EpiAgent: An Agent-Centric System for Ancient Inscription Restoration

Shipeng Zhu[1,2], Ang Chen[1,2], Na Nie[4,5], Pengfei Fang[1,2], Min-Ling Zhang[1,3], Hui Xue[1,2*]
[1]School of Computer Science and Engineering, Southeast University, China
[2]Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), Ministry of Education, China
[3]Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, China
[4]Nanjing University Museum, Nanjing University, China
[5]The China Centre for Linguistic and Strategic Studies, Nanjing University, China
{shipengzhu,chenang121}@seu.edu.cn , niena@nju.edu.cn , {fangpengfei,zhangml,hxue}@seu.edu.cn

Abstract

Ancient inscriptions, as repositories of cultural memory, have suffered from centuries of environmental and human-induced degradation. Restoring their intertwined visual and textual integrity poses one of the most demanding challenges in digital heritage preservation. However, existing AI-based approaches often rely on rigid pipelines, struggling to generalize across such complex and heterogeneous real-world degradations. Inspired by the skill-coordinated workflow of human epigraphers, we propose EpiAgent, an agent-centric system that formulates inscription restoration as a hierarchical planning problem. Following an Observe–Conceive–Execute–Reevaluate paradigm, an LLM-based central planner orchestrates collaboration among multimodal analysis, historical experience, specialized restoration tools, and iterative self-refinement. This agent-centric coordination enables a flexible and adaptive restoration process beyond conventional single-pass methods. Across real-world degraded inscriptions, EpiAgent achieves superior restoration quality and stronger generalization compared to existing methods. Our work marks an important step toward expert-level agent-driven restoration of cultural heritage. The code is available at https://github.com/blackprotoss/EpiAgent.

1. Introduction

From antiquity to the present, civilizations have inscribed their characters onto diverse materials to preserve information across time and space. Among these practices, ancient inscriptions, whether carved in stone or preserved as paper rubbings [25], encode an invaluable dual heritage: the irreplaceable textual records and the artistic essence of historical calligraphy from human civilization. Yet, these cultural artifacts are constantly threatened by environmental decay, material damage, and human intervention. The resulting coupled degradations pose a uniquely challenging problem: recovering both semantic integrity and glyph morphology under heterogeneous degradation patterns. Consequently, inscription restoration represents a critical frontier in digital humanities. Success in this domain would not only recover lost knowledge but also pioneer new paradigms for the intelligent preservation of documentary heritage [21].

Historically, trained over decades in script typology and artifact conservation, human epigraphers have been central to reconstructing both the textual content and visual form of ancient inscriptions [17]. However, the sheer volume of extant materials and the pace of ongoing degradation far exceed expert capacity. In response, recent AI-driven approaches

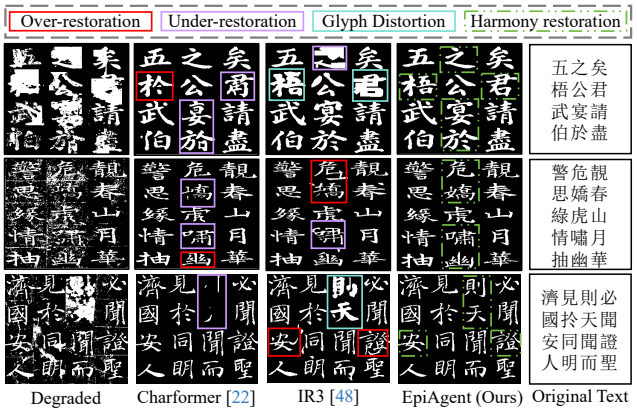

Figure 1. Illustration of the restored ancient inscription samples.

*Corresponding author

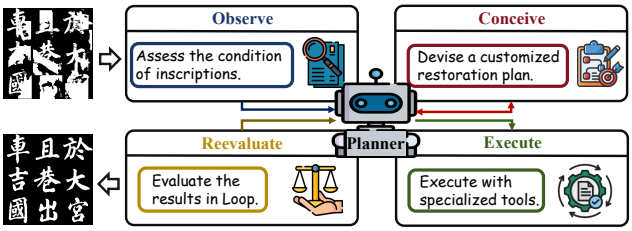

Figure 2. Illustration of the EpiAgent framework, which mimics the restoration workflow of human epigraphers.

have emerged to automate visual restoration [24]. In particular, most efforts have focused on single-character restoration [7, 22, 45]. These approaches do not scale to inscription-level cases, where degradation is spatially coupled and can fully obscure glyph sequences across a tablet. More recently, researchers have attempted full-inscription restoration using fixed pipelines with predefined workflows [44, 48]. However, this one-size-fits-all strategy lacks adaptability to heterogeneous degradation patterns. At a fundamental level, these methods are rooted in the image-to-image transfer paradigm, where the direct mapping from degraded to pristine characters often distorts the original glyph. As shown in Fig. 1, the resulting over-/under- restorations leave a critical gap between automated techniques and authentic restoration.

Therefore, we revisit the complete workflow of human epigraphers, where fine-grained analysis, specialized skills, and aesthetic judgment are orchestrated into a coherent reasoning process. Pursuing agentic systems that learn from expert behavior is thus both natural and necessary for preserving *textual authenticity* and *visual fidelity*. However, existing frameworks largely identify degradations and invoke generic tools [4, 46]. By contrast, epigraphic restoration demands a hierarchical process along three dimensions:

(1) *Multi-modal Analysis under Complex Degradation.* Inscriptions exhibit spatially varying, structurally entangled, and multi-scale degradations. The restoration system must therefore conduct multi-modal analysis: localizing characters, assessing precise visual damage, deciphering corrupted text, and reconciling results with the historical corpus. These requirements exceed the single degradation-aware schemes used for natural images. (2) *Adaptive Planning of Specialized Tools.* Inscription restoration seeks visual–textual harmony rather than isolated enhancement. This calls for a flexible and composable set of specialized tools, which can operate individually or be dynamically composed into task-specific routines. The system must then weigh evidence from text and appearance to invoke its toolkit, navigating the trade-off between textual authenticity and visual fidelity. Such state-dependent planning is incompatible with fixed pipelines. (3) *Multi-perspective Evaluation for Self-Refinement.* Authentic restoration requires judging not only pixel quality but also textual accuracy and aesthetic consistency, often with third-party expert review. Incorporating such perspectives into the decision loop enables iterative replanning, progressively steering the system toward expert-aligned epigraphic deliberation. This level of refinement is beyond the reach of single-pass or quality–centric pipelines.

