Towards Automated Chinese Ancient Character Restoration: A Diffusion-Based Method with a New Dataset

Haolong Li[1], **Chenghao Du**[1], **Ziheng Jiang**[1], **Yifan Zhang**[1], **Jiawei Ma**[1], **Chen Ye**[1, 2, *]

[1]Department of Computer Science and Technology, Tongji University, Shanghai, China
[2]The Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University
{furlongli322, zyf20020730, jiavve1.10}@gmail.com, {2133034, jiangzh, yechen}@tongji.edu.cn

Abstract

Automated Chinese ancient character restoration (ACACR) remains a challenging task due to its historical significance and aesthetic complexity. Existing methods are constrained by non-professional masks and even overfitting when training on small-scale datasets, which hinder their interdisciplinary application to traditional fields. In this paper, we are proud to introduce the Chinese Ancient Rubbing and Manuscript Character Dataset (ARMCD), which consists of 15,553 real-world ancient single-character images with 42 rubbings and manuscripts, covering the works of over 200 calligraphy artists spanning from 200 to 1,800 AD. We are also dedicated to providing professional synthetic masks by extracting localized erosion from real eroded images. Moreover, we propose DiffACR (Diffusion model for automated Chinese Ancient Character Restoration), a diffusion-based method for the ACACR task. Specifically, we regard the synthesis of eroded images as a special form of cold diffusion on uneroded ones and extract the prior mask directly from the eroded images. Our experiments demonstrate that our method comprehensively outperforms most existing methods on the proposed ARMCD. Dataset and code are available at https://github.com/lhl322001/DiffACR.

Introduction

Ancient character restoration has traditionally been a specialized skill reserved for restoration experts. With their in-depth knowledge in various disciplines, such as archaeology and art, restoration experts often dedicate days, months, or even years to meticulously restoring an artifact or even a single character (Tétreault 2013). In particular, the restoration of Chinese ancient characters has garnered widespread attention due to their distinctiveness and severe erosion. As empires rose and fell, the complex structures and strokes of various ancient Chinese characters, especially those on the inscriptions, have been significantly eroded, making their restoration increasingly challenging. Fortunately, with the development of computer technologies, automated Chinese ancient character restoration (ACACR) has become a viable solution.

Figure 1: Examples of automated Chinese ancient character restoration through our *erosionfication* and restoration processes.

ACACR is essentially a cross-disciplinary visual inpainting task aimed at automatically reconstructing ancient Chinese character images with domain-specific knowledge. Existing methods mostly focus on handwriting (Zhong et al. 2017; Song, Li, and Wang 2020; Wang et al. 2021) or stylized (Li et al. 2021) forms. The former usually use specific handwriting datasets (Wang et al. 2009; Liu et al. 2011), while the latter typically employ fonts in standardized formats (*e.g.* TrueType format) to generate datasets. However, these large-scale datasets differ significantly from Chinese

*Corresponding author

Figure 2: Real anicent characters(Real) compared with handwriting(HW) and standardized characters(SD).

ancient characters in spatial distribution and local details, as shown in Figure 2. The remaining ACACR methods are consistently hindered by insufficient paired data and non-professional masks. Xing *et al.* (Xing and Ren 2023) have to collect data solely from two inscriptions and mask the center block only. Shi *et al.* (Shi et al. 2022b) attempt to synthesize masks using mixed Gaussian noise, but only use 1467 ancient characters while training. Using such limited datasets and non-professional masks may lead to overfitting and poor performance in restoring real-world ancient characters.

To solve the problem, we are proud to introduce the Chinese Ancient Rubbing and Manuscript Character Dataset (ARMCD), which is specifically constructed for the ACACR task. Rubbings and manuscripts are important carriers of Chinese cultural heritage and historical information (QIu 2023), which are widely used in the fields of art, archaeology, historiography, cultural studies, *etc.* Our ARMCD is collected from 42 authoritative, well-known, and representative Chinese ancient rubbings and manuscripts, including 15,093 real-world images of uneroded ancient Chinese characters and 460 images of eroded characters. Our collections from these rubbings and manuscripts record the works of over 200 calligraphy artists from the Jin Dynasty to the Qing Dynasty. Moreover, we investigate and analyze the actual effects of erosion on real-world ancient characters as shown in Figure 4. Based on our research, we introduce a novel professional mask for the ACACR task, which is synthesized by overlaying local erosion region masks from real-world eroded images.

Furthermore, we propose a Diffusion Model for Automated Chinese Ancient Character Restoration (DiffACR) for the ACACR task. Our key insight is that real erosion can be characterized as a kind of image degradation, called *erosionfication*. Similar to *snowification* (Hendrycks and Dietterich 2019), *erosionfication* use local erosion masks to degrade an uneroded image to an eroded image, eventually into a completely black image. Inspired by Bansal *et al.* (Bansal et al. 2022), we regard the *erosionfication* as a form of cold diffusion. In addition, given the lack of auxiliary information such as masks in real-world scenarios (Li et al. 2021), we propose extracting the prior mask directly from eroded images rather than relying on synthetic masks provided by ARMCD. Our approach is simple yet effective: we use the set of white pixels in the eroded images as the prior mask to guide the restoration process, since the set of white pixels remains constant throughout the process.

