# Two-Stage Stain Reduction in Ancient Japanese Manuscripts via Mask-Guided Diffusion Models

**YUYA YOSHIZU[1], HAYATA KANEKO[1], RYUTO ISHIBASHI[1], LIN MENG[2] AND MINGCONG DENG[3]**

[1]Graduate School of Science and Engineering, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan
[2]Department of Electrical and Electronic Engineering, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan
[3]Tokyo University of Agriculture and Technology, 2-24-16 Nakacho, Koganei, Tokyo, Japan 184-8588

Corresponding author: L. Meng(e-mail: menglin@fc.ritsumei.ac.jp)

This paper is an extended version of our work presented at ICCSI 2024 [1].
Yuya Yoshizu and Hayata Kaneko contributed equally to this work.
This work was supported by JSPS KAKENHI Grant Number KP24KJ2152.

## ABSTRACT

Ancient Japanese manuscripts are invaluable cultural heritage, yet their readability has been severely degraded by stains, bleed-through, and aging. Conventional binarization produces unstable results under diverse degradations, and CNN-based approaches such as CycleGAN require task-specific training data, making dataset preparation costly. To address these limitations, we propose a two-stage restoration framework based on zero-shot Denoising Diffusion Restoration Models (DDRM). In the first stage, binarization with noise masking removes stains and isolates text regions, while mask-guided DDRM restores fine stroke details lost during thresholding. In the second stage, patch-wise restorations are reassembled and refined at the page level to ensure tonal and structural consistency. Experiments on *Tsurezuregusa* (322 pages) and *Isoho Monogatari* (198 pages) demonstrate that our method surpasses binarization and CycleGAN, achieving SSIM 0.715/0.776, with further improvement after gamma correction ($\gamma$=0.8). Analysis of brightness variance, kurtosis, and entropy confirms enhanced contrast, sharpened strokes, and reduced stain noise, demonstrating the effectiveness of the proposed approach for restoring manuscripts of cultural heritage.

**INDEX TERMS** Ancient manuscripts, Binarization, Cultural Heritage Preservation, Diffusion Model, Image Restoration

## I. INTRODUCTION

Ancient Japanese manuscripts are invaluable cultural heritage that have been preserved for centuries. They provide insights into the culture, thought and history of their eras, as well as information about the politics and economy of the time. Deciphering such manuscripts helps us understand the mindset and sentiments of people in the past. However, many of these manuscripts suffer from damage and stains caused by long-term deterioration, which greatly reduce their readability.

Fig.1 shows examples of characters and degradation types found in ancient Japanese manuscripts. Fig. 1-(1) presents character types: Fig.1-(1-b) illustrates bleed-through, where text from adjacent pages becomes visible on the current page. In Fig.1-(1-d), the main text overlaps with bleed-through characters, and stronger bleed-through severely impairs readability. This phenomenon also causes false edge detection in OCR-based recognition. Fig.1-(1-c) shows red characters, often seals indicating authentication or ownership, which may also serve structural, religious, or decorative purposes.

Fig.1-(2) (4) illustrate types of physical damage. Fig.1-(2) shows stains ranging from mild cases that do not overlap text Fig.1-(2-a) to severe ones that obscure characters Fig.1-(2-b), typically caused by paper degradation and long-term storage. Fig.1-(3) shows insect damage, typically caused by woodboring beetles, silverfish, and cockroaches. The beetles bore circular holes Fig.1-(3-a), while silverfish graze the surface irregularly, leaving Y-shaped defects Fig.1-(3-b). Fig.1-(4) shows character loss due to insect damage Fig.1-(4-a) or poor preservation Fig.1-(4-b). When multiple types of degradation coexist, restoration becomes technically challenging and costly.

Traditionally, such deterioration has been repaired man-

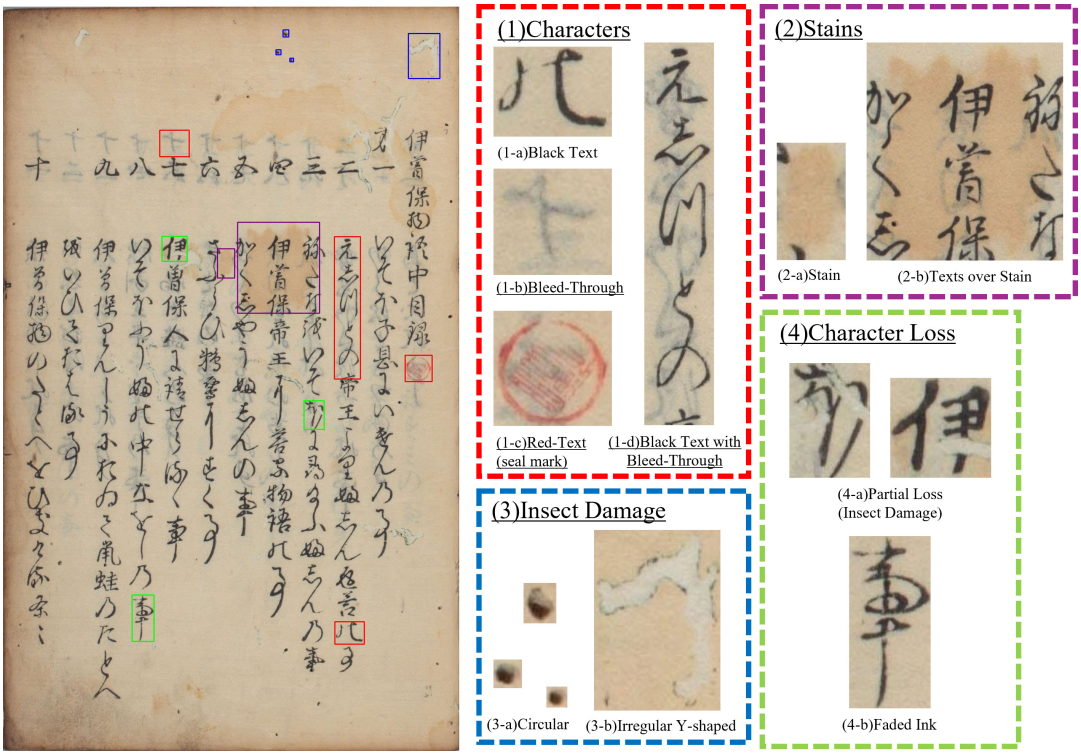



FIGURE 1: Characters and degradation types in an ancient Japanese manuscript. The left side shows the original page with bounding boxes indicating each category: 1)Characters (red), 2)Stains (purple), 3)Insect Damage (blue), and 4)Character Loss (green). The right side displays enlarged cropped images for each category.

ually. However, with recent advances in image processing, many institutions and researchers have begun restoring manuscripts using deep learning techniques [2], [5]–[11]. For example, the Library of Congress digitizes original manuscripts to promote collaboration between computer science and the humanities, while the Art Research Center (ARC) at Ritsumeikan University develops digitization technologies and publishes cultural heritage collections from various institutions [12]. Recent studies have applied deep learning to improve recognition accuracy for ancient scripts and complex typefaces [13]–[15].

Furthermore, the emergence of image generation models has greatly expanded the potential of digital restoration. Recent studies have attempted to restore missing characters using Generative Adversarial Networks (GANs) [16] and Diffusion Models [17]. However, research on AI-based stain removal remains limited, primarily due to the challenges of data collection and the diverse range of degradation types. Ancient manuscripts exhibit heterogeneous forms of contamination, such as yellowing, sunburn, bleed-through, mold, and ink diffusion, each with distinct causes and visual characteristics. This diversity complicates the optimization of deep learning models, making it challenging for a single network to generalize across different degradation patterns. Moreover, evaluation poses an additional challenge, as the absence of a perfectly clean reference (ground truth) makes it difficult to establish objective metrics for quantifying stain reduction.

To address these issues, we employ Denoising Diffusion Restoration Models (DDRM) [18], which perform zero-shot restoration without requiring paired data. DDRM leverages the generative prior of a pre-trained diffusion model that captures the statistical regularities and structural characteristics of natural images. This prior functions as a learned distribution of clean, plausible image structures, allowing the model to infer and reconstruct degraded regions even when specific stain types are not represented in the training dataset. As a result, the proposed two-stage framework mitigates data scarcity and achieves robust restoration performance across diverse degradation patterns.

For evaluation, we employ not only PSNR (Peak Signal-to-Noise Ratio) [19] and SSIM (Structural Similarity Index Measure) [20], but also kurtosis, entropy, and RMSE (Root Mean Square Error) as complementary metrics. While PSNR and SSIM are standard indicators of structural fidelity, the additional use of kurtosis, entropy, and RMSE enables a more comprehensive assessment of visual quality, statistical texture characteristics, and intensity consistency, providing a holistic evaluation of the proposed method.

First, binarization is applied to separate the background regions, followed by DDRM-based reconstruction of character details that were lost during thresholding. The framework consists of two stages: a patch stage, where each page is divided into overlapping patches for localized stain removal,

and a page stage, where restored patches are reassembled and the background is globally corrected.

To the best of our knowledge, this is the first study to apply diffusion models for stain removal in ancient Japanese manuscripts. The proposed method not only produces visually improved results but also achieves significant improvements in evaluation metrics such as kurtosis, brightness variance, and entropy, demonstrating closer similarity to the ground truth than binarization.

The remainder of this paper is organized as follows. Section II reviews prior work on binarization and learning-based restoration for historical documents. Section III describes our two-stage framework—phase 1 (patch-wise DDRM guided by binarization-based noise masking) and phase 2 (page-wise DDRM for global background consistency)—along with optional post-processing via $\gamma$-correction. Section IV details datasets (*Tsurezuregusa*, *Isoho Monogatari*), implementation settings, and quantitative results (PSNR, SSIM, kurtosis, brightness variance, entropy). Section V analyzes strengths, limitations (e.g., patch-boundary seams, occasional loss of red annotations), and the trade-offs introduced by $\gamma$-correction. Finally, Section VI summarizes the contributions and outlines directions for future work, including improved tonal control, patch blending, and expert-guided evaluation for cultural-heritage restoration.

In addition, the digital preservation of ancient manuscripts offers numerous advantages for managing cultural heritage. First, it reduces the need to handle fragile materials such as paper, cloth, and woodblock prints directly, thereby minimizing physical deterioration and allowing long-term preservation. Second, storing duplicates and backups in multiple locations mitigates the risk of loss due to disasters or accidents. Third, it allows rare and valuable materials to be accessed online by researchers and the general public worldwide, thus contributing to academic research and education without the constraints of time and place. Fourth, integrating metadata and optical character recognition (OCR) facilitates full-text search and cross-referencing between materials, while also enabling advanced analysis using AI and image processing techniques. Fifth, multilingual access and online exhibitions promote cultural transmission and international dissemination, forming a foundation for global collaboration and cultural exchange. Furthermore, providing a virtual restoration and damage-repair environment enables the evaluation of restoration processes and supports education through direct comparison between the original and restored data. This underscores the broader significance of our research and its potential contributions to the preservation and dissemination of cultural heritage.