Considering these challenges, we propose **EpiAgent**, an agent-centric system for ancient inscription restoration. At its core lies an LLM-based **Central Planner** that integrates both generalist and specialist analysis–restoration skills, accumulated restoration experience, and multi-perspective self-reflection. Operating within an *Observe–Conceive–Execute–Reevaluate* loop, the planner mirrors the collaborative workflow of human epigraphers and drives hierarchical closed-loop decision making throughout the process, as shown in Fig. 3. In the **Observe** stage, the planner collects multimodal signals from subordinate generalist–specialist hybrid modules and a historical corpus to establish a structured understanding of the inscription. Then, the **Conceive** stage fuses these cues with experience distilled from previous executions, thereby devising a customized restoration plan adaptively. During **Execution**, the planner governs a specialized modular toolkit that can be invoked individually or in combination to address complex and coupled degradations. Finally, in the **Reevaluate** stage, it closes the loop by evaluating the restored result using automatic metrics and optional expert feedback, updating its plan for subsequent iterations. Extensive experiments demonstrate that EpiAgent effectively handles real-world inscription degradation, achieving notable improvements in both textual authenticity and visual fidelity.

In a nutshell, the contributions are as follows:

- We introduce EpiAgent, a pioneering agent-centric system that formalizes the workflow of epigraphers within a unified Observe–Conceive–Execute–Reevaluate paradigm. An LLM-based central planner integrates multi-modal analysis, Specialized tools, and multi-perspective evaluation, enabling hierarchical closed-loop refinement.
- We decompose inscription restoration into atomic multi-modal operations. This allows our planner to dynamically assemble and schedule a specialized toolkit based on contextual analysis of degradation patterns and historical metadata, addressing complex coupled failures beyond the reach of static pipelines.
- Extensive experiments and ablation studies demonstrate the superior performance of EpiAgent over existing methods. Our work yields concrete insights for developing expert-level AI systems in cultural heritage preservation.

2. Related Work

2.1. Unified Image Restoration

Modern image restoration has shifted toward unified architectures for coupled degradations in natural images [12]. Early

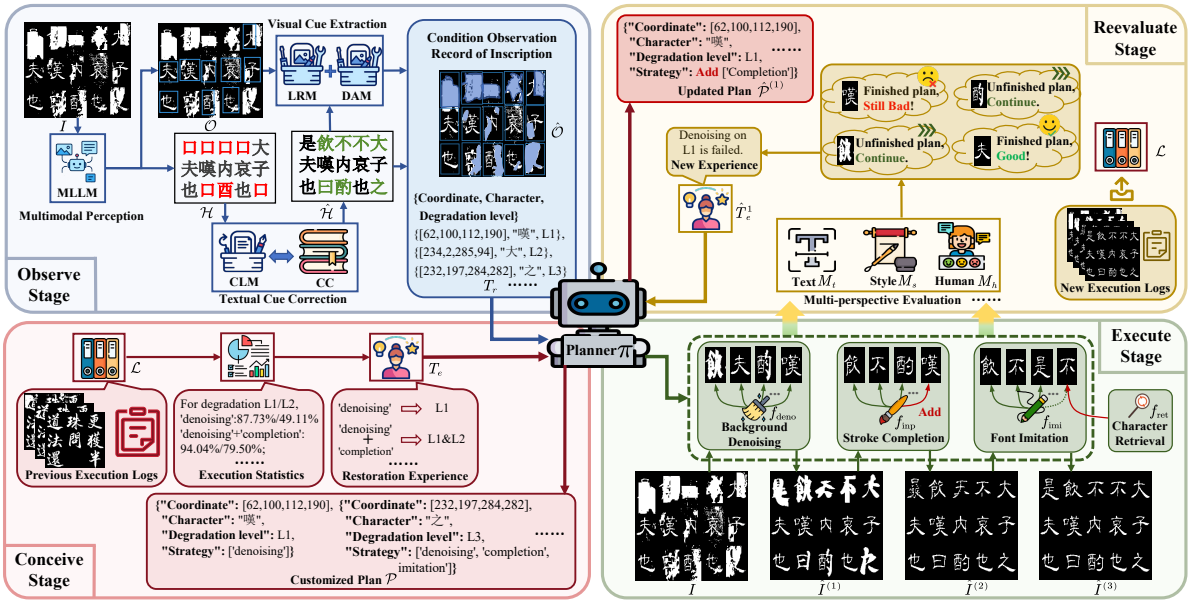

Figure 3. Illustration of the workflow of EpiAgent. The "MLLM", "LRM", "DAM", "CLM", and "CC" denote Multimodal Large Language Model, Layout Rectification Module, Degradation Assessment Model, Corrective Language Model, and Chinese Corpora, respectively.

backbones [9, 39] enabled unified modeling but still relied on manual degradation identification. Subsequent work moved toward universal restoration, where a single model adapts to diverse patterns. For example, PromptIR [20] injects degradation-aware prompts, while MoCE-IR [38] routes inputs via mixture-of-experts. With the rise of MLLMs, a new agentic paradigm has emerged. These methods leverage MLLMs to perceive degradation and dynamically invoke pre-trained restoration tools [4, 13, 15, 46]. Despite their flexibility, such general-purpose systems remain constrained in domain-specific contexts [6]. In ancient inscription restoration, this limitation becomes fundamental. Wherein, general degradation perception modules cannot capture the linguistic and stylistic nuances of ancient scripts, while off-the-shelf tools cannot preserve calligraphic authenticity. These challenges highlight the need for a domain-tailored framework.

2.2. Text Image Enhancement

Text images are inherently multimodal, with their semantics closely tied to visual structure [33]. In real-world scenarios, these images often suffer from multi-causal degradations [23]. Researchers thus explore restoration at multiple granularities. At the character level, CNN/Transformer models reconstruct the fine-grained strokes of individual characters [22, 30]. At the word level, methods typically incorporate structural priors to map degraded words to their clean counterparts [26, 47]. More recently, the focus has expanded to the document level, progressing toward unified restoration frameworks. For instance, DocDiff [36] employs frequency-domain priors for multi-task restoration, while DocRes [40] leverages a Restormer [39] to iteratively handle diverse degradations. However, most existing methods rely on image-to-image translation [12], which often compromises structural fidelity and introduces glyph distortion. Such drawbacks become unacceptable for inscriptions, where calligraphic consistency is integral to authenticity.

