process can be found in the supplementary materials.

Structured Degradation Image Restoration. Compared to unstructured degradation, researchers often describe structured degradation as the "image painting" task, which is more challenging. Thanks to recent developments in model architecture, *e.g.*, GAN [20], diffusion model [22], deep end-to-end methods have been able to learn semantic information from a large amount of data and fill in the missing content. By adding an irregular mask to image data, [33] firstly presents partial convolution to fill the missing content of the image. Following this method, many studies focus on considering local and global semantic contexts to get better inpainting results [56, 34, 41]. [43] and [26] introduce a unified framework to complete in/out painting tasks. However, in the mural-restoration task, repairing a terribly damaged image may use in&out painting techniques simultaneously. It is hard to combine different methods (*e.g.*, outpainting first, inpainting second) to repair it.

3. Proposed Dataset

In this section, we present a novel dataset designed to reflect the physical damage process of murals. Unlike previous work [57] using hard masks to cover the ground truth images, we use the 3D software `blender` to simulate the physical process. Detailed operations are present in section 3.1, and statistics are provided in section 3.3. We conduct experiments in section 5.1 to analysis the validity of our dataset compared to the previous mural dataset.

3.1. Simulation Process

Overview. The overall framework for generating damaged images is shown in Fig. 3. Given a restored mural image, we create a 3D mural plane and base wall according to the size of the image. (part-b in Fig. 3). Next, we perform 8 types of physical degradations on the built 3D model, which is roughly divided into two categories: unstructured and structured defects. Specifically, the unstructured defects appear on the superficial surface of murals, including color degradation (part-c) and surface degradation (part-d), while the structured defects reflect the destruction of mural walls (part-e). Different types of unstructured defects are mixed into a single texture with linear transformation (part-g) under various parameters (part-f). Finally, we generate the image by merging the damaged mural plane and mural wall (part-h). In this way, we can get amounts of corrupt murals by controlling the parameters (part-i).

Unstructured Degradation. In this paper, we mainly focus on four basic unstructured defects, including *color degradation* ψ_d, *grunge* ψ_g, *scratch* ψ_s and *blunt* ψ_b. For grunge degradation ψ_g, we use the Value noise algorithm [39] to generate a noisy mask image. Similarly, we use the Perlin noise [39] for blunt defects and Simplex noise [39] for scratch defects. To get better blending effects of defects, we use `blender` addon [2] for unstructured degradation. For color degradation ψ_d, we follow the instructions in [19] to summarize the reason and rules of color changes in Dunhuang murals.

---

[2]https://blendermarket.com/products/grungit
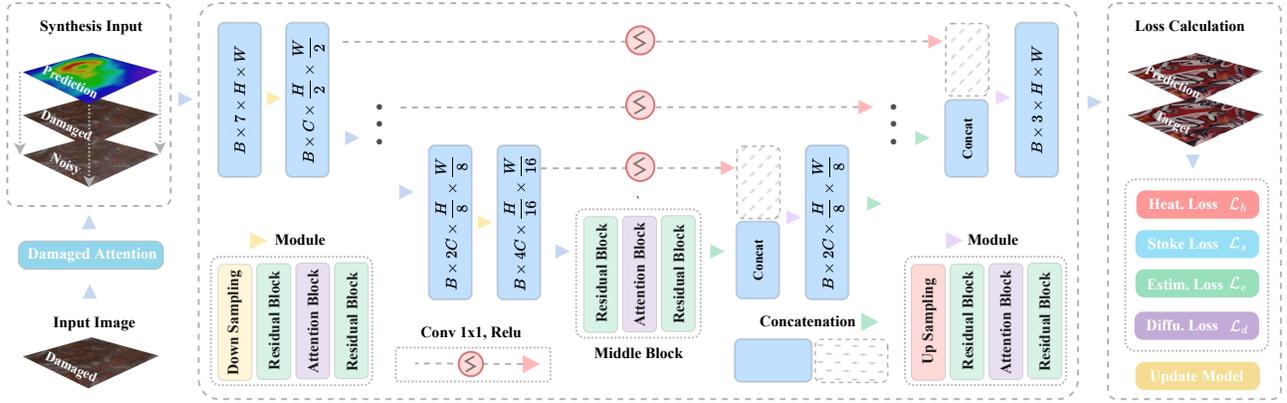
Figure 4: **An overview of model architecture.** Different from the guided diffusion model, we add a 1×1 convolution layer activated by ReLU on the skip connection. The model takes the synthesis images as the input and extracts high-level features through a U-like architecture. Then model outputs the result to calculate loss with the target image.

Opposite to the hard masking used in other datasets, the unstructured degradation is applied to all pixels; thus, we introduce a random linear transformation to mix these operations. Formally, then the whole unstructured degradation operation ψ is obtained by

$$
\begin{aligned}
\psi(\boldsymbol{x}) =& \boldsymbol{\alpha} \odot \psi_d(\boldsymbol{x}) + \boldsymbol{\beta} \odot \psi_s(\boldsymbol{x}) \\
& + \boldsymbol{\lambda} \odot \psi_b(\boldsymbol{x}) + \boldsymbol{\gamma} \odot \psi_g(\boldsymbol{x}),
\end{aligned}
\tag{1}
$$

where $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}\}$ are uniform random variables, and $\odot$ means element-wise product.