## II. RELATED WORKS

### A. BINARIZATION TECHNIQUES FOR ANCIENT DOCUMENT PRESERVATION

Binarization is a fundamental preprocessing step in the digital preservation and restoration of ancient documents.

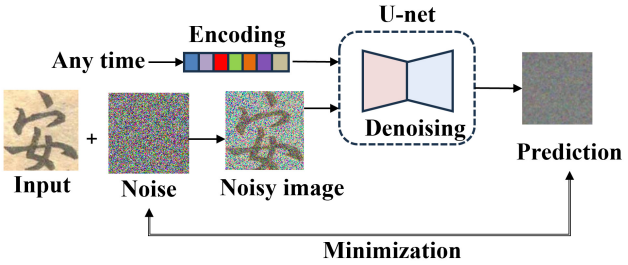



FIGURE 2: DDPM Architecture.

This process transforms grayscale or color images into binary images, separating text regions (foreground) from non-text areas (background). Such separation is indispensable for degraded manuscripts, where multiple deterioration factors—such as stains, bleed-through, ink fading, and uneven illumination—severely impair character legibility and reduce the accuracy of subsequent Optical Character Recognition (OCR) and restoration procedures.

Historically, binarization techniques relied primarily on simple global thresholding; however, advances in image processing have yielded more sophisticated approaches, including adaptive binarization and error diffusion methods. The emergence of image processing libraries such as OpenCV has further facilitated the implementation and dissemination of these techniques among researchers and practitioners.

Numerous binarization methods have been proposed to address the unique challenges of ancient and historical documents. Typical examples are as follows:

- **Effective and Fast Binarization for Combined Degradation on Ancient Documents (2019):** A four-stage framework incorporating histogram analysis, contrast enhancement, local adaptive thresholding, and artifact removal, demonstrating robustness against various forms of combined degradation [22].
- **A New Binarization Algorithm for Historical Documents:** An Otsu-based method with improved capability to separate text from complex backgrounds, yielding superior classification performance for historical manuscripts [23].
- **A Comparison of Binarization Methods for Historical Archive Documents:** A comparative analysis indicating that Niblack's method can outperform others for certain archival materials, owing to its adaptability to local illumination and contrast variations [24].
- **Restoration of Ancient Document Images Using Phase-Based Binarization:** An approach exploiting phase information to mitigate background interference in degraded manuscripts, thereby enabling clearer extraction of character contours [25].
- **Degraded Historical Document Binarization: A Review (2019):** A survey summarizing key challenges in binarizing degraded historical texts and outlining global, local, and hybrid thresholding strategies [26].

Recent advances in deep learning have further improved binarization performance on severely degraded historical documents. Ju et al. [27] proposed a three-stage GAN-based binarization framework that combines wavelet transforms, patch-level processing, and channel decomposition. Their model demonstrated strong generalization across complex degradation types such as bleed-through and yellowing, outperforming prior DIBCO benchmarks. Similarly, Zhang et al. [28] introduced DocRes, a unified restoration model capable of performing binarization alongside other restoration tasks such as dewarping and deblurring, guided by dynamic visual prompts.

Although conventional binarization methods offer high computational efficiency and can perform well under specific conditions, they encounter limitations when applied to severe or complex degradation patterns. In the context of ancient Japanese manuscripts, the preservation of fine and decorative strokes—such as those in *kuzushiji*—is particularly challenging. Furthermore, the diverse composition of inks, pigments, and paper substrates produces varying responses to binarization, complicating the application of a single universal method. To overcome these challenges, recent research has investigated advanced approaches such as color-space segmentation, unsupervised clustering, and deep learning-based binarization, all of which have demonstrated improvements in restoration fidelity and adaptability to heterogeneous degradation patterns.

### B. RESTORATION USING GAN-BASED MODELS

Virtual restoration of digitized cultural heritage, such as ancient manuscripts and murals, requires interpolation and reconstruction of information in damaged or missing areas [29] [30]. In this domain, Generative Adversarial Networks (GANs) have been widely adopted as a promising solution due to their powerful generative capabilities.

A GAN consists of a generator and a discriminator, trained in an adversarial manner, where the generator produces realistic outputs and the discriminator distinguishes between generated and real images [16]. In cultural heritage restoration, the objective is to predict the post-restoration state from the pre-restoration input, supplementing missing regions while preserving the original features, style, and cultural characteristics of the artifact. Restoration tasks in this context are particularly challenging because they require not only visual plausibility but also historical fidelity.

Various GAN architectures have been explored for image restoration tasks. **Pix2Pix** [31] performs supervised image-to-image translation and therefore requires paired degraded–clean data, which is often impractical for historical manuscripts. **CycleGAN** [32] relaxes this requirement by learning bidirectional mappings between domains, enabling restoration without paired samples and making it appealing for heritage documents where aligned pre- and post-degradation images do not exist. However, it may introduce artifacts due to cycle-consistency constraints and mode collapse in highly textured regions. **StyleGAN** [33] can synthesize high-fidelity textures by learning a disentangled latent space, and has been applied to tasks such as portrait and artwork restoration. Its powerful generative capacity, however, increases the risk of hallucinating visually plausible but historically inaccurate content. **SPADE** [34] incorporates spatially-adaptive normalization to preserve spatial layouts and semantic structures, enabling localized restoration while maintaining global coherence. Nonetheless, it typically requires semantic masks or labels, which are costly to prepare for ancient manuscripts with complex degradation patterns.

While GAN-based frameworks can produce visually compelling results, their reliance on training data and the potential for content hallucination highlight the need for methods that restore authentic information without extensive dataset preparation—motivating the adoption of diffusion-based approaches such as DDRM in this study.

In the field of cultural heritage restoration, Kaneko et al. [35] proposed a method for reconstructing missing characters by incorporating domain discriminator loss into the adversarial loss, enabling the model to learn differences between degraded and intact text regions. Zheng et al. [2] presented a restoration approach that takes two inputs—the image of the missing character and a reference character image—and uses the reference to guide the restoration process toward consistent results.

However, applying GANs to the restoration of cultural heritage presents several domain-specific challenges. Data scarcity is a major issue, as high-quality, high-resolution restoration datasets are limited, and historical artifacts are unique, restricting the diversity of training data. In Japanese ancient manuscripts, hiragana, katakana, and kanji characters are all present. While hiragana consists of 48 types and thus offers relatively more training samples, kanji exceeds 4,000 types, leading to significant data imbalance.

Style preservation is also crucial, as it is essential to maintain the original writing style, pigments, and materials specific to the artifact. In Japanese ancient manuscripts, *kuzushiji*—a cursive script—is a representative example of an author-specific style that must be preserved during restoration. Moreover, complex degradation patterns such as stains, cracks, fading, and bleed-through each require different restoration strategies, and various other types of deterioration occur depending on storage conditions. Finally, avoiding over-restoration is important: without proper constraints, GANs may generate non-existent patterns, resulting in historically inaccurate reconstructions.

Beyond GAN-based approaches, recent studies have applied diffusion models for document restoration. Li et al. [4] proposed DiffACR, a diffusion-based framework for reconstructing eroded ancient Chinese characters by simulating degradation via cold diffusion and reversing the process. The model incorporates prior masks to guide restoration and demonstrated superior performance over conventional inpainting methods. Such diffusion models offer enhanced stability and controllability in generation, and are especially effective for preserving fine structural details in historical

scripts.

Evaluation of GAN-based restorations requires quantitative and qualitative assessments. Quantitative metrics include PSNR and SSIM for fidelity, as well as LPIPS [3] for perceptual quality.

## III. PROPOSED METHOD

This section presents our two-stage stain reduction framework based on diffusion models. We first outline the overall pipeline, then summarize the diffusion preliminaries (DDPM) and the restoration formulation (DDRM) that our method builds upon, followed by the proposed two-stage restoration with noise masking and final $\gamma$-correction.

### A. OVERVIEW

As illustrated in Fig. 3, our pipeline consists of two stages: **Stage 1 (SR1)** restores local patches to preserve fine textual details, and **Stage 2 (SR2)** operates on the reconstructed full page to enforce global consistency and further reduce stains. We guide the generation process using noise masking derived from binarized text regions to retain character strokes while suppressing background stains.

### B. PRELIMINARIES: DENOISING DIFFUSION PROBABILISTIC MODELS (DDPM)

A Denoising Diffusion Probabilistic Model (DDPM) gradually adds Gaussian noise to a clean image $x_0$ through a Markov chain, transforming it into pure noise $x_T$. The forward diffusion process is defined as:

$$q(x_t \mid x_{t-1}) = \mathcal{N}\Big(x_t; \sqrt{1-\beta_t}\, x_{t-1},\, \beta_t \mathbf{I}\Big)\,, \tag{1}$$

where $\beta_t$ denotes the variance schedule. The closed-form expression after $t$ steps is:

$$q(x_t \mid x_0) = \mathcal{N}\big(x_t; \sqrt{\bar{\alpha}_t}\, x_0,\, (1-\bar{\alpha}_t)\mathbf{I}\big)\,, \tag{2}$$

with $\alpha_t := 1-\beta_t$ and $\bar{\alpha}_t := \prod_{s=1}^{t} \alpha_s$. The reverse process learns to approximate the data distribution $p_\theta(x_0)$ by predicting the added noise $\epsilon$ through a neural network $\epsilon_\theta(x_t, t)$, minimizing:

$$L(\theta) = \mathbb{E}_{x_0,\epsilon,t}\left[\|\epsilon - \epsilon_\theta(x_t, t)\|^2\right]. \tag{3}$$

This denoising objective enables the model to approximate the complex data distribution that cannot be analytically modeled.