2.3. Ancient Inscription Restoration

AI-driven ancient script restoration has become a growing focus in digital humanities [3, 5, 17, 37]. Among these, inscriptions are emerging as key research targets due to their dual nature as historical archives and calligraphic artifacts. Early research emphasized text deciphering and attribute classification [1, 2, 19], but neglected visual authenticity. This omission is a critical flaw for ideographic scripts, where semantics and morphology are inseparable [48]. Subsequent visual restoration approaches focused on character-level enhancement [16, 45], yet failed to maintain inscription-level coherence. Recent advances have extended to full-scale restoration: Duan et al. [7] introduced a context-aware approach limited to short phrases. Zhu et al. [48] proposed a global-local framework for full-inscription restoration that suffers from error propagation. In parallel, AutoHDR [44] has shown potential for damaged content prediction using LLMs, though its style-transfer backbones may cause calligraphic distortion under heterogeneous degradation.

3. EpiAgent

As illustrated in Fig. 3, EpiAgent is an *agent-centric restoration system* that operationalizes the deliberative workflow

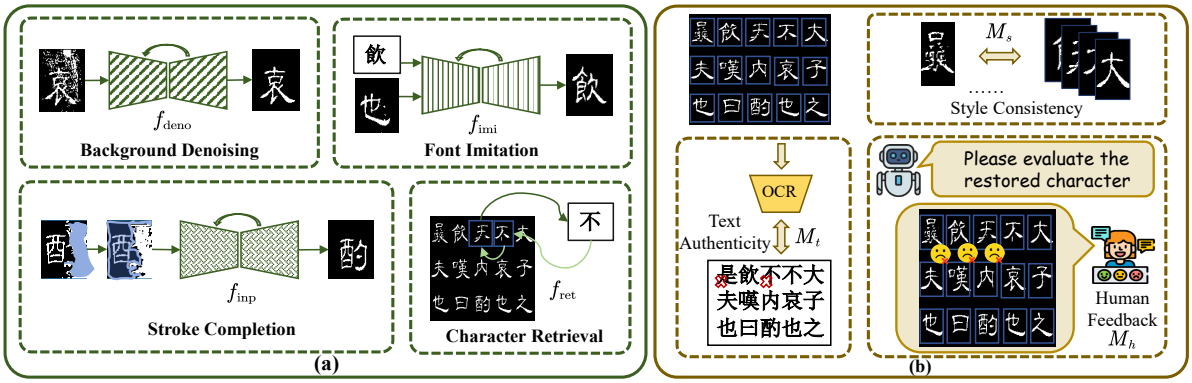

Figure 4. (a) Process of Specialized Restoration Tools; (b) Details of Multi-perspective Evaluation.

of human epigraphers. Given a degraded inscription image I affected by coupled degradation factors $\mathcal{D}$, our objective is to produce a restored output $\hat{I}$ that maximizes both textual authenticity and visual fidelity. Formally, the restoration process is governed by a **central planner** π, implemented as an agentic LLM, Kimi-K2 [28], which dynamically orchestrates analysis, reasoning, and tool invocation. Departing from one-pass pipelines, EpiAgent follows a dynamic four-stage process, *Observe-Conceive-Execute-Reevaluate*, which is executed iteratively until a stopping criterion.

Let $\mathcal{T}$ denote the restoration trajectory, comprising condition assessment T_r, experience priors T_e, execution plan P, and evaluation metrics $\mathcal{M}$. The central planner π operates on $\mathcal{T}$ to generate action sequences $\mathcal{F}$ adaptively. This closed-loop orchestration enables EpiAgent to evolve from reactive execution toward expert-informed deliberation by coupling procedural reasoning with domain-specific expertise.

3.1. Observe Stage

An effective restoration plan requires a fine-grained assessment of the textual content, calligraphic style, and degradation patterns in inscription I. Accordingly, the Observe stage builds a comprehensive record T_r via a two-step scheme:
Step 1: General multimodal perception. An MLLM [8] produces an initial layout $\mathcal{O}$ and textual hypotheses $\mathcal{H}$.
Step 2: Specialized visual and textual refinement. (a) Text Cue Correction. A Corrective Language Model (CLM), fine-tuned [11] on a 7B LLM [29] and equipped with Retrieval-Augmented Generation (RAG) [34], queries a large-scale Chinese corpus to produce the corrected reading $\hat{\mathcal{H}}$. Notably, the system allows human experts to verify $\hat{\mathcal{L}}$ for authentic correction. **(b) Visual Cue Extraction.** First, a Layout Rectification Module (LRM) consumes $\mathcal{O}$ and $\hat{\mathcal{H}}$ to predict a rectified layout $\hat{\mathcal{O}}$ that can explicitly account for fully missing or occluded regions. Second, a Degradation Assessment Module (DAM) delineates pixel-level degradation segmentation masks $\mathcal{S}_d$ and assigns discrete severity levels (none, slight, middle, severe) as $s \in \{0, 1, 2, 3\}$. Therefore, we formalize the observation record as: $T_r = \langle I, \mathcal{S}_d, s, \hat{\mathcal{H}}, \hat{\mathcal{O}} \rangle$.

3.2. Conceive Stage

Having observed the inscription status, the agent must translate this assessment T_r into an actionable plan $\mathcal{P}$. That is, it must decide how to select and sequence restoration tools for each character $c \in \mathcal{C}$, where $\mathcal{C}$ denotes the set of visual characters in the corrected reading $\hat{\mathcal{H}}$. Notably, EpiAgent plans at the level of fine-grained character restoration. Reliable layout and text predictions from the **Observe** stage isolate character regions, allowing global removal of cross-character background noise and leaving only local residuals for specialized character-level tools. Rather than relying on trial-and-error, the planner π should exploit previous restoration experience T_e to make adaptive decisions.

Concretely, T_e is distilled from historical execution logs $\mathcal{L}$. We mine $\mathcal{L}$ to extract statistical priors that map degradation patterns $\mathcal{S}_d$ to tool-efficacy distributions $p(f \mid \mathcal{S}_d)$, where $f \in \mathcal{F}$ denotes a restoration tool. For each character $c \in \mathcal{C}$, the planner π conditions on the concatenated input $[T_r; T_e]$ and produces an individual action sequence:

$$P_c = \pi(T_r, T_e, c) = \left(f_1^{(c)}, f_2^{(c)}, ..., f_{N_c}^{(c)}\right), \quad (1)$$

where each $f_i^{(c)}$ is a selected restoration tool for c, and N_c denotes the character-dependent length of the sequence; the index i specifies the execution order for c. The overall plan for inscription I is then given by $\mathcal{P} = \{P_c\}_{c \in \mathcal{C}}$.