Structured Degradation. Structured degradation mainly destroys the physical structure of the mural walls. In this paper, we focus on four primary structured defects, including *cracking* (relatively large cracks), *peeling* (wall skin breaks and bends), *breaking* (many pieces of wall skin are broken), and *wearing* (wall edge wear). We use the `blender` addon Cracker [3] and OCD [4] to make cracking and breaking defects. For peeling and wearing defects, we use the diamond square algorithm [28] to generate the random height map and delete parts of the 3D model that height over the threshold.

3.2. Degradation Hyper-Parameters

As shown in Fig. 3, the damaged mural image is generated by controlling the parameters of all types of degradation. How to choose the optimal hyper-parameters to build a suitable dataset is crucial. In a specific mural restoration task, we can adjust the degradation parameters to generate a dataset and fine-tune the model to fit this scenario. We analyze the effect of each degradation and conduct experiments to verify it (please see Appendix. B.2 for details). We make the following remarks: **(1)** for unstructured degradation, the parameters of color decay and grunge have the greatest impact on the final restoration effect. **(2)** for structured degra-

dation, higher cracking and peeling can help models significantly improve their ability to repaint patterns. **(3)** scratch, blunt, breaking, and wearing degradation are more suitable for restoration of the terribly damaged murals.

3.3. Statistics

We collect 170 large manual restored, or healthy mural images (minimum: 1464×1454, maximum: 6096×7944) from the public data [24, 57]. Firstly, we randomly split each large image into multiple patches (4,411 patches for all large images). To mimic the diversity of mural corruptions, we construct three subsets in our dataset with different degrees of extent, including Light, Medium and Terrible. All the control parameters and details of three subsets are listed in the appendix. We combine these sub-datasets (13,233 images) and split them into a training set (10,584 images) and a test set (2,649 images). We ensure that there is no overlap between the training and test sets. Finally, we scale all images to (256×256) to speed up the training. Please see Appendix. B for more details on our proposed dataset.

4. Attention Diffusion Framework

To the best of our knowledge, existing algorithms cannot be applied to the mural-restoration task directly. The main challenges can be roughly listed: **(1)** In/out-painting-based methods require a local hard mask to locate the area to restore. However, the area to restore in mural is described by a global soft heatmap, which is unavailable for unseen data since the physical process is unknown. **(2)** There are amounts of structured (missing pixel) and unstructured defects (noisy pixel), which are distributed in a mixed form. In this case, the single loss used in previous work [32, 27, 43] is incapable to restore both of these massive defects.

For the first challenge, we propose a simple yet effective light module to identify and locate the damaged areas. Detailed operations are present in section 4.1.

---

[3]https://blendermarket.com/products/cracker
[4]https://blendermarket.com/products/ocd
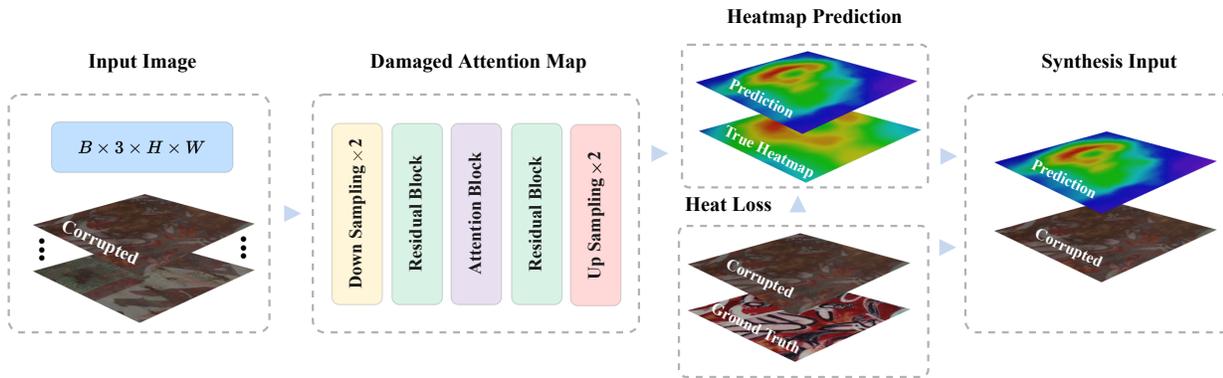
Figure 5: **The process of attention heatmap prediction.** The damaged attention map module takes corrupted images as input and extracts its high-level features to predict the damage map. Meanwhile, we will keep the content consistency between prediction and actual heatmap to improve the performance of DAM. Then the damage map of all types of defects is passed through the diffusion model.
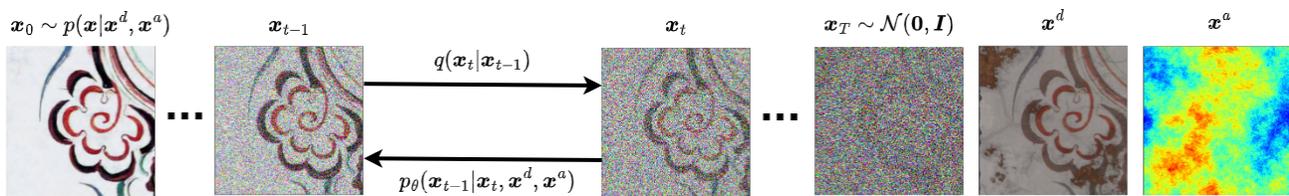
Figure 6: **The diffusion process of mural-restoration.** The forward process $q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ continually add Gaussian noise to $\boldsymbol{x}_{t-1}$ (from left to right), the reverse process $p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}^d, \boldsymbol{x}^a)$ aims to denoise the image $\boldsymbol{x}_t$ for given damaged image $\boldsymbol{x}^d$ and attention heatmap $\boldsymbol{x}^a$.