### C. DDRM FOR IMAGE RESTORATION

Denoising Diffusion Restoration Models (DDRM) extend DDPMs to solve linear inverse problems under a known degradation operator $\mathbf{H}$:

$$y = \mathbf{H}x + z, \tag{4}$$

where $y$ denotes the observed image and $z$ is Gaussian noise. where $\mathbf{H}$ is a known degradation operator and $z$ is noise. Using a conditional variational formulation,

$$q(x_{1:T} \mid x_0, y) := q^{(T)}(x_T \mid x_0, y) \prod_{t=0}^{T-1} q^{(t)}(x_t \mid x_{t+1}, x_0, y), \tag{5}$$

and the corresponding generative factorization,

$$p_\theta(x_{0:T}) = p_\theta^{(T)}(x_T) \prod_{t=0}^{T-1} p_\theta^{(t)}(x_t \mid x_{t+1}, y), \tag{6}$$

DDRM incorporates the singular value decomposition (SVD) of $\mathbf{H}$:

$$\mathbf{H} = U\Sigma V^T, \tag{7}$$

To operate in spectral space, we represent $\mathbf{H}$ using a pseudo-SVD. Let $\Sigma$ be a diagonal matrix where each element $s_i \in 0, 1$ indicates whether a pixel is observed ($s_i = 1$) or missing ($s_i = 0$). When $s_i = 0$, the corresponding component is unobserved and is generated unconditionally, whereas $s_i = 1$ means the generation is conditioned on the observed value $y$. $P$ denotes a permutation matrix that rearranges spatial positions.

Since exact SVD can be computationally expensive, task-specific pseudo-SVDs are employed. For inpainting tasks, we adopt the simplified form:

$$\mathbf{H} = I\Sigma P, \tag{8}$$

where $I$ is the identity matrix, $\Sigma$ is the binary diagonal visibility mask defined above, and $P$ permutes spatial positions. This formulation enables DDRM to reconstruct unobserved regions while aligning with the observed data.

### D. TWO-STAGE STAIN REDUCTION WITH NOISE MASKING

**Stain Reduction Stage 1 (Patch Stage).** The manuscript page is divided into smaller patches, each restored independently using DDRM with a DDPM pre-trained on randomly cropped manuscript patches. This preserves local character fidelity but may introduce slight inconsistencies in background tone between patches.

**Stain Reduction Stage 2 (Page Stage).** The restored patches are stitched together to form the full page, which is then refined by a second DDRM using a DDPM trained on full-page data. This stage improves global consistency and eliminates residual stains by referencing clean background statistics.

*a: Noise Masking for Character Guidance:* To preserve textual fidelity while suppressing stains, we introduce a *noise masking* mechanism that locally controls the denoising strength. A binary mask $\mathbf{H}$ is generated via fixed-threshold binarization, where text pixels are set to 1 and background to 0. We adopt a threshold of 128 following preliminary histogram analysis, where the luminance distributions of background paper and written strokes exhibited bimodal behavior

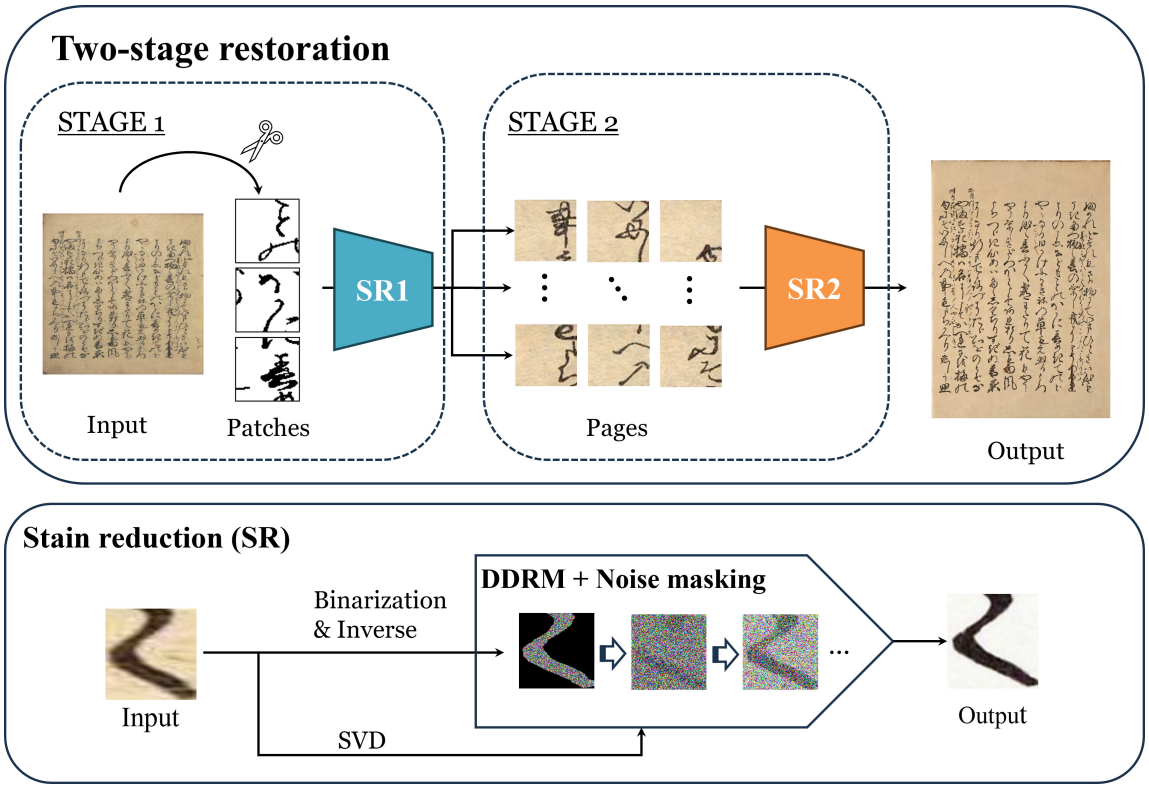



FIGURE 3: Two-stage stain reduction framework. **SR1:** patch-wise restoration with a DDPM trained on randomly cropped manuscript patches to preserve local character details. Patches are then stitched into a full page. **SR2:** page-wise restoration using a DDPM trained on full pages to enhance global consistency and remove residual stains. Noise masking ensures that binarized text regions are preserved while background stains are removed.

and the mid-range boundary fell near 128. As the mask is intended for coarse guidance rather than perfect segmentation, a fixed threshold provides stable behavior and avoids overadapting to local illumination variations, which can occur with adaptive methods such as Otsu thresholding. Thus, 128 serves as a robust midpoint for 8-bit document images and is widely used in historical-document binarization literature.

Given the network prediction $\mathbf{x}_{\theta,t}$ at diffusion step $t$, masked updating is defined as:

$$\mathbf{x}_{\text{mask}} = (1 - \mathbf{H}) \odot \mathbf{x}_{\theta,t}. \tag{9}$$

This selectively denoises background regions while suppressing updates in text regions. In practice, we replace background areas with a clean reference manuscript image, then apply DDRM with noise masking to refine character strokes. This enables locally adaptive restoration that retains stroke structure while removing stains and discoloration around text.

### *E. GAMMA CORRECTION*

Although DDRM reconstructs sharp strokes, the restored outputs sometimes appear low-contrast due to the limited training resolution. To enhance visual clarity, we apply post-processing gamma correction as follows:

$$y = x^{\gamma}, \quad \gamma = 0.8, \tag{10}$$

where $x$ and $y$ denote the DDRM-restored image and the gamma-corrected output, respectively. Note that these $x$ and $y$ differ from those in Eq. (4), where $x$ and $y$ represent the ground-truth (clean) image and the degraded observation modeled as $y = Hx + z$. In contrast, Eq. (10) is applied as a post-processing operation that adjusts luminance after restoration.

Gamma correction is a nonlinear transformation designed to match the human visual perception of brightness, which follows a power-law relationship rather than a linear one [37]. This operation redistributes luminance values so that mid-tone regions receive greater perceptual weight, improving readability and contrast without amplifying noise. Such nonlinearity has been widely used in the restoration of photographic and document images to compensate for faded luminance distributions and to optimize tone reproduction [36].

Following this principle, we examine gamma values in the range of 0.6–1.4 (Fig. 4). Empirically, $\gamma = 0.8$ yields the most perceptually balanced results: it enhances faded strokes while preserving the natural tone of the paper region. In contrast, $\gamma > 1.0$ tends to produce faded text, while $\gamma < 0.7$ results in excessive brightening of the background, diminishing tonal authenticity. Therefore, $\gamma = 0.8$ is adopted as a visually consistent and experimentally stable parameter that improves readability without altering the restored structure.

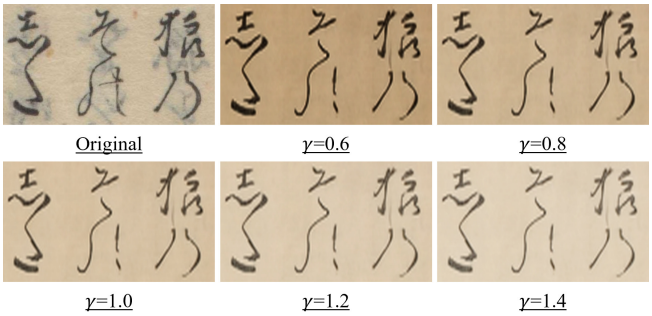



FIGURE 4: Visual comparison of gamma-corrected outputs ($\gamma = 0.6$–$1.4$) applied to the restored images produced by the proposed method. $\gamma = 0.8$ provides the best balance between character visibility and background tone.

The overall processing pipeline, including preprocessing, two-stage restoration, and gamma correction, is summarized in Algorithm III-E.

[t] Degraded manuscript image $I_{\text{raw}}$ (original size: $H_0 \times W_0$) Restored manuscript image $I_{\text{out}}$

$I_{\text{resize}} \leftarrow \text{RESIZE}(I_{\text{raw}},\ 704,\ 704)$ $I_{\text{gray}} \leftarrow \text{TO\_GRAYSCALE}(I_{\text{resize}})$ $I_{\text{bin}} \leftarrow \text{BINARIZE}(I_{\text{gray}};\ \text{thr} = 128)$ $\{P_i\}_{i=1}^{N} \leftarrow \text{PATCHIFY}(I_{\text{bin}};\ \text{size} = 64,\ \text{stride} = 64)$

**Phase 1: Character-level (Patch Stage)** $i \in \{1, \ldots, N\}$ $P_i' \leftarrow \text{DDRM\_RESTORE}(P_i;\ \text{model} = \mathcal{M}_{\text{char}},\ \text{diff\_steps} = 20,\ \text{mask\_steps} = 100)$ $I_{\text{char}} \leftarrow \text{RECONSTRUCT}(\{P_i'\};\ \text{out\_size} = 704 \times 704)$

**Intermediate:** $I_{\text{char\_bin}} \leftarrow \text{BINARIZE}(I_{\text{char}};\ \text{thr} = 128)$ $I_{\text{page\_input}} \leftarrow I_{\text{char\_bin}}$

**Phase 2: Page-level (Page Stage)** $I_{\text{page\_out}} \leftarrow \text{DDRM\_RESTORE}(I_{\text{page\_input}};\ \text{model} = \mathcal{M}_{\text{page}},\ \text{diff\_steps} = 20,\ \text{mask\_steps} = 100)$

**Postprocessing:** $I_\gamma \leftarrow \text{GAMMA\_CORRECT}(I_{\text{page\_out}};\ \gamma = 0.8)$ $I_{\text{out}} \leftarrow \text{RESIZE}(I_\gamma;\ \text{size} = H_0 \times W_0)$ $I_{\text{out}}$

## IV. EXPERIMENTS AND RESULTS

This section describes the datasets, experimental conditions, evaluation metrics, and quantitative/qualitative results.