To sum up, our major contributions can be concluded as:

- We introduce a new automated Chinese ancient character restoration dataset ARMCD, including 15,553 real-world ancient single-character images, and provide professional synthetic masks for the ACACR task.
- We propose a diffusion-based method called DiffACR, which performs *erosionfication* as a form of cold diffusion and utilizes the prior mask to improve the restoration performance.
- Our experiments demonstrate that our diffusion-based method achieves satisfactory restoration results on all metrics, and our synthetic masks are recognized by experts.

Related Work

Chinese Character Dataset

Existing Chinese character datasets are mainly used for Optical Character Recognition (OCR), such as HCL2000 (Zhang et al. 2009), SCUT-COUCH2009 (Jin et al. 2011), HWDB (Wang et al. 2009), CASIA-AHCDB (Xu et al. 2019), ACCID (Diao et al. 2023), *etc.* Due to the lack of a large-scale real-world Chinese ancient character restoration dataset, most current methods for the ACACR task are still trained on these datasets with common masks in image inpainting (*e.g.* rectangular block masks). The remaining methods (Xing and Ren 2023; Shi et al. 2022b) are also limited to insufficient paired data and non-professional masks. In contrast, we construct the ARMCD and provide professional synthetic masks.

Automated Chinese Ancient Character Restoration

Chinese ancient characters are valued and widely studied in art and archaeology. However, over a long period of time, they suffer from water, wind, *etc*. (Tétreault 2013). In this case, the restoration usually involves a highly complex, time-consuming, and professional workflow. Traditional methods are mostly based on stroke feature extraction (Xu et al. 2009; Wang et al. 2012). With the growth of generative models, VAE-base methods (Lv and Liu 2018; Shi et al. 2022a) and GAN-based methods (Zhong et al. 2017; Su et al. 2022; Li et al. 2021) have all achieved varying degrees of success in the ACACR task. Nevertheless, these methods suffer from poor restoration performance and disconnection from real scenarios. Different from the above works, we introduce DiffACR, which, to our knowledge, is the first diffusion-based method for the ACACR task.

Diffusion Models

Diffusion models (Sohl-Dickstein et al. 2015; Ho, Jain, and Abbeel 2020; Song, Meng, and Ermon 2020) recently emerged with remarkable results on image generation (Dhariwal and Nichol 2021; Rombach et al. 2022), image-to-image translation (Saharia et al. 2022a), image editing (Hertz et al. 2022), *etc.* However, these methods rely on the properties of the Gaussian noise. The cold diffusion method

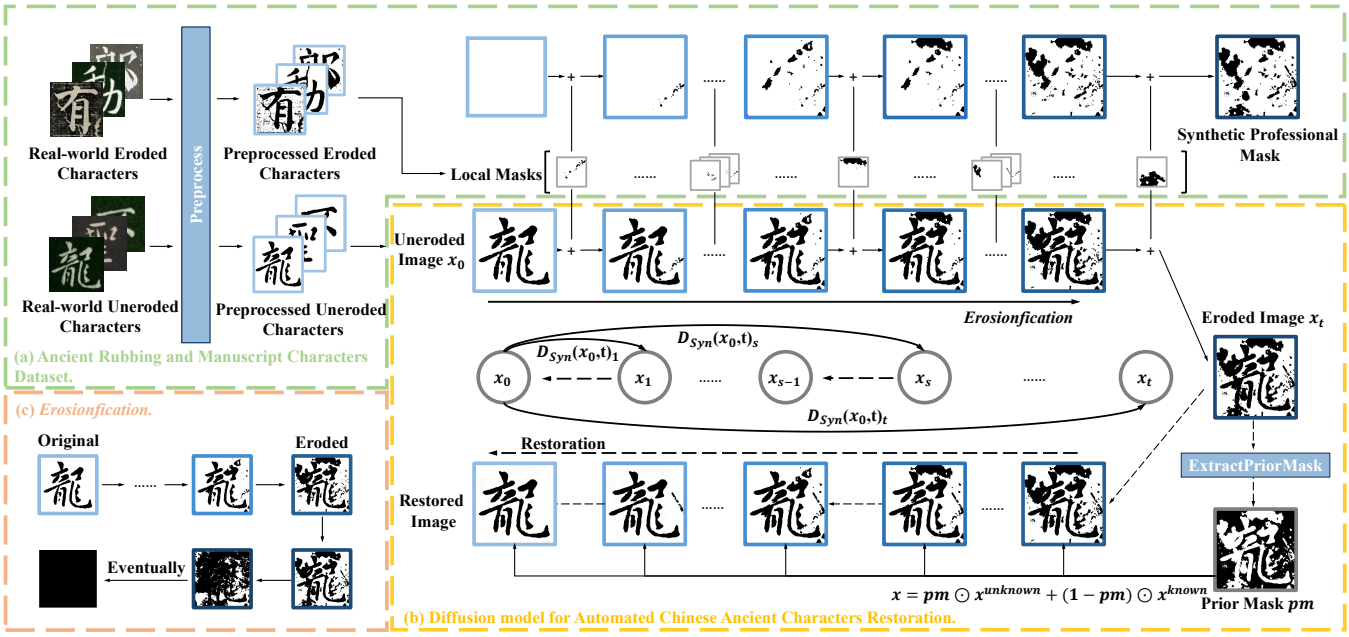


Figure 3: Overview of proposed ARMCD and DiffACR. (a) Chinese Ancient Rubbing and Manuscript Character Dataset (ARMCD) with professional mask synthesis. (b) Diffusion model for automated Chinese Ancient Character Restoration (DiffACR) with the forward *erosionfication* and the reverse restoration processes. (c) *Erosionfication*, where an uneroded image is iteratively degraded into an eroded image, eventually into a completely black image.