As for the second challenge, we propose three denoising losses to utilize variant information to improve the quality of output images as present in section 4.3. An overview of the attention diffusion framework is shown in Fig. 4, and the details of the model architecture are provided in section 4.4.

4.1. Damaged Attention Map Module

Different from the inpainting task, where most of the input pixels are copied from the original image, in the mural-restoration task, all pixels suffer from defects to some extent. On the other hand, some meaningful remnants are left in damaged areas, and the output image is expected to be consistent with these remnants. Such limitations make the task more challenging.

Moreover, the restoration methods depend on the damage extent across areas: for slightly damaged areas, only minor modifications are required, while for missing areas, we have to imagine consistent patches. Therefore, it is necessary to measure the damage first. Although the damage map can be obtained during the image generation process, it is unavailable during testing. To handle this issue, we propose a light module named ***Damage Attention Map (DAM)*** to automatically predict the damage map as shown in Fig. 5. First of all, to reduce the computational burden, the input image is downsampled into $1/4$ by two downsampling blocks, where each block consists of an average pooling layer and a 3×3 convolution layer. Next, the image fea-

ture is normalized and transferred into the locally-aware feature by a residual block. Afterward, to efficiently adjust the receptive field, we append an attention block for the globally-aware feature. Finally, we employ another residual block and an upsampling layer to predict the damaged attention map with the same size as the input image. Practically, the DAM module outputs a single heatmap for each image instead of one heatmap for each defect type to reduce the computation.

4.2. Conditional Denoising Diffusion Model

Diffusion model has achieved great success in image processing problems, including super-resolution [44, 58], in/outpainting [56, 34, 41] and colorization [18, 43]. Some studies show diffusion model fits the inverse problems well [27, 11], where the goal is to recover an image from noisy measurements. This model has been proved that can generate high-quality samples, is highly robust, and easy to train than other popular generative models [16], for example, GAN. Hence, we adopt Conditional Denoising Diffusion Model (CDDPM) as our backbone. It is worth noting that we can choose other models (*e.g.*, conditional GAN) as backbone equipped with our proposed module and still have good performance. We conduct some experiments in section 5.3 to verify it.

We further introduce the CDDPM [44] into the mural-restoration task. Given a damaged mural image $\boldsymbol{x}^d$ and

heatmap $\boldsymbol{x}^a$, we want to learn a parametric approximation $p(\boldsymbol{x}|\boldsymbol{x}^d, \boldsymbol{x}^a)$ through a stochastic iterative process. $p(\boldsymbol{x}|\boldsymbol{x}^d, \boldsymbol{x}^a)$ indicates a mapping from damaged image $\boldsymbol{x}^d$ to restored image $\boldsymbol{x}$. It is hard to approximate $p(\boldsymbol{x}|\boldsymbol{x}^d, \boldsymbol{x}^a)$ directly, hence, we use the diffusion process to divide the process into sub-tasks. We denote $\boldsymbol{x}_0 \sim p(\boldsymbol{x}|\boldsymbol{x}^d, \boldsymbol{x}^a)$ be restored image from $\boldsymbol{x}^d$ after T steps $(\boldsymbol{x}_T, \ldots, \boldsymbol{x}_t, \ldots, \boldsymbol{x}_0)$, $\boldsymbol{x}_T$ be noise image which drawn from Gaussian distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. CDDPM aims to learn a process $p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}^d, \boldsymbol{x}^a)$ which iteratively refines the noise image to get restored image $\boldsymbol{x}_0$. The process of CDDPM is shown in Fig. 6 and please see Appendix. A for more details.

4.3. Loss function

To accommodate the task characteristics of mural-restoration, we propose a series of loss functions in the rest of this section. Specifically, it is expected the generated image to **(1)** preserve the input image and its first-order information according to the extent of damage; **(2)** be consistent with the ground truth image distribution. To this end, we adopt the following losses. For target **(1)**, we propose *attention heatmap loss*, *estimation loss* and *edge loss*; for target **(2)**, we utilize *diffusion loss*.

Attention Heatmap Loss. In section 4.1, we propose DAM module to predict attention heatmap of different defects. To learn accurate attention, combining the damaged restored image, we propose the weighted heatmap loss:

$$\mathcal{L}_h = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} \exp(|\boldsymbol{x}_{(i,j)} - \boldsymbol{x}_{(i,j)}^d|) \\ \cdot \left[(\boldsymbol{x}_{(i,j)} - \boldsymbol{x}_{(i,j)}^d)^2 - \boldsymbol{x}^a \right]^2, \quad (2)$$

where $\boldsymbol{x}_{(i,j)}$ refers to the pixel of image $\boldsymbol{x}$ at position (i,j) (all image are scaled into $[0,1]$), and $\boldsymbol{x}^a$ is the predicted heatmap. Compared with the standard ℓ_2 regression loss, we multiply the loss by weights $\exp(|\boldsymbol{x}_{(i,j)} - \boldsymbol{x}_{(i,j)}^d|)$, such that severely damaged areas receive sufficient attention.