### A. DATASETS

We use page images from the "Kuzushi-ji Dataset" [38] provided by the Center for Open Data in the Humanities (CODH). Two works are targeted: *Isoho Monogatari* and *Tsurezuregusa*. *Isoho Monogatari* is a movable-type printed edition from the early Edo period (Kanei era, 1624–1644), consisting of three volumes with a total of 198 pages. It exhibits deterioration such as stains and insect damage; page images and character-level annotations are available from the National Institute of Japanese Literature. *Tsurezuregusa*, written in the early Keicho era (ca. 1596), consists of 322 pages and shows progressive degradation including color fading and accumulated dirt. For this work, only high-resolution scans are provided without page-level or character-level annotations. Therefore, for *Tsurezuregusa*, we manually crop the main document regions using the open-source annotation tool labelImg, removing background margins and surrounding objects to obtain clean document areas for evaluation.

TABLE 1: Test data used in this study

| Work | Period | Pages | Annotation |
|---|---|---|---|
| *Isoho Monogatari* | Kanei era (1624–1644) | 198 | Yes |
| *Tsurezuregusa* | Early Keicho era (ca. 1596) | 322 | No |

In this study, the term *Original (Raw Data)* refers to the publicly released manuscript images containing real degradations such as stains and fading. Since these images are actual historical documents, no clean ground truth exists. Hence, quantitative metrics such as PSNR and SSIM are computed with respect to these original images for relative comparison, indicating improvement in perceptual quality rather than absolute fidelity.

In addition to PSNR and SSIM, the following evaluation metrics were employed to provide a more comprehensive assessment of restoration performance: kurtosis, brightness, entropy, and RMSE.

Kurtosis represents the sharpness of the pixel intensity distribution in the input image, reflecting the clarity of ink contrast and edge definition. It is mathematically defined as:

$$Kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^2} - 3 \qquad (11)$$

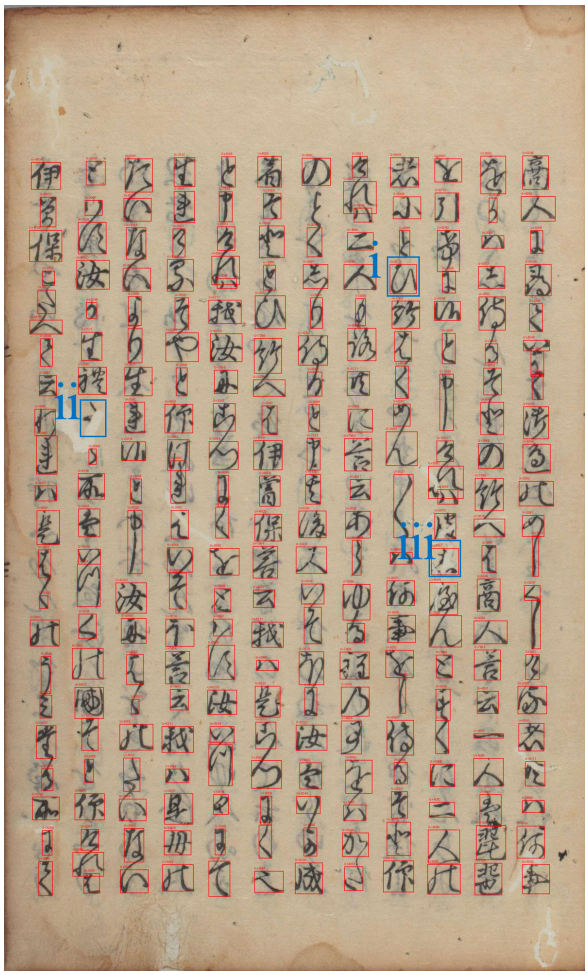
where $x_i$ denotes the pixel intensity, $\bar{x}$ is the mean intensity, and $N$ is the total number of pixels. A lower kurtosis value indicates that a wider range of pixel intensities is present, suggesting that stain components have been effectively reduced and the image has become visually smoother.

Brightness measures the dynamic range of luminance values; a higher brightness variance indicates a greater contrast between background and text regions, contributing to improved readability.

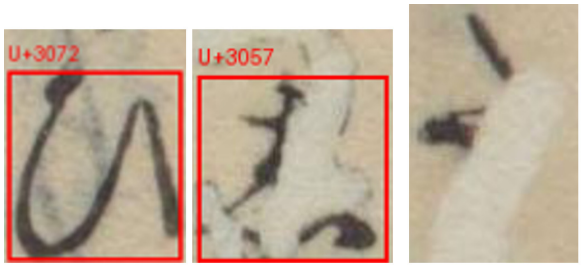
Entropy quantifies the overall complexity of image information. Higher entropy typically results from the presence of random noise, whereas lower entropy corresponds to cleaner and more homogeneous regions. Unlike kurtosis, which captures the peakedness of intensity distribution, entropy reflects the degree of randomness and textural variability across the image.

RMSE (Root Mean Square Error) measures the average deviation from the reference image. Since the original manuscripts already contain stains and deterioration, RMSE in this study serves as an indicator of how effectively those degradations are suppressed rather than absolute pixel-level accuracy.

These quantitative indices are analyzed alongside visual comparisons of the restored outputs to verify the relationship between numerical evaluation and perceptual improvement.



(a) Page-level annotation subset for *Isoho Monogatari*. Main-body characters are annotated; distorted or unrecognizable glyphs are excluded.




(b) Character-level annotation cases: (i) annotated character, (ii) degraded but annotated, (iii) unannotated.

FIGURE 5: Annotation overview for *Isoho Monogatari*.

### B. COMPARISON METHODS

To validate the effectiveness of the proposed method, we conduct comparative experiments with several baseline approaches that are commonly used in the restoration of historical documents.

**(1) Otsu's Binarization (Binary)** A classical non-learning global thresholding method that separates text and background regions based on pixel intensity distribution. It serves as the most fundamental baseline for stain removal and text extraction.

However, since the binarized images differ substantially in visual characteristics and dynamic range from other restored outputs, they are excluded from direct quantitative comparison in this study. Instead, they are used only for qualitative reference to illustrate the contrast between binarization and restoration-based methods.

**(2) Background-Filled Binarization (Binary_Fill):** An extended version of Otsu's method, in which the background regions obtained after binarization are filled with an estimated paper color derived from the original image. This process compensates for the visual discontinuity caused by harsh binarization and provides a simple background correction.

**(3) CycleGAN-Based Restoration (CycleGAN)** For a fair comparison, the CycleGAN baseline is trained under preprocessing and input conditions consistent with those used for our method. Each manuscript page is resized to $128\times128$ pixels, matching the resolution used in the page-level diffusion setting. A paired dataset consisting of 60 degraded manuscript images and their 60 clean counterparts (120 samples in total) is used for training. Because clean historical manuscript images are scarce, the training dataset is inherently limited, resulting in fewer samples for CycleGAN.

**(4) Two-Stage Restoration (Ours):** Our method integrates a two-stage Denoising Diffusion Restoration Model (DDRM) framework guided by binarization-based noise masking. The first stage (SR1) performs local patch restoration to recover character details, and the second stage (SR2) refines global consistency at the page level.

**(5) Proposed Method with Gamma Correction (Ours, $\gamma$=0.8):** A post-processed version of our two-stage DDRM framework that applies gamma correction to enhance visual contrast and readability while preserving structural fidelity.

These comparative settings allow us to evaluate the proposed method against both traditional image processing techniques and modern deep-learning-based restorations. The evaluation focuses on both quantitative metrics (PSNR, SSIM, kurtosis, brightness, entropy, RMSE) and qualitative visual consistency.

### C. EXPERIMENTAL CONDITIONS

**Hardware/OS**

Intel Xeon(R) Gold 6342 @ 2.80GHz (single node) and a single NVIDIA RTX A6000 GPU; Ubuntu 22.04.1 LTS.
CUDA 11.8, cuDNN 8.7.0, and PyTorch for deep-learning implementation.

**Training Data**

Two DDPMs were employed in this study: (i) a patch-level model trained on 100,240 randomly cropped $64\times64$ patch images, and (ii) a page-level model trained on 4,829 page images resized to $128\times128$ to accommodate computational resource constraints. Both models were integrated into the

DDRM-based two-stage restoration pipeline described in Sec. III.

No blending or smoothing algorithms were applied during patch reassembly. This omission was intentional, allowing us to directly evaluate the intrinsic restoration behavior and local consistency of DDRM without the influence of post-processing effects.

### D. RESULTS

#### 1) Overall Quantitative and Qualitative Evaluation

For Tsurezuregusa, as shown in Fig. 6, the proposed method achieves cleaner backgrounds and better preservation of shading and character sharpness than simple binarization or background-corrected binarization (Binary_Fill), effectively avoiding the excessive distortion typically caused by thresholding. As summarized in the upper table of Fig. 7, our method attains the highest SSIM (0.715) with well-balanced values in kurtosis, brightness variance, and entropy, even though its PSNR (15.967) is slightly lower than that of CycleGAN (21.536). After applying gamma correction ($\gamma$=0.8), PSNR further improves to 18.032, with slight decreases in kurtosis and brightness variance, while SSIM remains constant at 0.715. Regarding the RMSE (a)–(c) values shown in the figure, the proposed method achieves comparable or higher scores than CycleGAN, particularly in RMSE(a), indicating its ability to effectively suppress stains and local intensity inconsistencies within document regions.

For Isoho Monogatari, similar trends are observed (Fig. 9). As presented in the upper table of the same figure, the proposed method achieves a PSNR of 18.476, which is lower than CycleGAN's 21.176, but attains the highest SSIM (0.776) among all methods. In comparison with binarization, the brightness variance and entropy are substantially improved. With gamma correction, PSNR further increases to 20.683, while SSIM remains almost unchanged (0.775), indicating enhanced contrast without significant structural degradation.

Across both manuscripts, the proposed method consistently outperforms conventional binarization and CNN-based generative approaches such as CycleGAN. Gamma correction enhances visibility by increasing contrast and brightness variance, although slight trade-offs are observed in kurtosis, brightness, entropy, and RMSE. Overall, the proposed framework successfully balances stain removal and character preservation.