(Bansal et al. 2022), which has been applied to other fields (Yen et al. 2023; Gao et al. 2023), considers a broader family of degradation processes and constructs a generalized diffusion without previous theoretical limitations. The underlying properties of cold diffusion make it a promising framework for the ACACR task, where, in realistic conditions, the erosion is usually non-Gaussian. Moreover, Repaint (Lugmayr et al. 2022) introduces the mask region as a condition on unconditional denoising diffusion models. Inspired by these methods, we propose DiffACR, which follows the cold diffusion framework and uses the prior mask as a condition.

Chinese Ancient Rubbing and Manuscript Character Dataset (ARMCD)

In this section, we analyze the difficulties of collecting paired data and introduce our method of data collection, professional mask synthesis, data preprocessing, and statistics and analysis, as shown in Figure 3 (a).

Data Collection

Intuitively, one appealing approach to constructing a dataset for ACACR is to directly collect the pairs of eroded and uneroded (or restored) single-character images. However, this collection is arduous due to the following limitations:

- It is difficult to collect both uneroded and eroded versions of the same work. In ancient China, the characters were mostly recorded on steles and utensils, which have been eroded over time. The original uneroded copies were often damaged or lost, while only later records survived.
- The gold-standard restoration results of Chinese ancient characters are controversial, due to the different perspectives and standards of each individual.

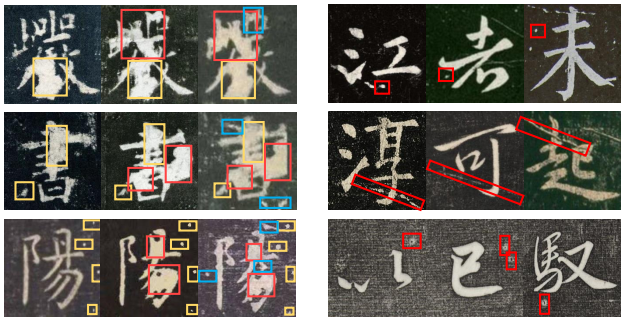
Therefore, we collect both eroded and uneroded single-character images of Chinese ancient characters from rubbings and manuscripts. We then attempt to extract (or synthesize) masks from the real-world eroded images to overlay on uneroded images, forming a paired dataset that closely resembles reality.

A total of 15093 uneroded and 460 eroded single-character images of Chinese ancient characters are collected from 42 authoritative, well-known, and representative Chinese ancient rubbings and manuscripts, covering the works of over 200 calligraphy artists spanning from 200 to $1,800$ AD. A detailed introduction to each of our sources is provided in Section Sources of the supplementary material.

Professional Mask Synthesis

Due to the limitations mentioned above, collecting uneroded images and providing masks is an alternative. However, there is an obvious gap between the existing image and character masks and real-world erosion. Erosion usually manifests as a random mixture of ink smears, noise, interfering patterns, and excavation damage (Shi et al. 2022b). Even the most similar irregular masks (Liu et al. 2018) and mixed Gaussian noise (Shi et al. 2022b) cannot fit correctly. Meanwhile, directly extracting a complete mask can result in inaccurate extraction and poor generalization. Therefore, we attempt to synthesize this special mask.

Guided by the expertise of restoration experts, we analyze



(a) Overlay of local erosion. (b) Similarity of local erosion.

Figure 4: Our key observations of real erosion. (a) The erosion of inscriptions from increasing historical periods is reflected in rubbings, which are respectively indicated by the yellow, red, and blue rectangles. Images from left to right are collected from *LiQi Rubbing Album*, *DuanFang Rubbing Album*, *LiHongYi Rubbing Album of Inscription of Sweet Spring in Jiucheng Palace*. (b) Areas of similar local erosion from different sources are highlighted by red rectangles.

massive real erosion from various perspectives and arrive at two key observations:

- Erosion is often a cumulative process of local erosion, which is consistent with natural laws. In nature, new erosion continues to overlay on the existing erosion rather than occurring all at once (see Figure 4 (a)).
- Erosion from different sources exhibits similar patterns of local erosion (see Figure 4 (b)).

We meticulously extract representative local erosion from collected eroded images as our local masks. These local masks are subjected to simple data augmentation and then overlaid. Specifically, given a set of local masks S_{LM} and a stacking times T_s, the professional mask M is synthesized in Algorithm 1 by:

$$M = Syn(S_{LM}, T_s, \theta_l, \theta_u, Sc_l, Sc_u, Ht, Vt), \quad (1)$$

where θ_l and θ_u denote the random rotation angle lower bound and upper bound, Sc_l and Sc_u denote the random scaling rate lower bound and upper bound, Ht and Vt denote the maximum absolute fraction for horizontal and vertical translations.

It is worth mentioning that this professional mask synthesis strategy is also adaptive for other ancient characters, such as ancient Greek and Egyptian characters, which provides a feasible means for constructing such restoration datasets. Moreover, this synthesis strategy supports the expansion of the existing local mask set to achieve better generalization.