3.3. Execute Stage

Given the complexity of coupled degradations, a single monolithic restorer is prone to under- or over-restoration or glyph distortion. Following the practice of human epigraphers, we instead factor restoration into four specialized composable tools that can be invoked independently or assembled on demand. Specifically, the Execute stage instantiates plan $\mathcal{P}$ via a composable toolkit $\mathcal{F}$. As shown in Fig. 4(a), $\mathcal{F}$ comprises three diffusion-based tools [10]:

Table 1. Inscription image restoration results on Testing Set S, R-I, and R-II. Comparison with state-of-the-art methods. The best and the second-best results are **highlighted** and underlined.

(a) Inscription Image Restoration on Testing Set S

Method / Metric	Quality							Recognition			End-to-End
	PSNR ↑	SSIM ↑	LPIPS ↓	CLIP-IQA ↑	MUSIQ ↑	MANIQA ↑	NIMA ↑	Top-1 Acc. ↑	Top-5 Acc. ↑	Macro Acc. ↑	1-NED ↑
CharFormer [22]	19.74	0.9503	0.0478	0.8763	52.77	0.4339	0.5547	0.9109	0.9533	0.5549	0.8313
DocDiff [36]	20.61	0.9565	0.0361	0.8962	53.31	0.4444	0.5559	0.9275	0.9622	0.5697	0.8439
GSDM [47]	20.37	0.9495	0.0390	0.8933	53.13	0.4422	0.5550	0.8948	0.9414	0.5368	0.8093
Restormer [39]	18.90	0.9390	0.0667	0.8891	52.82	0.4391	0.5509	0.8097	0.8824	0.4454	0.7523
MambaIR [9]	21.10	0.9599	0.0377	0.8923	53.22	0.4446	0.5556	0.9093	0.9492	0.5561	0.8251
PromptIR [20]	19.30	0.9464	0.0551	0.8882	52.95	0.4390	0.5542	0.8601	0.9214	0.4970	0.7741
MoCE-IR [38]	19.39	0.9462	0.0473	0.8955	53.24	0.4399	0.5553	0.8147	0.8929	0.4562	0.7526
IR3 [48]	21.15	0.9540	0.0388	0.8987	53.35	0.4429	0.5547	0.9626	0.9846	0.6459	0.8855
EpiAgent (Ours)	**22.14**	**0.9684**	**0.0254**	**0.9004**	**53.98**	**0.4553**	**0.5576**	**0.9889**	**0.9942**	**0.6877**	**0.9069**
Intact	-	-	-	-	-	-	-	0.9971	0.9996	0.7064	0.9120

(b) Inscription Image Restoration on Testing Set R-I

Method / Metric	Quality				End-to-End
	CLIP-IQA ↑	MUSIQ ↑	MANIQA ↑	NIMA ↑	1-NED ↑
CharFormer [22]	0.9320	49.34	0.3993	0.5398	0.5177
DocDiff [36]	0.9375	49.58	0.4059	0.5408	0.5080
GSDM [47]	0.9330	49.43	0.4027	0.5365	0.5245
Restormer [39]	0.9256	49.02	0.4029	0.5322	0.4866
MambaIR [9]	0.9346	49.40	0.4046	0.5356	0.5205
PromptIR [20]	0.9237	48.31	0.3951	0.5273	0.4312
MoCE-IR [38]	0.9317	48.53	0.3999	0.5323	0.4260
IR3 [48]	0.9370	49.77	0.4090	0.5334	0.5539
EpiAgent (Ours)	**0.9393**	**50.29**	**0.4179**	**0.5414**	**0.5766**

(c) Inscription Image Restoration on Testing Set R-II

Method / Metric	Quality				End-to-End
	CLIP-IQA ↑	MUSIQ ↑	MANIQA ↑	NIMA ↑	1-NED ↑
CharFormer [22]	0.9277	49.14	0.4009	0.5289	0.4642
DocDiff [36]	0.9352	49.45	0.4084	0.5307	0.4619
GSDM [47]	0.9373	49.26	0.4042	0.5302	0.4789
Restormer [39]	0.9081	48.32	0.4061	0.5225	0.4307
MambaIR [9]	0.9259	48.94	0.4104	0.5272	0.4697
PromptIR [20]	0.9030	47.61	0.3968	0.5208	0.3968
MoCE-IR [38]	0.9185	48.02	0.4012	0.5282	0.3892
IR3 [48]	0.9346	49.58	0.4082	0.5298	0.5374
EpiAgent (Ours)	**0.9388**	**49.94**	**0.4157**	**0.5381**	**0.5546**

- **Background Denoising** (f_{den}): Removes surface noise while preserving stroke structures via masked diffusion conditioned on $\mathcal{S}_d$.
- **Stroke Completion** (f_{inp}): Performs targeted inpainting of missing or severely degraded regions indicated by $\mathcal{S}_d$, avoiding deformation of intact strokes.
- **Font Imitation** (f_{imi}): For heavily corrupted characters, synthesizes stylistically consistent glyphs by learning style priors from high-quality exemplars of the same stele, thus maintaining calligraphic harmony.

Additionally, a **Character Retrieval** module (f_{ret}) serves as a fallback, searching for identical characters within I to replace irreparable ones without introducing style drift.

Given the per-character sequences $P_c = \big(f_i^{(c)}\big)_{i=1}^{N_c}$, the execution process applies the scheduled operators in order. At iteration k, the restored inscription $\hat{I}^{(k)}$ is obtained by:

$$\hat{I}^{(k)}[c] \;=\; f_{N_c}^{(c)} \circ \cdots \circ f_1^{(c)}\big(\hat{I}^{(k-1)}[c]\big), \quad \forall\, c \in \mathcal{C}, \qquad (2)$$

where $\hat{I}^{(0)} = I$. The overall execution trajectory is thus governed by the sequences $\{P_c\}_{c\in\mathcal{C}}$ and the toolkit $\mathcal{F}$.