As shown in Fig. 7 and Fig. 9, the degraded regions containing stains and bleed-through exhibit multiple noisy peaks compared with the background, reflecting luminance fluctuations due to deterioration. After restoration by the proposed method, these peaks are significantly reduced and their distributions approach those of the clean background, demonstrating effective suppression of stain-induced noise. This trend is consistently observed in both *Tsurezuregusa* and *Isoho Monogatari*.

Furthermore, Fig. 10 presents the evaluation results based on 45,358 annotated characters, where the character samples were randomly extracted from the test manuscripts. Without gamma correction, the proposed method achieves higher PSNR (16.970) and SSIM (0.700) than the binarization baselines, though slightly below CycleGAN. However, the entropy (3.812) is notably reduced, and visual inspection reveals that the proposed method successfully removes bleed-through artifacts near text regions without introducing the excessive deformation seen in Binary_Fill. In addition, kurtosis (1.627) and brightness variance (185.391) are markedly improved compared with binarization, indicating better contrast and more natural stroke restoration. When applying gamma correction ($\gamma$=0.8), PSNR increases to 19.042 and SSIM to 0.705, with minor decreases in brightness variance and kurtosis, reflecting a tunable trade-off between pixel-level fidelity and perceptual contrast.

## V. DISCUSSION

The experimental results confirm that the proposed two-stage DDRM-based restoration method effectively reduces stains and bleed-through while preserving character shapes. In both Tsurezuregusa and Isoho Monogatari, qualitative comparisons show that our method produces cleaner backgrounds and sharper strokes than conventional binarization approaches. Furthermore, the quantitative evaluation demonstrates that the proposed method consistently achieves superior SSIM performance, and also yields favorable results in terms of brightness variance and entropy, indicating improved contrast and reduced background noise compared with binarization.

The improvements observed in PSNR, SSIM, brightness variance, kurtosis, entropy, and RMSE can be explained by how the proposed diffusion-based restoration pipeline treats different types of degradation at both the patch level (SR1) and the page level (SR2).

### A. ANALYSIS OF QUANTITATIVE AND STRUCTURAL RESULTS

In heavily stained regions and bleed-through areas, the mask-guided diffusion process restricts updates primarily to background pixels, suppressing low-frequency discoloration from paper aging, oxidation, or ink transfer. This results in a more homogeneous, clean background, improving PSNR by reducing large-scale intensity noise and lowering entropy by removing random luminance fluctuations. Histograms in Fig. 7 and Fig. 9 confirm this effect: multi-modal degraded distributions collapse into unimodal distributions resembling clean paper.

For SSIM, preservation of character stroke geometry is key. Unlike hard binarization, which discards grayscale nuance and introduces jagged edges, the diffusion model leverages a generative prior to refine stroke boundaries smoothly. This maintains local contrast and edge continuity, yielding the highest SSIM in both *Tsurezuregusa* (0.715) and *Isoho Monogatari* (0.776), even when CycleGAN achieves higher PSNR. While CycleGAN increases global smoothness (in-


FIGURE 6: Restoration comparison for *Tsurezuregusa*. (i) Original image. (ii) Simple binarization using Otsu's method (non-learning baseline). (iii) Binary image with background correction, where white regions in (ii) are filled with the estimated paper color. (iv) CycleGAN-based restoration. (v) Proposed two-stage restoration guided by a binary mask. (vi) Proposed method with gamma correction ($\gamma$=0.8) for enhanced visual contrast. Red rectangles highlight enlarged character regions for qualitative comparison.

**Comparison of Evaluation Metrics for *Tsurezuregusa***

| | (i)Original | (ii)Binary | (iii)Binary_Fill | (iv)CycleGAN | (v)Ours | (vi)Ours ($\gamma$=0.8) |
|---|---|---|---|---|---|---|
| PSNR ↑ | – | 8.715 | 15.290 | 21.536 | 15.967 | 18.032 |
| SSIM ↑ | – | 0.622 | 0.669 | 0.656 | **0.715** | **0.715** |
| Kurtosis ↑ | 3.950 | 4.745 | 4.745 | 0.159 | **7.784** | **7.290** |
| Brightness ($\sigma$) ↑ | 31.730 | 228.000 | 178.968 | 164.546 | **199.430** | **188.070** |
| Entropy ↓ | 12.080 | 0.338 | 0.338 | 4.839 | **3.583** | **3.717** |
| RMSE(a) ↑ | – | 88.115 | 19.571 | 23.401 | **34.833** | **24.050** |
| RMSE(b) ↑ | – | 87.742 | 41.144 | 31.764 | **38.713** | 28.593 |
| RMSE(c) ↑ | – | 91.584 | 30.592 | 28.327 | **45.729** | **36.203** |

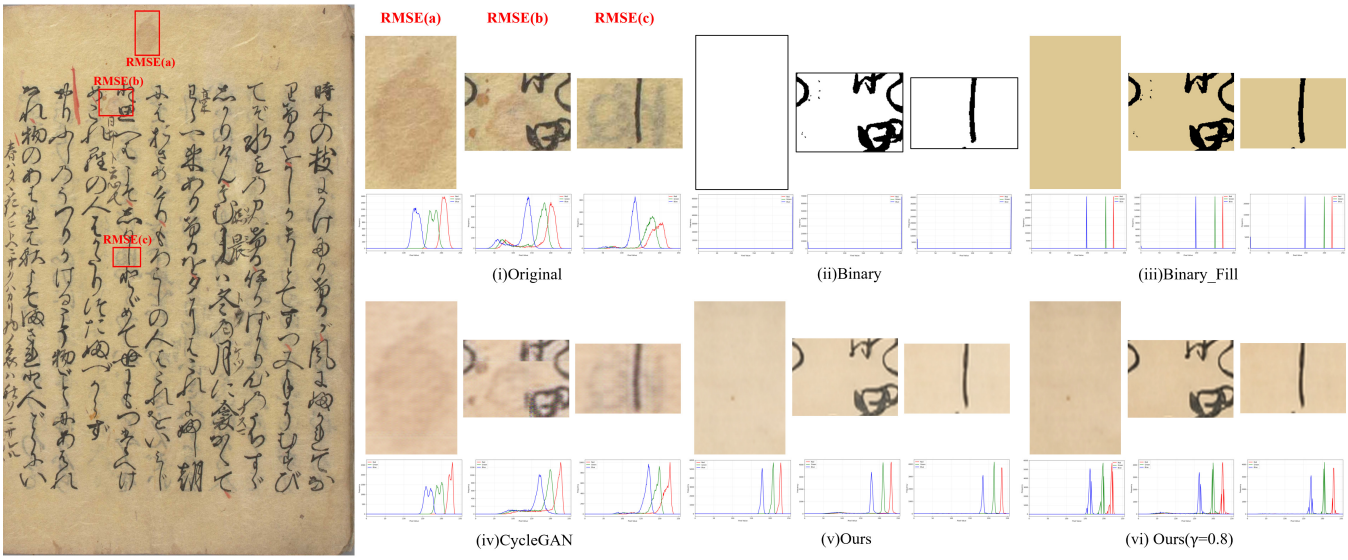



FIGURE 7: Quantitative and local evaluation for *Tsurezuregusa*. The subfigures (i)–(vi) correspond to the same restoration methods described in Fig. 6. Red rectangles indicate the regions used to compute RMSE, while the in-image histograms represent the RGB intensity distributions within those areas, reflecting the degree of stain removal. The upper table summarizes PSNR, SSIM, Kurtosis, Brightness ($\sigma$), Entropy, and RMSE (a)–(c) values across different methods.

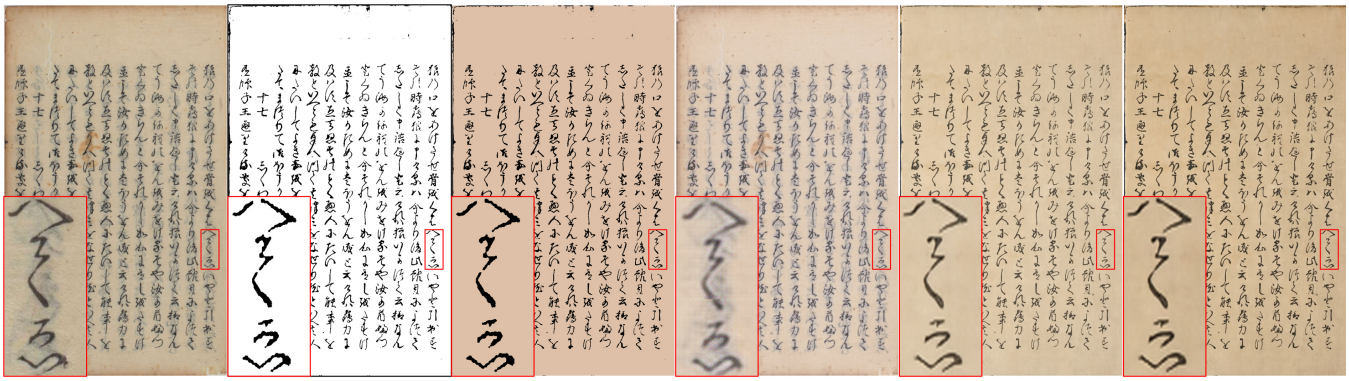



FIGURE 8: Restoration comparison for *Isoho Monogatari*. (i) Original image. (ii) Simple binarization using Otsu's method. (iii) Binary image with background correction, where white regions in (ii) are filled with the estimated paper color. (iv) CycleGAN-based restoration. (v) Proposed two-stage restoration guided by a binary mask. (vi) Proposed method with gamma correction ($\gamma$=0.8) for enhanced visual contrast. Red rectangles highlight enlarged character regions for qualitative comparison.