Data Processing

To standardize all the single-character images collected from different sources, we preprocess them, according to Xing *et al.* (Xing and Ren 2023) with a few adjustments. Specifically, given a collected image I, the preprocessing can be described in Algorithm 2 as:

$$I_p = Preprocess(I, E_c, Th_l, Th_u, K_s), \quad (2)$$

Algorithm 1: Professional Masks Synthesis Algorithm

Input: The Set of Local Masks S_{LM}, Stacking Times T_s
Parameter: θ_l, θ_u, Sc_l, Sc_u, Ht, Vt
Output: professional Mask M

M start with an empty image
$t \leftarrow 1$
while $t \leq T_s$ **do**
  Randomly select $LM \in S_{LM}$
  $LM \leftarrow RandomRotation(LM, \theta_l, \theta_u)$
  $LM \leftarrow RandomScaling(LM, Sc_l, Sc_u)$
  $LM \leftarrow RandomHorizontalTrans(LM, Ht)$
  $LM \leftarrow RandomVerticalTrans(LM, Vt)$
  $M \leftarrow M \odot LM$
  $t \leftarrow t + 1$
end while
return M

Algorithm 2: Data Preprocessing

Input: Image $I(= l \times l)$
Parameter: E_c, Th_l, Th_u, K_s
Output: Preprocessed Image I_p

$I \leftarrow I \times E_c$
$I_{Th_l} \leftarrow \{I(i,j) > Th_l : i \in [0,l), j \in [0,l)\}$
$I_{Th_u} \leftarrow \{I(i,j) < Th_u : i \in [0,l), j \in [0,l)\}$
$I_p \leftarrow I_{Th_l} \bigcup I_{Th_u}$
$I_p \leftarrow MedianBlur(I_p, K_s)$
return I_p

where I_p is prepossed by the input I, with the contrast enhancement ratio E_c, dual-threshold binarization (Chen et al. 2008) threshold lower bound Th_l and upper bound Th_u, and the kernel size K_s of Median Blur.

Statistics and Analysis

The statistics of ARMCD are presented in Table 1. Specifically, we divide all 15,093 uneroded images into a training set, validation set, and test set consisting of 10,000, 2,000, and 2,000 images respectively. Additionally, we provide these datasets with an equal number of synthetic professional masks and ensure that the distribution of mask ratio is similar across each dataset. Note that we also provide all collected images, as these may be used for other purposes. To provide a comprehensive representation of our ARMCD, we further analyze the number of regular/non-regular script images in each dataset and the number of images collected from manuscripts/rubbings as shown in Table 1.

	Total	Regular/Non-R	Rubbing/Manuscript
Train Set	10000	8879/1121	3952/6048
Validation Set	2000	1782/218	853/1147
Test Set	2000	1767/233	760/1240

Table 1: The statistics of our ARMCD.

Diffusion Model for Automated Chinese Ancient Character Restoration (DiffACR)

In this section, we first introduce the original cold diffusion method (Bansal et al. 2022), and then we propose our key insights of our DiffACR: *erosionfication* operator and prior mask. Based on our insights, we optimize the training and sampling process, as shown in Figure 3 (b).

Preliminaries: Cold Diffusion

The original cold diffusion method (Bansal et al. 2022) generalizes classical diffusion models (Ho, Jain, and Abbeel 2020) and extends the diffusion and sampling processes to arbitrary image transforms or degradation, such as *snowification* (Hendrycks and Dietterich 2019). Specifically, given an image $x_0 \in \mathbb{R}^N$ and a severity t, a customized operator D is defined to gradually degrade the image x_0 into the degraded image $x_t = D(x_0, t)$. Note that the operator D should satisfy $D(x_0, 0) = x_0$ and vary continuously in the severity t. A learnable restoration operator R is also defined to (approximately) invert operator D, which can be expressed as $\widehat{x}_0 = R(x_t, t) \approx x_0$. In practice, only the restoration operator R is implemented via a neural network parameterized by θ, which is trained via the minimization problem:

$$\min_{\theta} \mathbb{E}_{x\sim\mathcal{X}} \parallel R_\theta(D(x,t),t) - x \parallel_1, \tag{3}$$

where x denotes a random image sampled from distribution $\mathcal{X}$ and $\| \cdot \|_1$ denotes the $\mathcal{L}_1$ norm.

When the restoration operator is perfect (*i.e.* when $R_\theta(x_t, t) = x_0$ for all t), the $R_\theta(\cdot)$ can directly generate the restored image $\widehat{x}_0$ from the degraded image x_t. However, it should be emphasized that with an imperfect restoration operator, such direct restoration may produce a blurred image $\widehat{x}_0$ instead. Moreover, an improved sampling algorithm is proposed for cold diffusion. Specifically, given a descending sequence $s = \{t, t-1, ..., 1\}$, the sampling algorithm to reduce the accumulated error is defined by:

$$x_{s-1} = x_s - D(\widehat{x}_0, s) + D(\widehat{x}_0, s-1), \tag{4}$$

where $\widehat{x}_0 = R_\theta(x_s, s)$. The improved sampling algorithm can produce exact reconstruction (*i.e.* $x_s = D(x_0, s)$) even when the restoration operator R fails to perfectly invert D. Although the original cold diffusion method has been shown to produce high-quality images with common degradation, it is still not suitable for the ACACR task due to the special degradation of Chinese ancient characters.