3.4. Reevaluate Stage

To ensure both textual and visual harmony, it is crucial to introduce multi-perspective metrics after each execution pass. Such signals provide principled stopping and rollback criteria. Therefore, they allow the central planner to achieve self-refinement by revising plans for under-restored characters and continuously distilling experience for future decisions. As shown in Fig. 4(b), at iteration k we evaluate each character $c \in \mathcal{C}$ on the current $\hat{I}^{(k)}$ using metrics:

- **Text Authenticity**. This metric quantifies semantic correctness by comparing OCR text to the corrected reading: $M_t^{(k)}(c) = 1 - \text{CER}\big(\text{OCR}(\hat{I}^{(k)}[c]), \hat{\mathcal{H}}[c]\big) \in [0,1]$.
- **Style Consistency**.This aesthetic metric measures calligraphic conformity to the reference style distribution: $M_s^{(k)}(c) = \text{CosSim}\big(\phi(\hat{I}^{(k)}[c]), \phi_{\text{ref}}\big) \in [0,1]$, where ϕ_{ref} is a style embedding from high-quality exemplars.
- **Human Feedback (Optional)**: This metric provides an expert acceptance signal for ambiguous or high-stakes cases, serving as a hard decision criterion and calibrating thresholds for replanning or termination: $M_h^{(k)}(c) \in \{0,1\}$.

Given thresholds $\tau_t, \tau_s \in (0,1]$ and a maximum iteration budget $K_{\max}$, the failure set at iteration k is defined as:

$$\mathcal{F}^{(k)} = \Big\{ c \in \mathcal{C} \;\Big|\; \begin{array}{l} (M_h^{(k)}(c) = 0) \vee \\ (M_t^{(k)}(c) < \tau_t) \;\vee\; (M_s^{(k)}(c) < \tau_s) \end{array} \Big\}. \qquad (3)$$

If $\mathcal{F}^{(k)} \neq \varnothing$ and $k < K_{\max}$, the planner π uses $\mathcal{F}^{(k)}$, together with T_r and T_e, to generate a revised plan $\mathcal{P}^{(k+1)}$ that focuses on the failed characters; otherwise, the process terminates. Finally, we update the execution logs $\mathcal{L}$ to refine the empirical priors T_e. Thus, the strategy is dynamically updated during inference rather than fixed by a static prior.

4. Experiments

4.1. Evaluation Protocol

We evaluate EpiAgent on the Chinese Inscription Rubbing Images (CIRI) dataset [48], which comprises a wide range of real inscription rubbings featuring diverse calligraphic styles, complex character structures, and compound degradation pat-

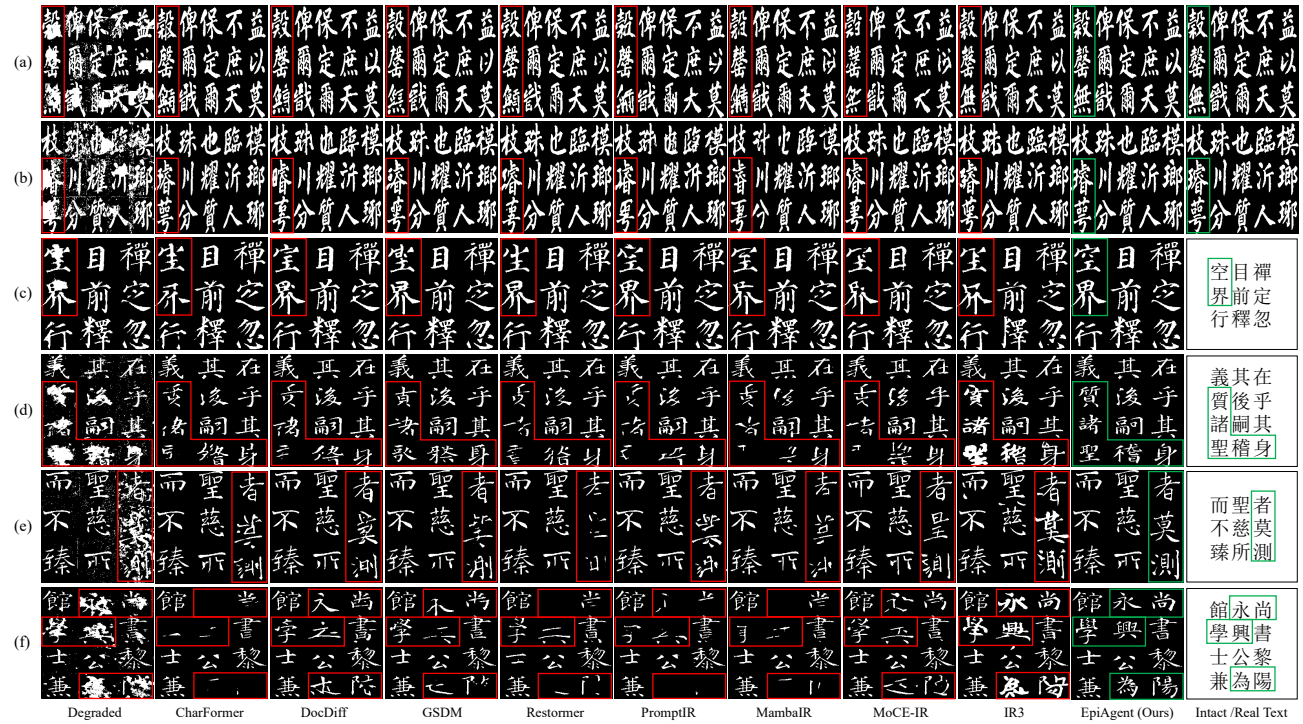

Figure 5. Restoration results of different methods on degraded inscription images. (a)-(b) are from Testing Set S, (c)-(d) belong to Testing Set R-I, and (e)-(f) belong to Testing Set R-II. The red borders denote the degraded patches and the restored patches by competing methods, while the green counterparts denote the ground-truth text and the restored patches by EpiAgent.

terns. CIRI contains 24k synthetic inscription images (20K for training, 4K for testing) and 2k real rubbings split into two test sets. Type I (R-I) includes images whose degradation patterns are partially reused as sources for synthesizing defects in the synthetic subset, while Type II (R-II) consists of fragments with entirely unseen degradation that remain unseen during synthesis and training. For comprehensive assessment, we adopt seven image quality metrics to evaluate the visual appearance of reconstructed characters: three full-reference metrics (PSNR, SSIM [32], LPIPS [41]) and four no-reference metrics (CLIP-IQA [31], MUSIQ [14], MANIQA [35], NIMA [27]), with particular emphasis on image aesthetics and stylistic fidelity. To quantify glyph fidelity and textual authenticity, we report Top-1 Accuracy, Top-5 Accuracy, and Macro Accuracy [43] for character-level recognition, as well as 1-NED [42] for image-level text similarity between predicted and ground-truth transcriptions.