Estimation Loss. In CDDPM, we denote $\hat{\boldsymbol{x}}_0$ as estimate restoration image during training period. Hence, it's intuitive to keep the content consistency between $\hat{\boldsymbol{x}}_0$ and $\boldsymbol{x}$:

$$\mathcal{L}_e = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} \boldsymbol{x}_{i,j}^a (\hat{\boldsymbol{x}}_{0,(i,j)} - \boldsymbol{x}_{(i,j)})^2. \quad (3)$$

Here the loss is weighted by the attention heatmap to focus on severely damaged areas. Moreover, $\boldsymbol{x}^a$ and $\hat{\boldsymbol{x}}_0$ are jointly learned for a better trade-off.

Edge Loss. We notice that simply using the estimation loss will lead to blurry outputs. To restore damaged lines clearly, we propose an edge loss to keep frequency domain consistency:

$$\mathcal{L}_s = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} \exp(|\boldsymbol{x}_{(i,j)} - \boldsymbol{x}_{(i,j)}^d|) \\ \cdot \left[grad(\hat{\boldsymbol{x}}_{0,(i,j)}) - grad(\boldsymbol{x}_{(i,j)}) \right]^2, \quad (4)$$

where the $grad$ module contains Sobel kernel convolution and a mixture of RGB channels (coefficients are R:0.257, G:0.504, B:0.097 [40]). Since the Sobel operator will introduce a little high-frequency noise [40, 14, 13], we replace the weights $\boldsymbol{x}^a$ with $\exp(|\boldsymbol{x}_{(i,j)} - \boldsymbol{x}_{(i,j)}^d|)$ to avoid affecting the learning of DAM.

Diffusion Loss. Following the guided diffusion model, we use the following diffusion loss to ensure distribution consistency (f_θ is a generative model in CDDPM which is defined in Appendix. A):

$$\mathcal{L}_d = \| f_\theta(\boldsymbol{x}^d, \boldsymbol{x}^a, \boldsymbol{x}_t, \gamma_t) - \boldsymbol{\epsilon} \|_2^2. \quad (5)$$

Finally, we get the training loss for CDDPM by combining the above losses:

$$\mathcal{L}_{all} = \mathcal{L}_h + \mathcal{L}_s + \mathcal{L}_e + \mathcal{L}_d. \quad (6)$$

4.4. Model Architecture

In this subsection, we focus on the model architecture of f_θ. As a common model for medical image segmentation, Unet has made great achievements in recent years. The model structure is split into three parts, which contain a series of residual layers and sampling modules. Benefiting from the architecture, Unet can capture the high-level semantic features (left part), transform these features (mid part) and reconstruct low-level images (right part). Hence, Unet is a suitable architecture for the image-to-image task to repair damaged textures, and edges [22, 16, 10, 9]. To better fit the mural-restoration task, we employ a modified guided diffusion model (as shown in Fig. 4) as f_θ. Numerous experimental results demonstrate the effectiveness of the proposed architecture in the mural-restoration task.

5. Experiment

In this section, we will conduct experiments to assess the effectiveness of ADF in the mural-restoration task.

Competitors. We compare ADF with recent SOTA methods, including Old Photo Restoration [46], DDRM [27], SwinIR [32], Real-ESRGAN [43], Restormer [58] and ALL-In-One [29] (please see Appendix. C.1 for details). Since existing mural restoration methods cannot be directly applied to our dataset, we do not make comparisons.