Erosionfication Operator

Inspired by the *snowification* (Hendrycks and Dietterich 2019) degradation, we propose to redefine our process of overlaying professional synthetic masks on uneroded images as a special degradation called *erosionfication*, where an uneroded image is iteratively degraded into an eroded image, eventually into a completely black image, as shown in Figure 3 (c). Specifically, given an uneroded image x_0 with the severity t, our *erosionfication* operator $D_{Syn(S_{LM},T_s,\theta_l,\theta_u,Sc_l,Sc_u,Ht,Vt)}$ can be described in Algorithm 3 as:

$$\{x_1, ..., x_t\} = D_{Syn(S_{LM},T_s,\theta_l,\theta_u,Sc_l,Sc_u,Ht,Vt)}(x_0, t), \tag{5}$$

Algorithm 3: *Erosionfication* Algorithm

Input: Uneroded Image x_0, Severity t
Function: $Syn(S_{LM}, T_s, \theta_l, \theta_u, Sc_l, Sc_u, Ht, Vt)$
Output: List of Eroded Images with Increasing Severity $\{x_1, ..., x_t\}$

$i \leftarrow 1$
while $i \leq t$ **do**
  $LM \leftarrow Syn(S_{LM}, T_s, \theta_l, \theta_u, Sc_l, Sc_u, Ht, Vt)$
  $x_i \leftarrow x_{i-1} \odot LM$
  $i \leftarrow i + 1$
end while
return $\{x_1, ..., x_t\}$

where $\{x_1, ..., x_t\}$ is a list of eroded images with increasing severity ($t' \leq t$) synthesized by the function $Syn(\cdot)$. We have so far used the subscript $D_{Syn(S_{LM},T_s,\theta_l,\theta_u,Sc_l,Sc_u,Ht,Vt)}$ to emphasize the dependence of the function $Syn(\cdot)$ on the hyperparameters and inputs, but we will omit this symbol for simplicity, as $D_{Syn}(\cdot)$ in the discussion below.

Prior Mask

We propose using the prior mask as a condition to highlight the unrestored region and provide prior guidance. Note that the prior mask differs from the professional synthetic mask in Subsection Professional Mask Synthesis because the ACACR task is essentially an end-to-end task that requires restoring an eroded image to its uneroded state without utilizing any additional information. We have to directly extract the prior mask from the eroded images.

Intuitively, the set of white pixels in the eroded image is a natural constant that should remain unchanged throughout the whole process. Instead, the set of non-white pixels can be regarded as unrestored regions, which may be changed during restoration. Moreover, we cannot further narrow down the unrestored regions due to the high fusion of erosion and strokes on the eroded images. Therefore, we choose the set of white pixels as our prior mask which can be directly extracted from the eroded images. Specifically, given an eroded image x, the prior mask pm is extracted by:

$$pm = ExactPriorMask(x), \tag{6}$$

where $ExactPriorMask(\cdot)$ is a simple function that outputs the inverted image after exacting all white pixels.

Optimizing Training and Sampling Process

Training with Prior Mask. By using the prior mask region as a condition, our goal is redefined to predict the unknown (or unrestored) pixels of an eroded image. Specifically, we propose reducing the distance on the unknown region instead of the whole image during training. Following (Bansal et al. 2022), given the operator D_{Syn} and R_θ, our training objective is defined in Algorithm 4 by:

$$\min_{\theta} \mathbb{E}_{x_0\sim\mathcal{X},x_{t'},pm\sim D_{Syn}(x,t)} \parallel R_\theta(x_{t'}, t') \odot pm - x_0 \odot pm \parallel_1, \tag{7}$$

Algorithm 4: Training a Restoration Network R_θ

Input: Training Distribution $\mathcal{X}$, Severity t
Function: $ExactPriorMask(\cdot)$
Operator: D_{Syn}, R_θ

repeat
  $x_0 \sim \mathcal{X}$
  $x_1, ..., x_t \leftarrow D_{Syn}(x_0, t)$
  $pm = ExactPriorMask(x_t)$
  $t' \sim Uniform(\{1, ..., t\})$
  $\widehat{x} \leftarrow R_\theta(x_{t'}, t')$
  Take a gradient descent step on $\nabla_\theta \| \widehat{x} \odot pm - x_0 \odot pm \|_1$
until converged

where $pm = ExactPriorMask(x_t)$, $x_{t'}$ and x_t is the t'th and the last image of the output list of $D_{Syn}(x_0, t)$, with the uneroded image x_0 randomly sampled from the training distribution $\mathcal{X}$ and a random severity t' selected from $Uniform(\{1, ..., t\})$.

Sampling for Restoration. Considering the prior mask pm, we denote the ground truth restored image x, consisting of the unknown pixels as $pm \odot x$ and the known pixels as $(1-pm)\odot x$. Since every restoration step from x_s to x_{s-1} depends solely on x_s, we can alter the unknown regions as long as we keep the correct prediction of the corresponding regions. Inspired by (Lugmayr et al. 2022), we sample the known regions from our forward operator D_{Syn}. Thus, using x_{s-1}^{known} and $x_{s-1}^{unknown}$ for the known and unknown regions for the time step s, the x_{s-1} is defined by:

$$x_{s-1}^{known} = D_{Syn}(\widehat{x}_0, t)_{s-1}, \tag{8a}$$

$$x_{s-1}^{unknown} = x_s - D_{Syn}(\widehat{x}_0, t)_s + D_{Syn}(\widehat{x}_0, t)_{s-1}, \tag{8b}$$

$$x_{s-1} = pm \odot x_{s-1}^{unknown} + (1 - pm) \odot x_{s-1}^{known}, \tag{8c}$$