4.2. Comparison with State-of-the-Art Methods

We compare EpiAgent against a series of state-of-the-art open-source methods from related areas. These include Restormer [39], MambaIR [9], PromptIR [20], and MoCE-IR [38] from unified image restoration; CharFormer [22], GSDM [47], and DocDiff [36] from text image enhancement; a tailored baseline [48] (denoted as IR3) from inscription restoration. All comparison methods are trained on the CIRI training set using their official implementations and default hyperparameters, and their performance is evaluated on the three CIRI test splits. It is worth noting that existing agentic frameworks [4, 46] for natural images cannot be directly compared, since their general perception modules and restoration tools cannot be readily transferred to inscription restoration without substantial redesign. In addition, recent tailored methods [37, 44] are difficult to reproduce faithfully due to the unavailability of some components.

Table 2. User study of inscription image restoration results. The best and the second best results are **highlighted** and underlined.

Method / Metric	User Study (%) ↑		
	Top-1 Ranking	Top-3 Ranking	Mean Ranking [18]
Charformer [22]	3.72	26.57	52.39
DocDiff [36]	8.38	46.63	63.46
GSDM [47]	5.75	39.55	60.70
Restormer [39]	0.89	9.20	28.10
MambaIR [9]	3.88	32.74	57.55
PromptIR [20]	0.63	7.75	20.96
MoCE-IR [38]	1.52	11.33	36.62
IR3 [48]	<u>15.60</u>	<u>51.14</u>	<u>67.41</u>
EpiAgent (Ours)	**59.66**	**84.18**	**82.11**

We conduct quantitative experiments on the three CIRI test sets to evaluate both the visual fidelity and textual authenticity of images restored by different methods, with results

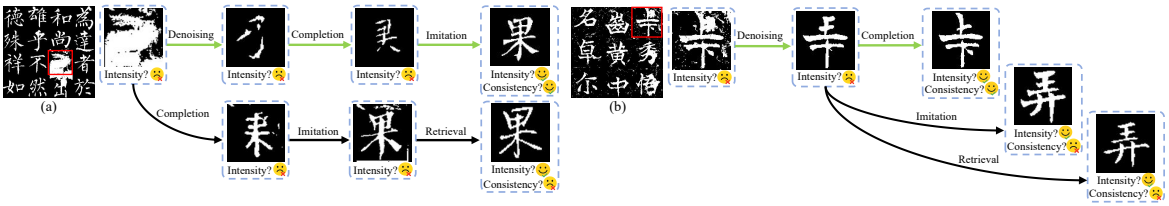

Figure 6. Exemplary comparison between different tool invocation sequences faced with (a) severely degraded (L3) and (b) slightly degraded (L1) character blocks. The green lines mark the optimal restoring sequence.

summarized in Tab. 1 and Fig. 5. In parallel, we perform a user study involving over twenty epigraphers and scholars who provided professional subjective assessments of restoration quality, as reported in Tab. 2.

The results in Tab. 1 show that EpiAgent achieves consistently superior performance on inscription restoration. Across all metrics on the synthetic split S and the real splits R-I and R-II, EpiAgent surpasses all competing methods, demonstrating the advantages of an agent-centric restoration strategy in both recovering the visual appearance of damaged regions and preserving textual authenticity. The user study in Tab. 2 further demonstrates that human experts overwhelmingly prefer EpiAgent over all baselines. Together, these results indicate that EpiAgent delivers SOTA performance under both metric- and expert-centered evaluation.

Qualitative comparisons in Fig. 5 provide additional insights: (1) EpiAgent maintains high restoration quality under both minor degradation (e.g., the region highlighted in Fig. 5(c)) and severe degradation (e.g., Fig. 5(d)). This robustness arises from degradation-aware and experience-guided planning and execution, which enable flexible deployment of appropriate tools for specific degradation patterns; (2) EpiAgent achieves strong style perception and alignment, both when completing lightly damaged characters (Fig. 5(c)) and when fully reconstructing missing characters in large degraded areas (Fig. 5(f)). This preservation of calligraphic consistency is crucial for aesthetically convincing inscription restoration; (3) EpiAgent demonstrates robust generalization across synthetic and real inscription images. While methods such as CharFormer [22] and IR3 [48] perform competitively on synthetic test images (Fig. 5(a),(b)), their performance drops markedly on real inscriptions.

4.3. Ablation Study

4.3.1. The Impact of Multimodal Analysis

Multimodal analysis in the **Observe** stage is critical to EpiAgent, as it provides a comprehensive assessment of degraded inscriptions on which all downstream planning depends. We therefore ablate the multimodal analysis modules and study their impact on overall restoration performance. The evaluated modules include: (i) an MLLM for general perception, (ii) a fine-tuned Corrective Language Model (CLM) for text correction, and (iii) a Retrieval-Augmented Generation (RAG) module over a Chinese corpus. Tab. 3 reports end-to-end spotting performance under different combinations. The results reveal three main observations: (1) a standalone MLLM struggles to handle the complex visual–textual cues and coupled degradations of inscriptions; (2) correcting the recognized script with a specialized CLM yields a clear gain in 1-NED, highlighting the importance of linguistic priors for textual authenticity; (3) adding RAG achieves the best performance, indicating that corpus-level retrieval further improves robustness and generalization.

Table 3. Ablation studies of the analysis modules used in the Observation stage. The best and the second-best results are **highlighted** and underlined.

Analysis Module			1-NED ↑		
MLLM	CLM	RAG	Set S	Set R-I	Set R-II
✓	✗	✗	0.6481	0.5509	0.4336
✓	✓	✗	0.9211	0.8759	0.8486
✓	✓	✓	**0.9742**	**0.9694**	**0.9606**

Table 4. Ablation study of different planning strategies for inscription restoration. "Random" refers to random tool invocation from the toolkit. "Fixed" denotes predefined Scheme A (Denoising-Completion) and Scheme B (Denoising-Completion-Imitation). "Experience-guided" customizes the restoration scheme based on distilled experience priors. The best and the second best results are **highlighted** and underlined.