Training Details. All experiments are conducted on an Ubuntu 16.04.1 server equipped with an Intel Xeon(R) Silver 4110 CPU and eight RTX 3090 GPUs.
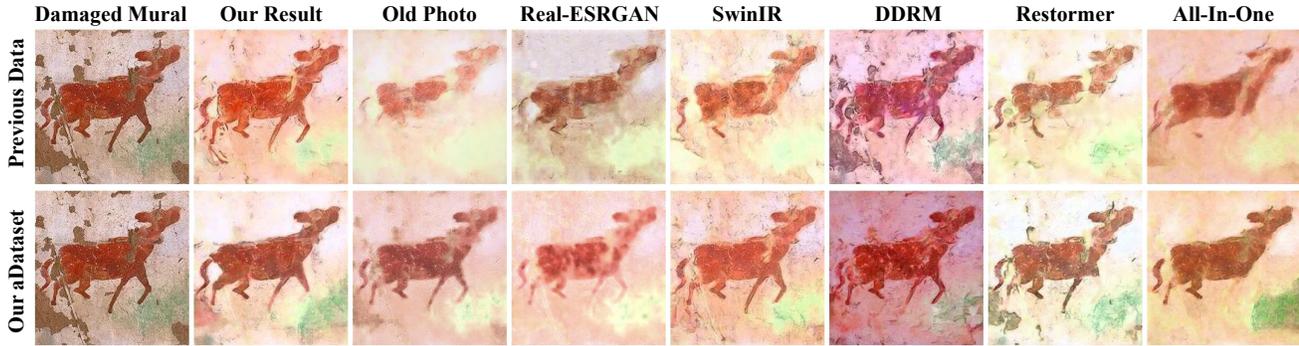
Figure 7: **The real world experiment of mural-restoration.** The first column contains the damaged mural; the first and second rows show the restoration results of the model which is trained on the previous and our proposed dataset respectively.
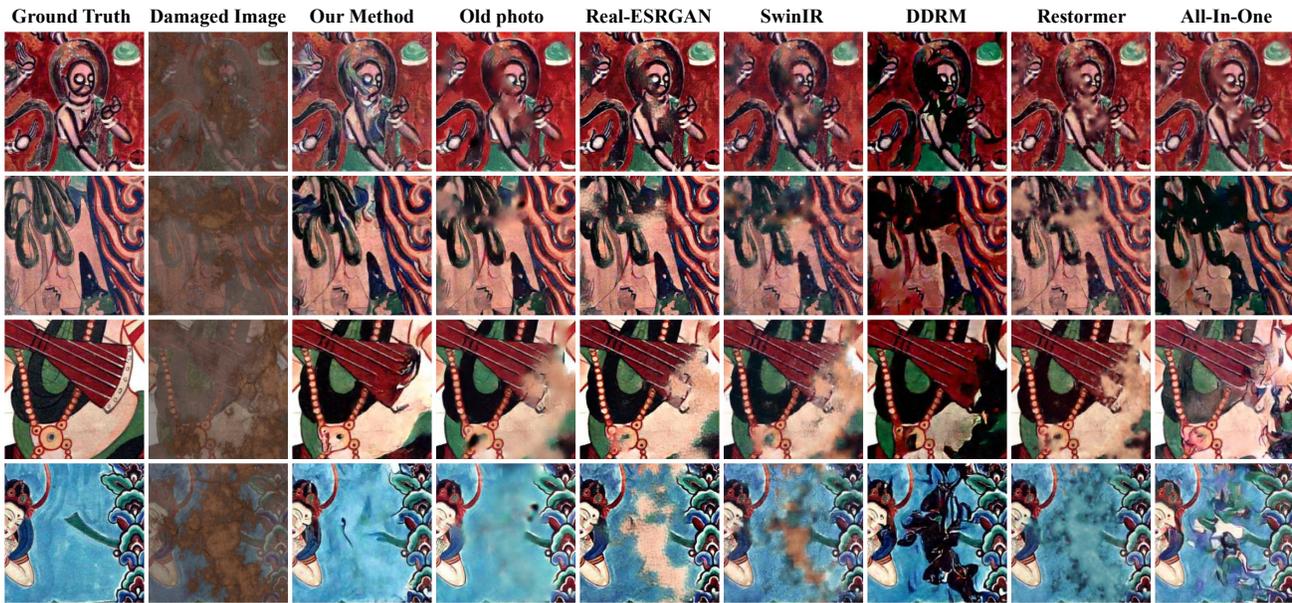
Figure 8: **The qualitative results of mural-restoration experiments.**

All codes are developed in Python 3.7.13, PyTorch 1.12.1 and torchvision 0.13.1 environment. We train our ADF model for 1000 iterations with a batch size of 64 (8 per GPU). Specially, we train our method from scratch and do not perform early stopping. Following [44], we choose Adam optimizer with a linear warmup schedule. The learning rate starts from 5×10^{-5} and increases to 10^{-4} over 500 iterations. We set weight decay and dropout rate to 10^{-5} and 0.2, respectively. All images are resized into 256×256 to reduce the computational burden.

5.1. Real-world Experiments

We collect some real-world paired images (including damaged and repaired images) to build test dataset. We conduct real-world experiments to demonstrate the validity of our proposed mural disruption simulation framework and restoration method ADF.

We train all baselines on our proposed and previous datasets [57] respectively. The comparison is shown in Fig. 7. The quantitative results of all baselines on real-world are listed in Tab. 1 and Tab. 2. According to the result of experiments, we make the following observations:

(1) Dataset Comparison. In Fig. 7, it's clear that most of the baselines trained on our dataset (second row) have better results compared to the previous dataset (first row).

This indicates that the previous pipeline does not describe the damage to the mural realistically. The existing mural dataset can not be used as training data to repair the real-world damaged mural. Therefore, our proposed synthesis pipeline is necessary and practical.

(2) Baseline Comparison. In Tab. 1 and Tab. 2, both for the previous dataset and ours, the ADF model achieves outstanding performance and generates satisfactory restoration images. This shows that our proposed DAM module and a series of loss functions are significantly effective in the mural restoration task. These modules can improve the model's ability to locate the damaged area and the samples' quality.

Method	PSNR ↑	SSIM ↑	IS ↑	FID ↓
Old Photo [46]	12.18	0.41	1.02	73.12
DDRM [27]	12.74	0.43	1.05	69.94
SwinIR [32]	13.51	0.46	1.04	72.18
ESRGAN [50]	11.83	0.36	1.01	74.27
Restormer [58]	12.95	0.38	1.03	71.82
All-in-one [29]	12.49	0.40	1.02	73.09
Ours method	**17.65**	**0.62**	**1.09**	**57.13**

Table 1: **The results on the real damaged mural dataset.** The performance of model which trained on the previous dataset. Upward arrows indicate that the image quality is proportional to metric, downward is otherwise. The best results are highlighted in **bold**.

Method	PSNR ↑	SSIM ↑	IS ↑	FID ↓
Old Photo [46]	15.23	0.46	1.05	57.54
DDRM [27]	15.91	0.49	1.06	59.93
SwinIR [32]	17.83	0.56	1.08	54.74
ESRGAN [50]	14.98	0.52	1.06	49.82
Restormer [58]	16.94	0.57	1.09	50.38
All-in-one [29]	15.26	0.47	1.05	51.23
Ours method	**19.08**	**0.68**	**1.12**	**47.59**

Table 2: **The results on the real damaged mural dataset.** The performance of the model trained on our dataset.