where $\widehat{x}_0 = R_\theta(x_s, s)$ and $D_{Syn}(\cdot)_i$ denotes the ith image of the output list of $D_{Syn}(\cdot)$. Moreover, due to the natural constant of the prior mask mentioned in Subsection Prior Mask, the known region $1 - pm$ of the image $x_k^{known} = D_{Syn}(\widehat{x}_0, t)_k$ constantly represents a white region. Therefore, given a fixed pm, the calculation of x_{s-1} simplifies in Algorithm 5 to:

$$\begin{aligned} x_{s-1} &= pm \odot x_{s-1}^{unknown} + (1 - pm) \odot x_{s-1}^{known} \\ &= pm \odot x_{s-1}^{unknown} + 1 - pm. \end{aligned} \tag{9}$$

Architecture. Our DiffACR uses a U-Net architecture (Ho, Jain, and Abbeel 2020) as the operator R_θ. The network architecture is based on the 224×224 ConvNeXt (Liu et al. 2022) with time embeddings.

Experiments

In this section, we first describe the experimental settings. Then we perform extensive experiments, compare to the state-of-the-art solutions, and conduct an ablation study. A user study (2AFC) is also introduced to assess the professionality of our synthetic masks.

Algorithm 5: Improved Sampling for the ACACR Task

Input: Eroded Image x_t, Severity t
Function: $ExactPriorMask(\cdot)$
Operator: D_{Syn}, R_θ

$pm = ExactPriorMask(x_t)$
$s \leftarrow t$
while $s \geq 0$ **do**
  $\widehat{x}_0 \leftarrow R_\theta(x_s, s)$
  $\widehat{x}_1, ..., \widehat{x}_t \leftarrow D_{Syn}(\widehat{x}_0, t)$
  $x_{s-1}^{unknown} \leftarrow x_s - \widehat{x}_s + \widehat{x}_{s-1}$
  $x_{s-1} \leftarrow pm \odot x_{s-1}^{unknown} + 1 - pm$
  $s \leftarrow s - 1$
end while

Experimental Settings

Implementation Details. We evaluate all the experiments on our proposed ARMCD. All images are resized and cropped to 224×224, and all the experiments are conducted with one NVIDIA Tesla V100 SXM2 32GB GPU.

Evaluation Metrics. The mean absolute error (MAE), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) index are used to quantitatively evaluate the restoration performance. We additionally introduce Fréchet Inception Distance (FID) and perceptual metric LPIPS (Zhang et al. 2018), which is a learned distance metric based on the deep feature space of AlexNet.

Performance Comparison

We compare our DiffACR with nine state-of-the-art open-source inpainting methods on our ARMCD, including DNCNN (Zhang et al. 2017), Cycle-Dehaze (Engin, Genç, and Kemal Ekenel 2018), VDN (Guo et al. 2019), CIDG (Zhang, Guo, and Fan 2020), SCCGAN (Liu et al. 2021), SGGAN (Li et al. 2021), SwinIR (Liang et al. 2021), IPT (Chen et al. 2021), and CharFormer (Shi et al. 2022a). Note that we modify the code of these methods provided by the researchers and make them trainable on ARMCD. As shown in Table 2, our DiffACR method achieves the best and second-best performance on most metrics. Moreover, we additionally synthesize an equal number of masks (not included in ARMCD) with varying mask rates by adjusting the stacking times T_s. We then test our DiffACR under different mask ratios as shown in Table 3. Interestingly, we achieve satisfactory results under 21%-40% and 41%-60% mask ratios, which matches the distribution of our masks in ARMCD.

Ablation Study

We compare our *erosionfication* operator with other degradation (*e.g.* the Gaussian noise and *snowification*) and examine the effectiveness of the prior mask by removing its guidance in different stages. Table 4 shows the ablation study results. The results demonstrate that using *erosionfication* operator and the prior mask region enhances performance.

Method	MAE ↓	PSNR ↑	SSIM ↑	FID ↓	LPIPS ↓
DNCNN (Zhang et al. 2017)	0.0873	21.04	0.9065	75.12	0.3925
Cycle-Dehaze (Engin, Genç, and Kemal Ekenel 2018)	0.1025	16.97	0.8862	92.19	0.4215
VDN (Guo et al. 2019)	0.0619	21.46	0.9457	64.65	0.3078
CIDG (Zhang, Guo, and Fan 2020)	0.0567	21.88	0.9271	49.96	0.2623
SCCGAN (Liu et al. 2021)	0.0324	17.72	0.8976	36.59	0.1914
SGGAN (Li et al. 2021)	0.0308	19.92	0.9673	33.24	0.0842
IPT (Chen et al. 2021)	**0.0169**	23.73	0.9727	22.68	0.0777
SwinIR (Liang et al. 2021)	0.0195	<u>24.08</u>	<u>0.9983</u>	18.53	**0.0483**
CharFormer (Shi et al. 2022a)	0.0226	**24.38**	0.9886	<u>15.44</u>	0.0557
DiffACR(Ours)	<u>0.0187</u>	22.25	**0.9988**	**12.87**	<u>0.0494</u>

Table 2: Performance comparison with state-of-the-art methods and our DiffACR on the ARMCD. The best and second-best results are highlighted and underlined.