Strategy / Metric	Quality			End-to-End
	PSNR ↑	SSIM ↑	LPIPS ↓	1-NED ↑
Random	18.53	0.9078	0.0869	0.7702
Fixed (Scheme A)	21.19	0.9605	0.0371	0.8814
Fixed (Scheme B)	20.78	0.9526	0.0401	0.8935
Experience-guided (Ours)	**22.14**	**0.9684**	**0.0254**	**0.9069**

4.3.2. The Effect of Adaptive Planning

We next evaluate how adaptive planning in the **Conceive** stage affects restoration quality. As illustrated in Fig. 6, improper tool ordering can severely degrade results, e.g., skipping background denoising and directly applying stroke completion compromises glyph integrity and style consistency (Fig. 6(a)), while inappropriate use of completion and imita-

Table 5. Ablation studies of multi-perspective evaluation module. The best and the second best results are **highlighted** and underlined.

Automatic Metric		Human Feedback	Quality							End-to-End
Text Authenticity	Style Consistency	Score of Human Experts	PSNR ↑	SSIM ↑	LPIPS ↓	CLIP-IQA ↑	MUSIQ ↑	MANIQA ↑	NIMA ↑	1-NED ↑
✗	✗	✗	21.48	0.9593	0.0305	0.8976	53.63	0.4454	0.5549	0.8969
✓	✗	✗	21.59	0.9616	0.0292	0.8974	53.66	0.4465	0.5552	0.9026
✗	✓	✗	21.73	0.9632	0.0284	0.8979	53.70	0.4477	0.5559	0.8994
✓	✓	✗	22.02	0.9668	0.0279	0.8983	53.73	0.4492	0.5566	0.9041
✓	✓	✓	**22.14**	**0.9684**	**0.0254**	**0.9004**	**53.98**	**0.4553**	**0.5576**	**0.9069**

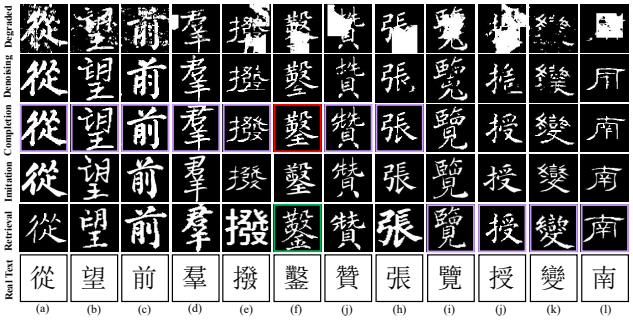

Figure 7. Comparison of restoration outcomes across different tools. (a)-(d), (e)-(h), (i)-(l) correspond slightly, middely, and severely degraded character. Red and green borders indicate the restorations preferred by EpiAgent and human experts, respectively, while purple borders indicate cases where their choices coincide.

tion leads to over-restoration with fake structures (Fig. 6(b)). These cases highlight that effective planning is non-trivial and must be guided by domain knowledge and accumulated restoration experience. To quantify this effect, we compare three planning strategies on Testing Set S: "Random", "Fixed", and our experience-guided adaptive planning based on distilled priors. As reported in Tab. 4, the experience-guided strategy achieves superior scores across all three image-quality metrics and spotting accuracy, confirming the benefit of adaptive planning for handling dual-modal cues and complex degradations in inscriptions.

4.3.3. The Impact of Specialized Restoration Toolkit

Fig. 7 provides a character-level comparison of restorations obtained with different tools in the specialized toolkit. In Fig. 7(a)–(h), the background denoising and stroke completion tools effectively handle slight to medium degradation. Specifically, denoising removes scattered noise and artifacts, while the completion tool then refines and reconstructs damaged stroke structures. However, under severe spalling or complete absence, as in Fig. 7(i)–(l), denoising and completion alone become insufficient since the limited local evidence makes inpainting ill-posed. In such cases, font imitation and character retrieval provide reliable alternatives. Nonetheless, these two tools can still produce undesirable results, as shown in Fig. 7(h), highlighting the necessity of dynamic tool combination.

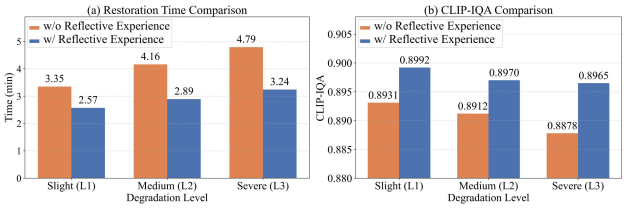

Figure 8. Quantitative comparison of restoration time and CLIP-IQA with and without reflective experience.

4.3.4. The Role of Evaluation and Refinement

We investigate the multi-perspective evaluation used in the **Reevaluate** stage. From Tab. 5, we obtain three observations: (1) adding the text authenticity metric markedly improves end-to-end spotting accuracy, indicating that target prediction deviations provide an effective signal to refine each restoration pass; (2) incorporating the style consistency metric further boosts visual quality across all seven full- and no-reference IQA measures, showing that explicitly scoring calligraphic coherence helps the agent better preserve glyph aesthetics; (3) expert review yields the highest gains, as expert-in-the-loop feedback more faithfully reflects actual restoration quality and injects valuable domain knowledge.

Fig. 8 examines the effect of disabling the refinement mechanism, i.e., evaluating restorations without recording and reusing feedback. Removing this refinement mechanism leads to a clear increase in average restoration time and a decline in quality, confirming that accumulated reflective experience is both reusable and beneficial for enhancing the inscription restoration capability of EpiAgent.

5. Conclusion

We introduced **EpiAgent**, an agent-centric system for ancient inscription restoration that operationalizes the workflow of epigraphers. An LLM-based central planner coordinates the dynamic four-stage process, enabling hierarchical closed-loop decision-making that safeguard textual authenticity and visual fidelity. Experiments and ablation studies show that EpiAgent consistently surpasses strong baselines, highlighting the effectiveness of our domain-aware design. Beyond inscriptions, our framework offers a concrete blueprint for integrating agentic AI into cultural heritage preservation.