Method	PSNR ↑	SSIM ↑	IS ↑	FID ↓
Old Photo [46]	21.12	0.70	1.13	48.11
DDRM [27]	17.39	0.62	1.17	54.19
SwinIR [32]	22.03	0.78	1.16	43.95
Real-ESRGAN [50]	22.36	0.77	1.15	37.36
Restormer [58]	21.71	0.75	1.14	37.23
All-in-one [29]	22.52	0.68	1.16	39.27
Ours method	**23.64**	**0.80**	**1.18**	**35.88**

Table 3: **Quantitative results on our proposed dataset.**

5.2. Benchmark Experiments

Main Results. We conduct similar image restoration studies on mural images comparing our proposed method ADF. Fig. 8 shows the results of all competitors on our benchmark dataset. According to the figures, we have the following observations: **(1)** Most methods can process simple unstructured degradation. However, for terribly damaged defects, *e.g.*, peeling, breaking, and cracking, the previous restoration methods prefer to blur them instead of repainting them. **(2)** Our proposed method ADF outperforms all the competitors in most cases. This approach can repair the missing content and restore the texture and stoke of the damaged mural. The results demonstrate that our method can be applied in practice.

Tab. 3 shows some automated metrics, including PSNR,

Model	Loss function			Metrics			
	$\mathcal{L}_h$	$\mathcal{L}_s$	$\mathcal{L}_e$	PSNR↑	SSIM↑	IS↑	FID↓
Origin		✓		19.28	0.67	1.09	54.00
			✓	22.30	0.72	**1.15**	49.95
		✓	✓	**22.59**	**0.76**	1.14	**47.07**
Ours	✓			19.07	0.77	1.12	47.92
		✓		20.02	0.76	1.16	45.31
	✓	✓		20.65	0.72	1.17	44.74
	✓		✓	22.76	0.79	1.17	38.63
	✓	✓	✓	**23.64**	**0.80**	**1.18**	**35.88**

Table 4: **Performance of ablation studies under different settings.** Origins indicate that the guided diffusion without DAM module.
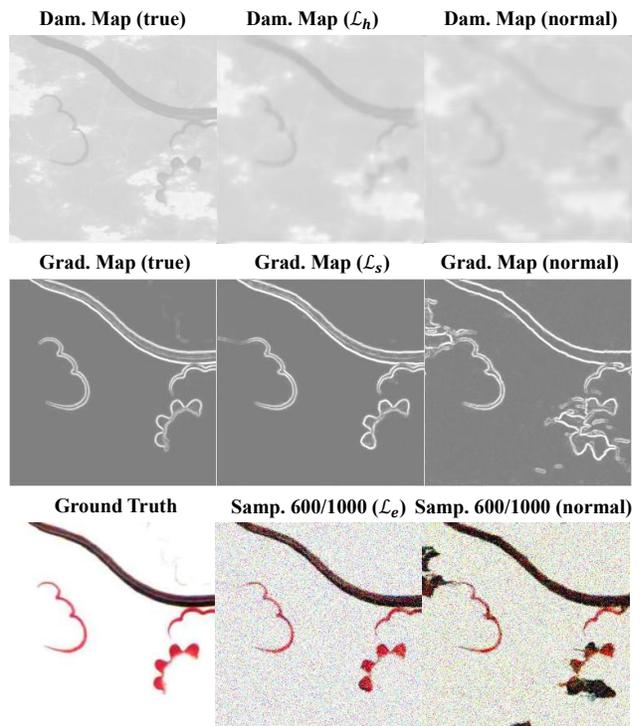
Figure 9: Abalation studies on our proposed techniques.

SSIM, IS, and FID for mural-restoration task [51]. We can conclude that our method ADF can achieve the highest sample quality scores than other competitors. The main reason is that previous works are designed to restore specific or several defects, which are hard to apply in this benchmark.

5.3. Ablation Studies.

In this section, we conduct extensive experiments to validate the effect of three components: DAM module, loss functions, and the backbone (result is shown in Tab. 4).

(1) DAM module. In general, when the loss functions are all set to the same, our proposed framework will perform better than the original model in most metrics. This demon-
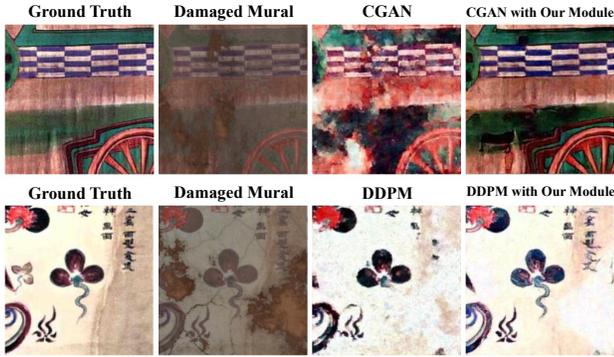
Figure 10: Abalation studies on different backbone.

strates our proposed DAM module can efficiently overcome the challenges in mural restoration.

(2) Loss function. When $\mathcal{L}_h$ and DAM are used together, the effect is even better than when they exist alone. This result indicates that the $\mathcal{L}_h$ we proposed is reasonably matched with DAM (as shown in Fig. 9, first row).

Moreover, the edge of the generated image, which uses $\mathcal{L}_s$ (Fig. 9, second row, middle) is clear than the normal one (right). This demonstrates $\mathcal{L}_s$ can efficiently reconstruct the damaged edge of the mural image.