Mask Ratio	MAE ↓	PSNR ↑	SSIM ↑	LPIPS ↓
0-20%	0.0234	20.08	0.9780	0.0738
21%-40%	0.0169	22.22	0.9984	0.0509
41%-60%	0.0151	24.16	0.9990	0.0371
61%-80%	0.0524	16.35	0.9513	0.1533
81%+	0.2165	8.35	0.9109	0.4546

Table 3: Performance under different mask ratios.

Method	MAE ↓	SSIM ↑	LPIPS ↓
Gaussian noise	0.0563	0.9677	0.2127
Snowification	0.0356	0.9809	0.0832
w/o prior mask in all stages	0.0287	0.9577	0.0925
w/o prior mask in training stage	0.0259	0.9632	0.0782
w/o prior mask in sampling stage	0.0236	0.9643	0.0665
Ours(*Erosionfication*)	**0.0187**	**0.9988**	**0.0494**

Table 4: Ablation study results on the ARMCD. The upper part compares our *erosionfication* with other degradation, while the lower part compares the effects of removing the prior mask guidance at different stages.

	Deception Rate	Recognition Rate
Graduate Students	55%	45%
Restoration Experts	63%	37%

Table 5: User study results of 2AFC test. The deception rate represents the proportion of synthesized images that are regarded as real erosion among the combinations by participants, while the recognition rate is the opposite.

User Study

We introduce the forced-choice method of selecting the better of two options (2AFC) to assess the professionality of our synthetic masks mentioned in Subsection Professional Mask Synthesis, which has been widely applied in psychophysics, neuroscience, and cognitive science (Saharia et al. 2022b). For each participant, we randomly selected 60 synthetic masks overlaying on the uneroded images and 60 real eroded images from our collection. All images were combined and then randomized. The participants were presented with these combinations one by one and asked, "Which one do you think is the real damaged image?" They were allowed to observe the images for 3 seconds (Zhang et al. 2018) before making a choice within 5 seconds after observation. Finally, the proportion of synthesized images being mistaken for real eroded images was calculated. This proportion represents the deception rate of participants, which indirectly reflects the professionality of our synthetic masks. We invited a total of 37 manual restoration experts and 200 graduate students in related majors to participate in the test and aggregated the results from all the participants. The results in Table 5 reveal that, on average, 63% of the eroded images synthesized by our masks were recognized by restoration experts, while 55% by graduate students.

Conclusion

In this paper, we introduce the ARMCD for the ACACR task, which is, to our knowledge, the largest publicly available ACACR dataset, surpassing existing datasets in both the diversity and quantity of ancient Chinese real-world artifacts. We also provide synthetic masks, which have been verified for professionality through our user study. We further introduce DiffACR, which introduces *erosionfication* as a form of cold diffusion and uses the prior mask to guide the restoration process. To our knowledge, DiffACR is also the first diffusion-based method for the ACACR task. We believe the proposed ARMCD and DiffACR can serve as a foundation for future ACACR works and stimulate research with a broader perspective.

References

Bansal, A.; Borgnia, E.; Chu, H.-M.; Li, J. S.; Kazemi, H.; Huang, F.; Goldblum, M.; Geiping, J.; and Goldstein, T. 2022. Cold diffusion: Inverting arbitrary image transforms without noise. *arXiv preprint arXiv:2208.09392*.

Chen, H.; Wang, Y.; Guo, T.; Xu, C.; Deng, Y.; Liu, Z.; Ma, S.; Xu, C.; Xu, C.; and Gao, W. 2021. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12299–12310.

Chen, Q.; Sun, Q.-s.; Heng, P. A.; and Xia, D.-s. 2008. A double-threshold image binarization method based on edge detector. *Pattern recognition*, 41(4): 1254–1267.

Dhariwal, P.; and Nichol, A. 2021. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34: 8780–8794.

Diao, X.; Shi, D.; Li, J.; Shi, L.; Yue, M.; Qi, R.; Li, C.; and Xu, H. 2023. Toward Zero-shot Character Recognition: A Gold Standard Dataset with Radical-level Annotations. *arXiv preprint arXiv:2308.00655*.

Engin, D.; Genç, A.; and Kemal Ekenel, H. 2018. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 825–833.

Gao, Q.; Li, Z.; Zhang, J.; Zhang, Y.; and Shan, H. 2023. CoreDiff: Contextual Error-Modulated Generalized Diffusion Model for Low-Dose CT Denoising and Generalization. *arXiv preprint arXiv:2304.01814*.

Guo, S.; Yan, Z.; Zhang, K.; Zuo, W.; and Zhang, L. 2019. Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1712–1722.

Hendrycks, D.; and Dietterich, T. 2019. Benchmarking neural network robustness to common corruptions and perturbations. *arXiv preprint arXiv:1903.12261*.

Hertz, A.; Mokady, R.; Tenenbaum, J.; Aberman, K.; Pritch, Y.; and Cohen-Or, D. 2022. Prompt-to-prompt image editing with cross attention control. *arXiv preprint arXiv:2208.01626*.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Jin, L.; Gao, Y.; Liu, G.; Li, Y.; and Ding, K. 2011. SCUT-COUCH2009—a comprehensive online unconstrained Chinese handwriting database and benchmark evaluation. *International Journal on Document Analysis and Recognition (IJDAR)*, 14: 53–64.

Li, H.; Zhong, Z.; Guan, W.; Du, C.; Yang, Y.; Wei, Y.; and Ye, C. 2021. Generative character inpainting guided by structural information. *The Visual Computer*, 37: 2895–2906.

Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; and Timofte, R. 2021. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1833–1844.

Liu, C.-L.; Yin, F.; Wang, D.-H.; and Wang, Q.-F. 2011. CASIA online and offline Chinese handwriting databases. In *2011 international conference on document analysis and recognition*, 37–41. IEEE.

Liu, G.; Reda, F. A.; Shih, K. J.; Wang, T.-C.; Tao, A.; and Catanzaro, B. 2018. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European conference on computer vision (ECCV)*, 85–100.

Liu, R.; Wang, X.; Lu, H.; Wu, Z.; Fan, Q.; Li, S.; and Jin, X. 2021. SCCGAN: style and characters inpainting based on CGAN. *Mobile networks and applications*, 26: 3–12.

Liu, Z.; Mao, H.; Wu, C.-Y.; Feichtenhofer, C.; Darrell, T.; and Xie, S. 2022. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11976–11986.

Lugmayr, A.; Danelljan, M.; Romero, A.; Yu, F.; Timofte, R.; and Van Gool, L. 2022. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11461–11471.

Lv, D.; and Liu, Y. 2018. The Restoration of Style Chinese Characters Based on Deep Learning. In *2018 International Conference on Network, Communication, Computer Engineering (NCCE 2018)*, 426–430. Atlantis Press.

QIu, Y. 2023. Appreciation of ouyang xun's" ninety percent palace feels ashamed springs inscription retraces reprint instructions" inscription. *Beauty and Times (Medium)(03)*, 132–134.

Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10684–10695.

Saharia, C.; Chan, W.; Chang, H.; Lee, C.; Ho, J.; Salimans, T.; Fleet, D.; and Norouzi, M. 2022a. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, 1–10.

Saharia, C.; Ho, J.; Chan, W.; Salimans, T.; Fleet, D. J.; and Norouzi, M. 2022b. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4713–4726.

Shi, D.; Diao, X.; Shi, L.; Tang, H.; Chi, Y.; Li, C.; and Xu, H. 2022a. CharFormer: A glyph fusion based attentive framework for high-precision character image denoising. In *Proceedings of the 30th ACM International Conference on Multimedia*, 1147–1155.

Shi, D.; Diao, X.; Tang, H.; Li, X.; Xing, H.; and Xu, H. 2022b. RCRN: Real-world Character Image Restoration Network via Skeleton Extraction. In *Proceedings of the 30th ACM International Conference on Multimedia*, 1177–1185.

Sohl-Dickstein, J.; Weiss, E.; Maheswaranathan, N.; and Ganguli, S. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, 2256–2265. PMLR.

Song, G.; Li, J.; and Wang, Z. 2020. Occluded offline handwritten Chinese character inpainting via generative adversarial network and self-attention mechanism. *Neurocomputing*, 415: 146–156.

Song, J.; Meng, C.; and Ermon, S. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*.

Su, B.; Liu, X.; Gao, W.; Yang, Y.; and Chen, S. 2022. A restoration method using dual generate adversarial networks for Chinese ancient characters. *Visual Informatics*, 6(1): 26–34.

Tétreault, J. 2013. Agent of deterioration: Pollutants. *Agents of Deterioration*.

Wang, D.-H.; Liu, C.-L.; Yu, J.-L.; and Zhou, X.-D. 2009. CASIA-OLHWDB1: A database of online handwritten Chinese characters. In *2009 10th International Conference on Document Analysis and Recognition*, 1206–1210. IEEE.

Wang, J.; Pan, G.; Sun, D.; and Zhang, J. 2021. Chinese Character Inpainting with Contextual Semantic Constraints. In *Proceedings of the 29th ACM International Conference on Multimedia*, 1829–1837.

Wang, X.; Liang, X.; Hu, J.; and Sun, L. 2012. Stroke-based Chinese character completion. In *2012 Eighth International Conference on Signal Image Technology and Internet Based Systems*, 281–288. IEEE.

Xing, C.; and Ren, Z. 2023. Binary Inscription Character Inpainting Based on Improved Context Encoders. *IEEE Access*.

Xu, S.; Jin, T.; Jiang, H.; and Lau, F. C. 2009. Automatic generation of personal chinese handwriting by capturing the characteristics of personal handwriting. In *Twenty-First IAAI Conference*.

Xu, Y.; Yin, F.; Wang, D.-H.; Zhang, X.-Y.; Zhang, Z.; and Liu, C.-L. 2019. CASIA-AHCDB: A large-scale Chinese ancient handwritten characters database. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 793–798. IEEE.

Yen, H.; Germain, F. G.; Wichern, G.; and Le Roux, J. 2023. Cold diffusion for speech enhancement. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Zhang, H.; Guo, J.; Chen, G.; and Li, C. 2009. HCL2000-A large-scale handwritten Chinese character database for handwritten character recognition. In *2009 10th international conference on document analysis and recognition*, 286–290. IEEE.

Zhang, J.; Guo, M.; and Fan, J. 2020. A novel generative adversarial net for calligraphic tablet images denoising. *Multimedia Tools and Applications*, 79: 119–140.

Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; and Zhang, L. 2017. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7): 3142–3155.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–595.

Zhong, Z.; Yin, F.; Zhang, X.-Y.; and Liu, C.-L. 2017. Handwritten Chinese character blind inpainting with conditional generative Adversarial Nets. In *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, 804–809. IEEE.