**Comparison of Evaluation Metrics for *Isoho Monogatari***

| | (i)Original | (ii)Binary | (iii)Binary_Fill | (iv)CycleGAN | (v)Ours | (vi)Ours ($\gamma$=0.8) |
|---|---|---|---|---|---|---|
| PSNR ↑ | – | 9.737 | 17.564 | 21.176 | 18.476 | 20.683 |
| SSIM ↑ | – | 0.694 | 0.736 | 0.761 | **0.776** | **0.775** |
| Kurtosis ↑ | 5.679 | 10.517 | 10.517 | 0.543 | **11.023** | **5.679** |
| Brightness ($\sigma$) ↑ | 173.803 | 237.273 | 184.067 | 170.524 | **199.440** | **173.803** |
| Entropy ↓ | 4.174 | 0.251 | 0.251 | 4.838 | **3.543** | **4.170** |
| RMSE(a) ↑ | – | 83.268 | 22.202 | 6.735 | **27.400** | **18.947** |
| RMSE(b) ↑ | – | 96.245 | 45.400 | 14.104 | **38.035** | **29.396** |
| RMSE(c) ↑ | – | 89.012 | 44.549 | 25.273 | **28.368** | 21.618 |

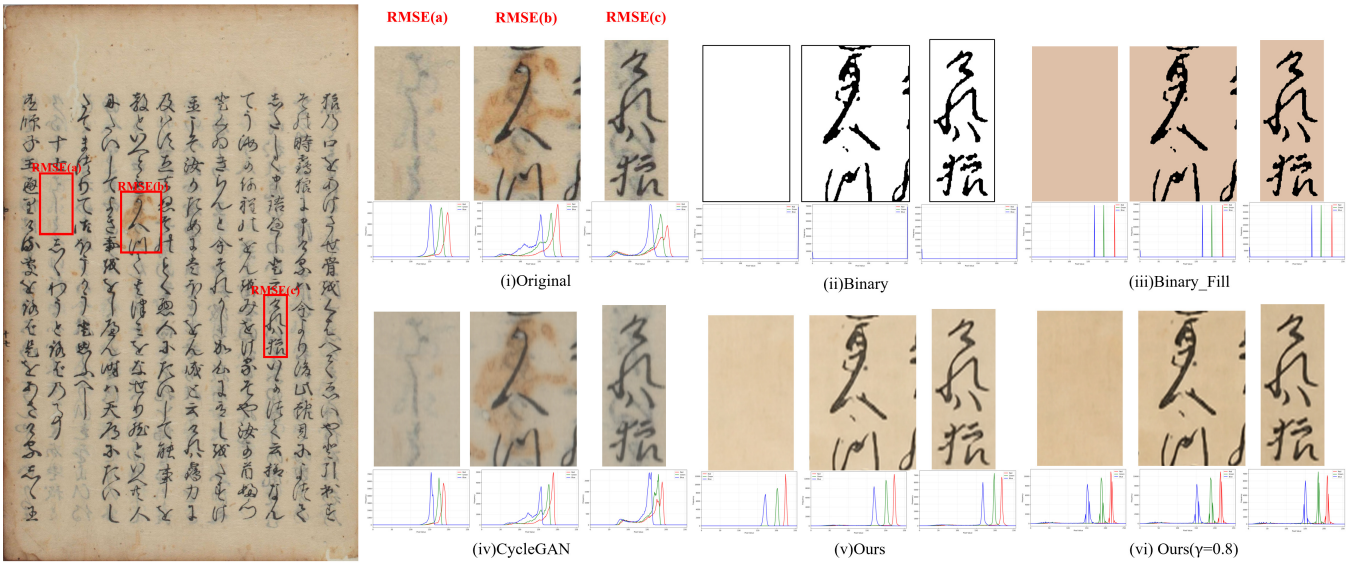



FIGURE 9: Quantitative and local evaluation for *Isoho Monogatari*. The subfigures (i)–(vi) correspond to the same restoration methods described in Fig. 9. Red rectangles indicate the regions used to compute RMSE, while the in-image histograms represent the RGB intensity distributions within those areas, reflecting the degree of stain removal. The upper table summarizes PSNR, SSIM, Kurtosis, Brightness ($\sigma$), Entropy, and RMSE (a)–(c) values across different methods.

**Character-level Evaluation Metrics for *Isoho Monogatari***

| | (i)Original | (ii)Binary | (iii)Binary_Fill | (iv)CycleGAN | (v)Ours | (vi)Ours ($\gamma$=0.8) |
|---|---|---|---|---|---|---|
| PSNR ↑ | – | 9.076 | 14.562 | 20.611 | 16.970 | 19.042 |
| SSIM ↑ | – | 0.492 | 0.543 | 0.751 | 0.700 | 0.705 |
| Kurtosis ↑ | 0.331 | 1.000 | 1.000 | 1.509 | **1.627** | 1.488 |
| Brightness ($\sigma$) ↑ | 153.251 | 206.702 | 159.687 | 165.230 | **185.391** | **172.791** |
| Entropy ↓ | 4.462 | 0.476 | 0.476 | 4.811 | **3.812** | **3.903** |
| RMSE ↑ | – | 91.520 | 52.194 | 29.518 | **37.639** | **29.924** |

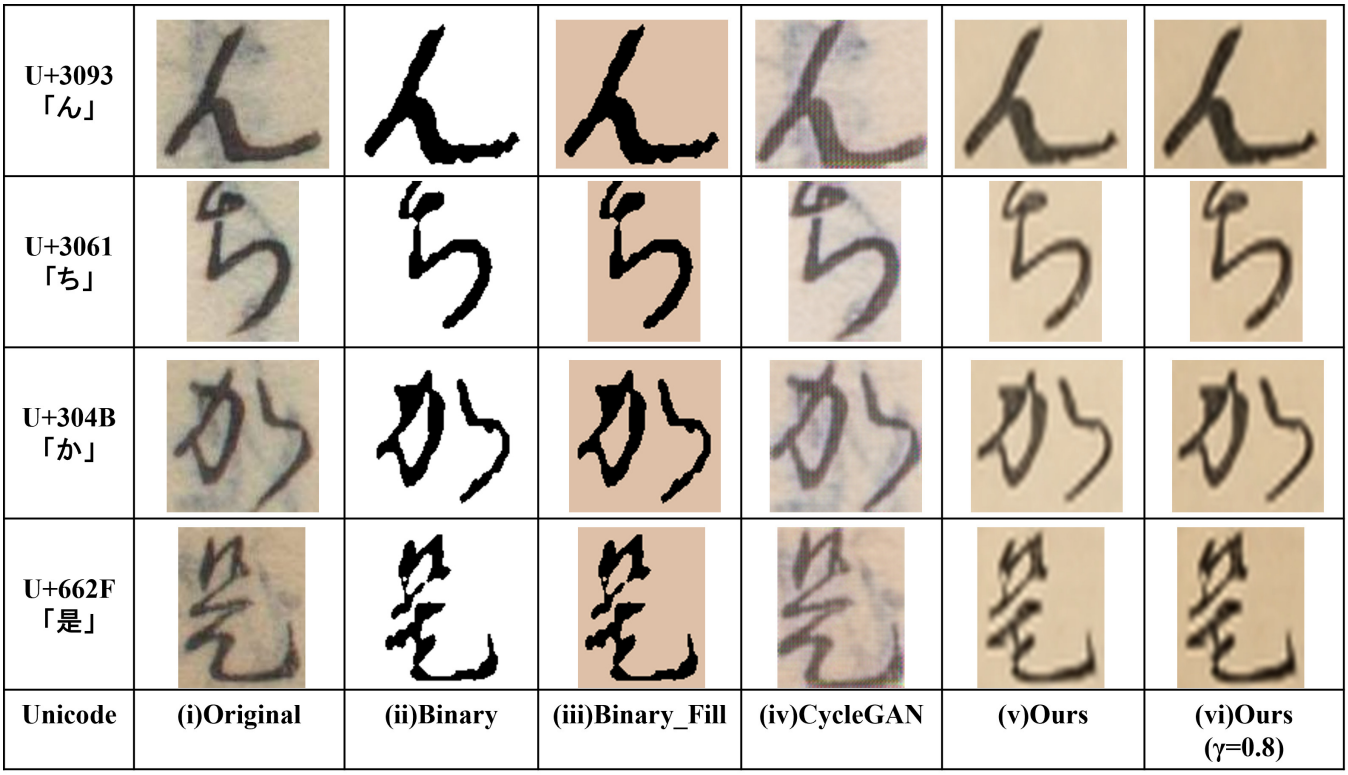


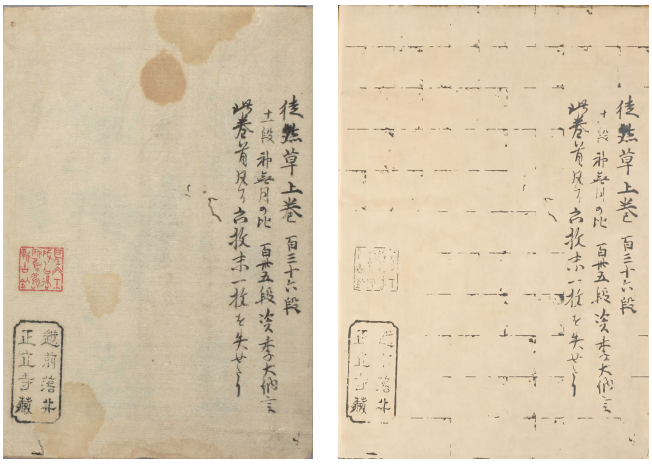

FIGURE 10: Character-level restoration comparison for *Tsurezuregusa*: (a) Ground truth. (b) Simple binarization using Otsu's method (non-learning baseline). (c) Binary image with background correction, where white regions of (b) are filled with estimated paper color. (d) CycleGAN-based restoration. (e) Proposed two-stage restoration guided by a binary mask. (f) Proposed method with gamma correction ($\gamma$=0.8).

flating PSNR), it sometimes oversmooths or hallucinates strokes, whereas our method maintains stroke fidelity and removes background clutter, leading to superior local structural similarity.

Brightness variance further supports these observations. Higher luminance variance compared to binarization indicates enhanced text–background separation: stains are reduced, ink strokes remain dark, and contrast improves. Unlike global contrast stretching, this selective enhancement arises naturally from stain removal while preserving stroke intensity.

Kurtosis also increases under our method, reflecting sharper, edge-like intensity structures. Although binarized images have only two intensity levels, their distributions are dominated by uniform regions, producing flatter histogram peaks. In contrast, our method produces smooth backgrounds with steep transitions at stroke edges, yielding statistically higher kurtosis. Concurrently, entropy decreases relative to CycleGAN and degraded inputs, indicating reduced stochastic noise without loss of textual detail.

RMSE trends corroborate these findings. Since the "reference" includes stains, methods that remain visually close to the degraded input can have lower RMSE. Nevertheless, our method achieves comparable or better RMSE in cropped regions, while actively removing stains. This indicates that DDRM reconstructs plausible paper texture rather than introducing flat fills, unlike Binary_Fill. Although Binary_Fill may achieve lower RMSE in some cases, it destroys grayscale stroke structure and produces unnatural tone transitions. In contrast, higher SSIM and sharper edge features in our method reflect accurate preservation of stroke curvature and tapering, even when RMSE increases due to necessary



(a) Original manuscript image containing red text.

(b) Restoration result with visible seams between patches.