In most cases, $\mathcal{L}_e$ can improve the performance of the model and accelerate the model generation process (Fig. 9, third row). Compared to $\mathcal{L}_h$ (locate damaged areas) and $\mathcal{L}_s$ (enhance edges), $\mathcal{L}_e$ is the only loss that constrains content consistency directly, while others focus on auxiliary information. Hence, $\mathcal{L}_e$ is critical to the performance improvement of the model.

(3) Backbone. From the result of Fig. 10, it's clear that for both types of origin model, our module can efficiently improve the performance. This means our proposed DAM and losses are effective and can be migrated to other backbones besides CDDPM, *e.g,* Conditional GAN. Please see the Appendix. C.2 for details of the experiment setup.

6. Conclusion

In conclusion, we focus on a complex image restoration task, namely mural restoration. To address the lack of suitable paired mural-restoration datasets, with the help of the physics and texture engine of the 3D software `blender`, we propose a systematic framework to simulate the actual damage process and generate images. Moreover, we introduce an effective framework called ADF to complete mural restoration tasks. Finally, we conduct numerous experiments to demonstrate that our proposed method and benchmark dataset are valid for this task.

The benchmark dataset proposed by us belongs to synthetic images, and there is still a certain gap between it and the real damaged images. Therefore, manual annotation of real damaged datasets significantly contributes to the mural-restoration task. Our future work will focus on improving the simulation framework of the damage process.

References

[1] S Derin Babacan, Rafael Molina, and Aggelos K Katsaggelos. Total variation super resolution using a variational approach. In *ICIP*, pages 641–644. IEEE, 2008. 2

[2] Marcelo Bertalmio, Andrea L Bertozzi, and Guillermo Sapiro. Navier-stokes, fluid dynamics, and image and video inpainting. In *CVPR*, volume 1, pages I–I. IEEE, 2001. 2

[3] Folkmar Bornemann and Tom März. Fast image inpainting based on coherence transport. *Journal of Mathematical Imaging and Vision*, 28(3):259–278, 2007. 2

[4] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *CVPR*, volume 2, pages 60–65. Ieee, 2005. 2

[5] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE TIP*, 25(11):5187–5198, 2016. 2

[6] Haoming Cai, Jingwen He, Yu Qiao, and Chao Dong. Toward interactive modulation for photo-realistic image restoration. In *CVPR*, pages 294–303, 2021. 2

[7] Jianfang Cao, Zibang Zhang, Aidi Zhao, Hongyan Cui, and Qi Zhang. Ancient mural restoration based on a modified generative adversarial network. *Heritage Science*, 8(1):1–14, 2020. 2

[8] Tony F Chan and Jianhong Shen. Nontexture inpainting by curvature-driven diffusions. *Journal of visual communication and image representation*, 12(4):436–449, 2001. 2

[9] Zuyao Chen, Runmin Cong, Qianqian Xu, and Qingming Huang. Dpanet: Depth potentiality-aware gated attention network for rgb-d salient object detection. *IEEE Transactions on Image Processing*, 30:7012–7024, 2020. 6

[10] Zuyao Chen, Qianqian Xu, Runmin Cong, and Qingming Huang. Global context-aware progressive aggregation network for salient object detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 10599–10606, 2020. 6

[11] Hyungjin Chung, Dohoon Ryu, Michael T McCann, Marc L Klasky, and Jong Chul Ye. Solving 3d inverse problems using pre-trained 2d diffusion models. *arXiv preprint arXiv:2211.10655*, 2022. 5

[12] Irina-Mihaela Ciortan, Sony George, and Jon Yngve Hardeberg. Colour-balanced edge-guided digital inpainting: Applications on artworks. *Sensors*, 21(6):2091, 2021. 2

[13] Runmin Cong, Qinwei Lin, Chen Zhang, Chongyi Li, Xiaochun Cao, Qingming Huang, and Yao Zhao. CIRNet: Cross-modality interaction and refinement for RGB-D salient object detection. *IEEE Trans. Image Process.*, 31:6800–6815, 2022. 6

[14] Runmin Cong, Qi Qin, Chen Zhang, Qiuping Jiang, Shiqi Wang, Yao Zhao, and Sam Kwong. A weakly supervised learning framework for salient object detection via hybrid labels. *IEEE Trans. Circuits Syst. Video Technol.*, 33(2):534–548, 2023. 6

[15] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE TIP*, 16(8):2080–2095, 2007. 2

[16] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *NeurIPS*, 34:8780–8794, 2021. 5, 6

[17] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, pages 184–199. Springer, 2014. 2

[18] Elisa Mariarosaria Farella, Salim Malek, and Fabio Remondino. Colorizing the past: Deep learning for the automatic colorization of historical aerial images. *Journal of Imaging*, 8(10):269, 2022. 5

[19] Qingping Feng, Xiaojun Ma, Xiaojun Zhang, Zuixiong Li, and Shi Li. Studies on microbial factor on color change of dunhuang mural. i. classification of microbes on color-changed mural and property of some typical species. *Wei Sheng wu xue bao= Acta Microbiologica Sinica*, 38(1):52–56, 1998. 3

[20] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. 3

[21] Junlin Han, Mehrdad Shoeiby, Lars Petersson, and Mohammad Ali Armin. Dual contrastive learning for unsupervised image-to-image translation. In *CVPR*, pages 746–755, 2021. 2

[22] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 33:6840–6851, 2020. 3, 6, 12

[23] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, pages 1125–1134, 2017. 2

[24] istockphoto. istockphoto. https://www.istockphoto.com/. 2022. 4

[25] Yeying Jin, Wending Yan, Wenhan Yang, and Robby T Tan. Structure representation network and uncertainty feedback learning for dense non-uniform fog removal. *arXiv preprint arXiv:2210.03061*, 2022. 2

[26] Changho Jo, Woobin Im, and SungEui Yoon. In-n-out: Towards good initialization for inpainting and outpainting. In *The British Machine Vision Conference (BMVC)*, 2021. 3

[27] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *arXiv preprint arXiv:2201.11793*, 2022. 4, 5, 6, 8, 16

[28] Oleg Kiselyov and Toshihiro Nakayama. New view on plasma fractals. *okmij.org*, 2021. 4

[29] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17452–17462, 2022. 6, 8, 16

[30] Jiamin Li, Hui Zhang, Zaixuan Fan, Xiang He, Shimin He, Mingyuan Sun, Yimin Ma, Shiqiang Fang, Huabing Zhang, and Bingjian Zhang. Investigation of the renewed diseases on murals at mogao grottoes. *Heritage Science*, 1(1):1–9, 2013. 1

[31] Stan Z Li. *Markov random field modeling in image analysis*. Springer Science & Business Media, 2009. 2

[32] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCV*, pages 1833–1844, 2021. 2, 4, 6, 8, 16

[33] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In *ECCV*, pages 85–100, 2018. 2, 3

[34] Hongyu Liu, Bin Jiang, Yi Xiao, and Chao Yang. Coherent semantic attention for image inpainting. In *ICCV*, pages 4170–4179, 2019. 3, 5

[35] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman. Non-local sparse models for image restoration. In *ICCV*, pages 2272–2279. IEEE, 2009. 2

[36] Xiaojiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. *NeurIPS*, 29, 2016. 2

[37] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. 16

[38] Siqi Ni, Xueyun Cao, Tao Yue, and Xuemei Hu. Controlling the rain: From removal to rendering. In *CVPR*, pages 6328–6337, 2021. 2

[39] Ken Perlin. Making noise. *GDC Talk, 1999*, 1999. 3

[40] Maria MP Petrou and Costas Petrou. *Image processing: the fundamentals*. John Wiley & Sons, 2010. 6

[41] Yurui Ren, Xiaoming Yu, Ruonan Zhang, Thomas H Li, Shan Liu, and Ge Li. Structureflow: Image inpainting via structure-aware appearance flow. In *ICCV*, pages 181–190, 2019. 3, 5

[42] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. Encoding in style: a stylegan encoder for image-to-image translation. In *CVPR*, pages 2287–2296, 2021. 2

[43] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–10, 2022. 3, 4, 5, 6, 12, 16

[44] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE TPAMI*, 2022. 2, 5, 7

[45] Meijun Sun, Dong Zhang, Zheng Wang, Jinchang Ren, Bo-long Chai, and Jizhou Sun. What's wrong with the murals at the mogao grottoes: a near-infrared hyperspectral imaging method. *Scientific reports*, 5(1):1–10, 2015. 1

[46] Ziyu Wan, Bo Zhang, Dongdong Chen, Pan Zhang, Dong Chen, Jing Liao, and Fang Wen. Bringing old photos back to life. In *CVPR*, pages 2747–2757, 2020. 2, 6, 8, 16

[47] Ziyu Wan, Bo Zhang, Dongdong Chen, Pan Zhang, Dong Chen, Fang Wen, and Jing Liao. Old photo restoration via deep latent space translation. *IEEE TPAMI*, 2022. 2

[48] Nianyi Wang, Weilan Wang, Wenjin Hu, Aaron Fenster, and Shuo Li. Thanka mural inpainting based on multi-scale adaptive partial convolution and stroke-like mask. *IEEE TIP*, 30:3720–3733, 2021. 2

[49] Xiaoguang Wang, Ningyuan Song, Lu Zhang, and Yanyu Jiang. Understanding subjects contained in dunhuang mural images for deep semantic annotation. *Journal of documentation*, 2017. 1

[50] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *ICCV*, 2021. 8

[51] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004. 8

[52] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *CVPR*, pages 17683–17693, 2022. 2

[53] Yair Weiss and William T Freeman. What makes a good model of natural images? In *CVPR*, pages 1–8. IEEE, 2007. 2

[54] Hong wu. Reborn in paradise: a case study of dunhuang sūtra painting and its religious, ritual and artistic context. *Orientations*, 23(5):52–60, 1992. 1

[55] Yuntong Ye, Yi Chang, Hanyu Zhou, and Luxin Yan. Closing the loop: Joint rain generation and removal via disentangled image translation. In *CVPR*, pages 2053–2062, 2021. 2

[56] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In *CVPR*, pages 5505–5514, 2018. 3, 5

[57] Tianxiu Yu, Shijie Zhang, Cong Lin, Shaodi You, Jian Wu, Jiawan Zhang, Xiaohong Ding, and Huili An. Dunhuang grottoes painting dataset and benchmark. *arXiv preprint arXiv:1907.04589*, 2019. 3, 4, 7

[58] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, pages 5728–5739, 2022. 2, 5, 6, 8, 16

[59] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *CVPR*, pages 3929–3938, 2017. 2

[60] Pan Zhang, Bo Zhang, Dong Chen, Lu Yuan, and Fang Wen. Cross-domain correspondence learning for exemplar-based image translation. In *CVPR*, pages 5143–5153, 2020. 2