FIGURE 11: Examples of failure cases in high-resolution inference. (a) shows the original manuscript image containing red annotations. (b) presents the restoration result, where visible seams appear along patch boundaries due to insufficient global consistency and the red annotations are not reproduced. These cases highlight current challenges in maintaining seamless page integration and preserving colored annotations.

deviation from stained pixels. Thus, RMSE alone is insufficient to measure restoration quality; SSIM and perceptual structure better capture historical stroke fidelity.

Overall, higher SSIM and favorable brightness variance and kurtosis values arise from accurate reconstruction of stroke detail and contrast, while reduced entropy and competitive RMSE reflect effective stain suppression without tone loss. These results demonstrate that the proposed method performs structured, region-aware restoration aligned with text–background semantics.

Finally, performance varies by degradation type. For stains and bleed-through, masking successfully restricts denoising to background regions, improving background uniformity while preserving character structure. However, faint or partially missing strokes may be excluded by binarization, occasionally causing stroke fragmentation. Future work could employ adaptive or confidence-weighted masks to preserve low-intensity strokes while suppressing noise.

### B. CHALLENGES IN HIGH-RESOLUTION INFERENCE AND MASK DEPENDENCY

In high-resolution inference, seams occasionally appear at the boundaries between patches (Fig. 11(a)). Since blending and smoothing algorithms were intentionally not applied in this study, these artifacts reflect the intrinsic behavior of DDRM rather than any post-processing effect. This phenomenon suggests that while DDRM performs stable restoration within individual patches, it lacks global coherence across neighboring regions.

Moreover, the proposed method is highly dependent on the binary mask. When generating the second-stage mask, the seams that appeared during the first-stage restoration were partially inherited through binarization. Consequently, these artifacts propagated into the second-stage restoration, resulting in the visible seams observed in Fig. 11(a). This indicates that the binarization step can amplify discontinuities introduced during the initial stage, underscoring the need for more consistent mask generation across patch boundaries.

Future improvements should therefore focus on developing patch-fusion or multi-scale inference strategies to enhance overall page-level consistency.

As shown in Fig. 11(b), red annotations and decorative elements are sometimes not reconstructed. This limitation likely arises because the current training process primarily emphasizes the restoration of black text. Explicitly including "colored-text classes" in training would help preserve these annotations. Alternatively, a post-restoration compositing strategy—where red or decorative elements detected prior to restoration are reattached to the reconstructed output—could also be effective. These omissions may have also contributed to the observed decrease in similarity-based metrics such as PSNR and SSIM, despite the improvement in perceptual clarity.

### C. FADING OF CHARACTER STROKES

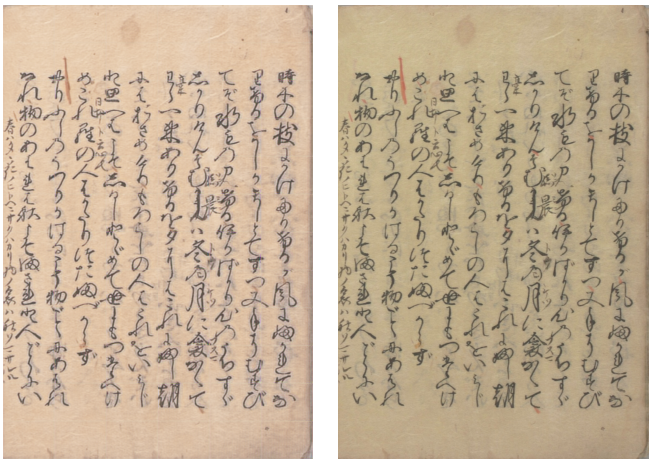
During the diffusion process, reconstructed images tend to converge toward the average tone of the training background, causing characters to appear slightly lighter or thinner. Adjusting the number of reverse diffusion steps and the weighting of noise masking is essential to balance stain reduction and ink density preservation.

### D. EFFECT OF GAMMA CORRECTION

Gamma correction is adopted as a post-processing step to improve contrast. While it enhances visual readability and increases brightness variance, it slightly decreases SSIM and kurtosis. This indicates minor structural distortion, suggesting that gamma correction should be interpreted as a complementary perceptual enhancement rather than a substitute for DDRM-based restoration.

### E. RESOLUTION CONSTRAINTS IN TRAINING

Due to computational limitations, the page-level model is trained on $128\times128$ images and the patch-level model on $64\times64$ patches. This inevitably limits the model's ability to reconstruct fine-grained texture and subtle ink variations, occasionally leading to slightly blurred edges or uneven tonal transitions in dense text regions. If computational resources allow, training at higher resolutions would be desirable. In particular, resizing during preprocessing may have introduced jagged edges in both input and binarized images, which subsequently propagated into the restoration results. Combining global low-resolution context with local high-resolution refinement through multi-scale training could therefore improve both structural fidelity and tonal stability.



(a) Trained on real-stain data (b) Trained on synthetic-stain data

FIGURE 12: Comparison of CycleGAN trained on different degradation sources. (a) CycleGAN trained on real historical stains, and (b) CycleGAN trained on synthetic stains generated using the ocrodeg algorithm. Despite realistic textures in the synthetic degradation, the model trained on synthetic stains fails to remove real stains effectively, indicating a domain gap between pseudo and authentic degradation patterns.

### F. LIMITATIONS OF EVALUATION METRICS

Since truly clean, degradation-free ground truth manuscripts rarely exist, conventional similarity metrics such as PSNR and SSIM provide only partial insight. Stain-reduction indicators such as kurtosis, brightness variance, and entropy describe background restoration but do not necessarily correlate with human readability. Future work should establish task-specific metrics—such as OCR or kuzushiji-recognition accuracy—as proxy measures for practical readability.

### G. LIMITATIONS OF SYNTHETIC DEGRADATION FOR GAN-BASED RESTORATION

To further analyze the robustness of GAN-based restoration, we conducted an additional experiment comparing CycleGAN trained on synthetically degraded images with its performance on real historical degradation. Specifically, we applied synthetic stain degradation using the ocrodeg library provided by NVLabs under the InstructIR framework, which simulates realistic deterioration patterns such as stains, abrasion, and mold. In this synthetic-degradation training, we followed the same resolution setting as the DDRM page-level learning: clean manuscript pages were resized to $128\times128$, synthetic stains were applied to these downsampled images, and CycleGAN was trained on the resulting pseudo-degraded pairs. This ensured a fair comparison under the same input resolution constraints as DDRM.

Figure 12 presents a qualitative comparison between (a) real-stain input and (b) synthetically degraded input. Although the synthetic degradation was designed to replicate real-world aging, CycleGAN trained on synthetic-stain data failed to remove stains effectively in real cases. This gap suggests that artificially generated degradations, despite their realism, do not fully capture the complex spatial, chromatic, and textural characteristics of genuine manuscript aging.

These findings highlight a key limitation in relying solely on synthetic data for training GAN-based restoration models. Under the current training conditions, the CycleGAN trained on pseudo-degraded images could not effectively remove real stains, suggesting that synthetic degradations, while visually realistic, do not fully capture the complex spatial, chromatic, and textural characteristics of authentic historical aging.

To summarize, DDRM demonstrated high effectiveness in the digital restoration of cultural heritage manuscripts. Future research should focus on improving tonal consistency, developing adaptive pre-/post-processing techniques, implementing patch-level blending, and establishing expert-informed evaluation frameworks tailored to ancient documents.

Recent advances in diffusion-based methods have sought to overcome the limitations of GANs and synthetic degradation. For instance, DiffACR [4] employs a cold diffusion process to iteratively restore eroded ancient Chinese characters by modeling the degradation as a reversible corruption process. Their approach demonstrates improved stroke fidelity compared to traditional GAN-based restoration, especially for severely faded single characters. While DiffACR is effective for isolated glyphs, it remains limited to character-level reconstruction and does not address broader page-level degradation patterns such as bleed-through or multi-region stains. In contrast, our method performs two-stage, mask-guided restoration across full manuscript pages, enabling semantic-aware and region-specific stain suppression without requiring paired clean data.

## VI. CONCLUSIONS

This study proposes a two-stage stain reduction framework for ancient Japanese manuscripts by combining conventional binarization with Denoising Diffusion Restoration Models (DDRM). The framework integrates binarization-based noise masking with patch-level and page-level diffusion restoration, enabling selective suppression of stains and bleed-through while preserving character structure and tonal consistency. Experimental evaluations on *Tsurezuregusa* and *Isoho Monogatari* demonstrated that the proposed method outperforms conventional binarization and generative baselines in both quantitative and qualitative aspects.

The mechanistic analysis reveals that the improvements in PSNR, SSIM, brightness variance, and kurtosis stem from the diffusion model's ability to restore smooth background luminance and sharp edge transitions simultaneously. The noise-masked diffusion selectively removes low-frequency stain components in background regions, leading to increased PSNR and reduced entropy, while preserving stroke continuity that contributes to higher SSIM. Moreover, the enhanced kurtosis and brightness variance indicate clearer foreground–background separation and better-defined ink bound-

aries, providing both structural and perceptual benefits.

In addition to visual enhancement, the proposed method contributes to the practical preservation and readability of cultural heritage materials. However, some limitations remain—such as the slightly pale appearance of restored characters, seams between patches, and the challenge of maintaining colored annotations. Addressing these issues through adaptive mask generation, multi-scale inference, and color-aware training will be crucial for future improvements.

Further experiments with gamma correction ($\gamma = 0.8$) showed that moderate contrast enhancement improves PSNR and perceptual clarity, even though it slightly reduces SSIM and kurtosis due to local structural deformation. These findings suggest that gamma correction acts as a complementary post-processing technique that enhances human readability without altering overall restoration fidelity. Incorporating adaptive gamma control and expert-informed perceptual evaluation will enable a more balanced optimization between quantitative accuracy and visual accessibility.

In conclusion, this work establishes a diffusion-based framework as a promising approach for restoring degraded historical manuscripts. Beyond Japanese heritage documents, the proposed methodology can be extended to multilingual and multimaterial archives, contributing to both scholarly research and long-term digital preservation of global cultural heritage.

## REFERENCES


[1] Y. Yoshizu, H. Kaneko, R. Ishibashi, L. Meng, and M. Deng, "Two-Stage Stains Reduction With Diffusion Models for Ancient Japanese Manuscripts," Proc. International Conference on Cyber-Physical Social Intelligence (ICCSI), pp. 1–6, 2024.

[2] W. Zheng, B. Su, R. Feng, X. Peng, and S. Chen, "EA-GAN: Restoration of Text in Ancient Chinese Books Based on an Example Attention Generative Adversarial Network," Heritage Science (Springer Nature), Vol. 11, No. 23, pp. 1–14, Jan. 2023.

[3] Zhang, R., Isola, P., Efros, A. A., Shechtman, E. & Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 586–595 (2018).

[4] Li, X., & Chen, M. (2024). DiffACR: Diffusion-Based Restoration of Eroded Ancient Characters. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

[5] Z. Xu, C. Zhang, and Y. Wu, "Digital Inpainting of Mural Images Based on DC-CycleGAN," Multimedia Tools and Applications (Springer), Vol. 82, No. 21, pp. 31247–31267, Nov. 2023.

[6] J. Cao, Y. Li, Q. Zhang, and H. Cui, "Restoration of an Ancient Temple Mural by a Local Search Algorithm of an Adaptive Sample Block," Journal of Cultural Heritage, Vol. 39, pp. 190–198, Mar. 2019.

[7] M. Jmal, W. Souidene, and R. Attia, "Efficient Cultural Heritage Image Restoration with Nonuniform Illumination Enhancement," Journal of Cultural Heritage, Vol. 28, pp. 55–64, Jan. 2017.

[8] S. Poornapushpakala, S. Barani, M. Subramoniam, and T. Vijayashree, "Restoration of Tanjore Paintings Using Segmentation and In-painting Techniques," Journal of Physics: Conference Series, Vol. 2161, No. 1, Article ID 012014, pp. 1–8, Jan. 2022.

[9] C. Lv, Z. Li, Y. Shen, J. Li, and J. Zheng, "SeparaFill: Two Generators Connected Mural Image Restoration Based on GAN With Skip Connect," Heritage Science (Springer Nature), Vol. 10, No. 122, pp. 1–14, Dec. 2022.

[10] X. Deng and Y. Yu, "Ancient Mural Inpainting via Structure Information Guided Two-Branch Model," Journal of Cultural Heritage, Vol. 61, pp. 72–81, Jan. 2023.

[11] H. Ren, K. Sun, F. Zhao, and X. Zhu, "Dunhuang Murals Image Restoration Method Based on Generative Adversarial Network," Heritage Science (Springer Nature), Vol. 12, No. 23, pp. 1–12, Feb. 2024.

[12] R. Akama, "Digitization of Classical Japanese Books by Ritsumeikan University Art Research Center: On the ARC International Model," Proc. International Conference on Digital Humanities (CODH), pp. 1–6, 2015.

[13] X. Yue, H. Li, Y. Fujikawa, and L. Meng, "Dynamic Dataset Augmentation for Deep Learning-based Oracle Bone Inscriptions Recognition," ACM Journal on Computing and Cultural Heritage (ACM), Vol. 15, No. 4, pp. 1–20, Jul. 2022.

[14] B. Lyu, X. Yue, and L. Meng, "Japanese Literature Organization and Spatiotemporal Database System Creation for Natural Disaster Analysis," Heritage Science (Springer Nature), Vol. 12, No. 194, pp. 1–15, Jun. 2024.

[15] L. Meng, "Two-Stage Recognition for Oracle Bone Inscriptions," Proc. International Conference on Image Analysis and Processing (ICIAP), pp. 672–682, Sep. 2017.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," Advances in Neural Information Processing Systems (NeurIPS), Vol. 27, pp. 2672–2680, Dec. 2014.

[17] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," Advances in Neural Information Processing Systems (NeurIPS), Vol. 33, pp. 6840–6851, Dec. 2020.

[18] B. Kawar, M. Elad, S. Ermon, and J. Song, "Denoising Diffusion Restoration Models," Advances in Neural Information Processing Systems (NeurIPS), Vol. 35, pp. 23190–23204, Dec. 2022.

[19] Huynh-Thu, Q. & Ghanbari, M. Scope of validity of PSNR in image/video quality assessment. *Electron. Lett.* **44**, 800–801 (2008).

[20] Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).

[21] M. Diem and R. Sablatnig, "Recognizing Characters of Ancient Manuscripts," Proc. 12th International Conference on Document Analysis and Recognition (ICDAR), pp. 471–475, Jul. 2010.

[22] K. Saddami, K. Munadi, Y. Away, and F. Arnia, "Effective and Fast Binarization Method for Combined Degradation on Ancient Documents," Journal of Physics: Conference Series (IOP Publishing), Vol. 1235, No. 1, Article 012021, pp. 1–8, 2019.

[23] M. Almeida, R. D. Lins, R. Bernardino, D. Jesus, and B. Lima, "A New Binarization Algorithm for Historical Documents," Proc. ACM Symposium on Document Engineering (DocEng), pp. 25–34, Aug. 2018.

[24] J. He, Q. D. M. Do, A. C. Downton, and J. Kim, "A Comparison of Binarization Methods for Historical Archive Documents," Proc. Eighth Int. Conf. on Document Analysis and Recognition (ICDAR), Vol. 1, pp. 538–542, Aug. 2005.

[25] V. Supaja, S. N. Afreen, P. Thanmai, and S. Varsha, "Restoration of Ancient Document Images Using Phase Based Binarization," Proc. Int. Conf. on Intelligent Systems and Sustainable Computing (Springer), pp. 145–154, Oct. 2022.

[26] A. Sulaiman, K. Omar, and M. F. Nasrudin, "Degraded Historical Document Binarization: A Review on Issues, Challenges, Techniques, and Future Directions," Journal of Theoretical and Applied Information Technology, Vol. 97, No. 3, pp. 785–797, Feb. 2019.

[27] Ju, M., Chen, X., & Wang, T. (2024). A Three-Stage GAN-Based Binarization Framework for Historical Documents. *Pattern Recognition*.

[28] Zhang, W., Li, H., & Nguyen, T. (2024). DocRes: A Generalist Document Restoration Model with Dynamic Visual Prompts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[29] P. Jeevan, D. S. Kumar, and A. Sethi, "WavePaint: Resource-efficient Token-mixer for Self-supervised Inpainting," Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV) Workshops, pp. 456–465, Oct. 2023.

[30] H. Liu, Z. Wan, W. Huang, Y. Song, X. Han, and J. Liao, "PD-GAN: Probabilistic Diverse GAN for Image Inpainting," Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 9371–9380, Jun. 2021.

[31] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 5967–5976, Jul. 2017.

[32] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," Proc. IEEE Int. Conf. on Computer Vision (ICCV), pp. 2242–2251, Oct. 2017.

[33] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 4401–4410, 2019.

[34] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic Image Synthesis with Spatially-Adaptive Normalization," Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 2337–2346, 2019.
[35] H. Kaneko, R. Ishibashi, and L. Meng, "Deteriorated Characters Restoration for Early Japanese Books Using Enhanced CycleGAN," Heritage (MDPI), Vol. 6, No. 5, Article 230, pp. 1–15, May 2023.
[36] V. Bruni, A. Neri, and D. Vitulano, "Context-based defading of archive photographs," EURASIP Journal on Image and Video Processing, Vol. 2010, No. 1, Article 620379, pp. 1–12, 2010.
[37] G. Tejashwini and K. S. Rani, "Restoration of Old Documents That Suffer From Degradation," International Research Journal of Engineering and Technology (IRJET), Vol. 4, No. 7, pp. 1622–1626, 2017.
[38] Center for Open Data in the Humanities (CODH), "Kuzushiji Dataset (National Institute of Japanese Literature)," CODH Open Data Repository, 2024. [Online]. Available: http://codh.rois.ac.jp/char-shape/book/

YUYA YOSHIZU received his B.E. degree in Electronic and Computer Engineering from Ritsumeikan University (RU), Shiga, Japan, in 2024. He is currently pursuing his M.E. degree at the Graduate School of Science and Engineering, RU. His research focuses on AI-based restoration of ancient Japanese manuscripts. His research interests include diffusion models, document restoration, computer vision, and cultural heritage preservation.

HAYATA KANEKO (Student Member, IEEE) received the M.E. degree from the Graduate School of Science and Engineering, RU, Kusatsu, Japan, in 2024, where he is currently pursuing the Ph.D. degree. He is also a Research Fellow of the Japan Society for the Promotion of Science (JSPS DC1). His research interests include deep learning, computer vision, and computer architecture.

RYUTO ISHIBASHI (Student Member, IEEE) received the B.E. degree in Electronic and Computer Engineering from the College of Science and Engineering, RU, Kusatsu, Japan, in 2023, and the M.E. degree in Electronic Systems from the Graduate School of Science and Engineering, RU, in 2025. He is currently pursuing a Ph.D degree in the Department of Electronic Systems at the Graduate School of Science and Engineering, RU. His research interests include deep learning and computer vision, with a particular focus on the design of lightweight AI models.

MINGCONG DENG (IEEE Fellow) received his Ph.D in Systems Science from Kumamoto University, Japan, in 1997. From 1997 to 2010, he was with Kumamoto University; University of Exeter, UK; NTT Communication Science Laboratories; Okayama University. From 2010, he has been with Tokyo University of Agriculture and Technology, Japan, as a professor. Now he is the Chair of Department of Electrical and Electronic Engineering. Prof. Deng specializes in three complementary areas: Operator based nonlinear fault detection and fault tolerant control design; System design on human factor based robot control; Learning based nonlinear control. Prof. Deng has over 650 publications including 220 journal papers in peer reviewed journals, including IEEE Press and other top-tier outlets. He serves as a chief editor for two international journals, and associate editors of 6 international journals. Prof. Deng is a co-chair of agricultural robotics and automation technical committee, IEEE Robotics and Automation Society; Also a chair of the environmental sensing, networking, and decision making technical committee, IEEE SMC Society. He was the recipient of 2014 & 2019 Meritorious Services Award of IEEE SMC Society, 2020 IEEE RAS Most Active Technical Committee Award (IEEE RAS Society), and 2024 IEEE Most Active SMC Technical Committee Award (IEEE SMC Society). He is a fellow of The Engineering Academy of Japan.

LIN MENG (Senior Member, IEEE) received his Ph.D. degree from the Graduate School of Science and Engineering, RU, Kusatsu, Japan, in 2012. He served as a Research Associate (2011–2012), Assistant Professor (2013–2017), and Lecturer (2018) in the Department of Electronic and Computer Engineering at RU. In 2015, he was a Visiting Scholar with the Department of Computer Science and Engineering, University of Minnesota, Twin Cities, Minneapolis, MN, USA. He is currently a Professor with the College of Science and Engineering, RU. He has published approximately 400 academic papers, including around 100 journal articles. His research interests include computer architecture, parallel and high-performance computing, the Internet of Things (IoT), artificial intelligence, and cultural heritage preservation. He is also a member of ACM, IPSJ, IEICE, and IEE.