References

[1] Yannis Assael, Thea Sommerschield, and Jonathan Prag. Restoring ancient text using deep learning: a case study on greek epigraphy. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 6368–6375, 2019. 3

[2] Yannis Assael, Thea Sommerschield, Brendan Shillingford, Mahyar Bordbar, John Pavlopoulos, Marita Chatzipanagiotou, Ion Androutsopoulos, Jonathan Prag, and Nando de Freitas. Restoring and attributing ancient texts using deep neural networks. *Nature*, 603(7900):280–283, 2022. 3

[3] Songxiao Cao, Zichao Shu, Zhipeng Xu, Dailiang Xie, and Ya Xu. Character segmentation and restoration of qin-han bamboo slips using local auto-focus thresholding method. *Multimedia Tools and Applications*, 81(6):8199–8213, 2022. 3

[4] Haoyu Chen, Wenbo Li, Jinjin Gu, Jingjing Ren, Sixiang Chen, Tian Ye, Renjing Pei, Kaiwen Zhou, Fenglong Song, and Lei Zhu. Restoreagent: Autonomous image restoration agent via multimodal large language models. In *Advances in Neural Information Processing Systems*, pages 110643–110666, 2024. 2, 3, 6

[5] Xiaolei Diao, Daqian Shi, Wei Cao, Ting Wang, Ruihua Qi, Chuntao Li, and Hao Xu. Oracle bone inscription image restoration via glyph extraction. *npj Heritage Science*, 13(1): 321, 2025. 3

[6] Zhe Dong, Zhengning Zhang, Yuzhe Sun, Haochen Jiang, Tianzhu Liu, and Yanfeng Gu. Phydae: Physics-guided degradation-adaptive experts for all-in-one remote sensing image restoration. *IEEE Transactions on Geoscience and Remote Sensing*, 64:1–18, 2026. 3

[7] Siyu Duan, Jun Wang, and Qi Su. Restoring ancient ideograph: A multimodal multitask neural network approach. In *Proceedings of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, pages 14005–14015, 2024. 2, 3

[8] Dong Guo, Faming Wu, Feida Zhu, Fuxing Leng, Guang Shi, Haobin Chen, Haoqi Fan, Jian Wang, Jianyu Jiang, Jiawei Wang, et al. Seed1. 5-vl technical report. *arXiv preprint arXiv:2505.07062*, 2025. 4

[9] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. In *Proceedings of the European Conference on Computer Vision*, pages 222–241, 2024. 3, 5, 6

[10] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 4

[11] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *Proceedings of the International Conference of Learning Representation*, 2022. 4

[12] Junjun Jiang, Zengyuan Zuo, Gang Wu, Kui Jiang, and Xianming Liu. A survey on all-in-one image restoration: Taxonomy, evaluation and future trends. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 2, 3

[13] Xu Jiang, Gehui Li, Bin Chen, and Jian Zhang. Multi-agent image restoration. *arXiv preprint arXiv:2503.09403*, 2025. 3

[14] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5148–5157, 2021. 6

[15] Bingchen Li, Xin Li, Yiting Lu, and Zhibo Chen. Hybrid agents for image restoration. *arXiv preprint arXiv:2503.10120*, 2025. 3

[16] Yunjing Liu, Erhu Zhang, Guangfeng Lin, and Jinghong Duan. A structural information-guided cross-modal method for damaged inscription inpainting via vision-language models. *npj Heritage Science*, 13(1):485, 2025. 3

[17] Mallory E Matsumoto. Archaeology and epigraphy in the digital era. *Journal of Archaeological Research*, 30(2):285–320, 2022. 1, 3

[18] Eva Ostertagova, Oskar Ostertag, and Jozef Kováč. Methodology and application of the kruskal-wallis test. *Applied Mechanics and Materials*, 611:115–120, 2014. 6

[19] Katerina Papavassileiou, Dimitrios I Kosmopoulos, and Gareth Owens. A generative model for the mycenaean linear b script and its application in infilling text from ancient tablets. *ACM Journal on Computing and Cultural Heritage*, 16(3): 1–25, 2023. 3

[20] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. *Advances in Neural Information Processing Systems*, 36:71275–71293, 2023. 3, 5, 6

[21] Erasmo Purificato, Danai Bili, Robert Jungnickel, Serra Victoria Ruiz, Josefina Faniani, Dias Abendroth, Llorca David Fernandez, and Emilia Gomez. *The role of artificial intelligence in scientific research*. Publications Office of the European Union, 2025. 1

[22] Daqian Shi, Xiaolei Diao, Lida Shi, Hao Tang, Yang Chi, Chuntao Li, and Hao Xu. Charformer: A glyph fusion based attentive framework for high-precision character image denoising. In *Proceedings of the ACM International Conference on Multimedia*, pages 1147–1155, 2022. 2, 3, 5, 6, 7

[23] Yan Shu, Weichao Zeng, Fangmin Zhao, Zeyu Chen, Zhenhang Li, Xiaomeng Yang, Yu Zhou, Paolo Rota, Xiang Bai, Lianwen Jin, et al. Visual text processing: A comprehensive review and unified evaluation. *arXiv preprint arXiv:2504.21682*, 2025. 3

[24] Thea Sommerschield, Yannis Assael, John Pavlopoulos, Vanessa Stefanak, Andrew Senior, Chris Dyer, John Bodel, Jonathan Prag, Ion Androutsopoulos, and Nando De Freitas. Machine learning for ancient languages: A survey. *Computational Linguistics*, 49(3):703–747, 2023. 2

[25] Kenneth Starr. Black tigers: A grammar of chinese rubbings. In *Black Tigers*. University of Washington Press, 2018. 1

[26] Jiande Sun, Fanfu Xue, Jing Li, Lei Zhu, Huaxiang Zhang, and Jia Zhang. Tsinit: A two-stage inpainting network for incomplete text. *IEEE Transactions on Multimedia*, 25:5166–5177, 2022. 3

[27] Hossein Talebi and Peyman Milanfar. Nima: Neural image assessment. *IEEE Transactions on Image Processing*, 27(8): 3998–4011, 2018. 6

[28] Kimi Team, Yifan Bai, Yiping Bao, Guanduo Chen, Jiahao Chen, Ningxin Chen, Ruijue Chen, Yanru Chen, Yuankun Chen, Yutian Chen, et al. Kimi k2: Open agentic intelligence. *arXiv preprint arXiv:2507.20534*, 2025. 4

[29] Qwen Team et al. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*, 2024. 4

[30] Jiahao Wang, Gang Pan, Di Sun, and Jiawan Zhang. Chinese character inpainting with contextual semantic constraints. In *Proceedings of the ACM International Conference on Multimedia*, pages 1829–1837, 2021. 3

[31] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the Annual AAAI Conference on Artificial Intelligence*, pages 2555–2563, 2023. 6

[32] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 6

[33] Baole Wei, Yuxuan Zhou, Liangcai Gao, and Zhi Tang. Glyphsr: A simple glyph-aware framework for scene text image super-resolution. In *Proceedings of the Annual AAAI Conference on Artificial Intelligence*, pages 8277–8285, 2025. 3

[34] Shitao Xiao, Zheng Liu, Peitian Zhang, and Niklas Muennighoff. C-pack: Packaged resources to advance general chinese embedding. *arXiv preprint arXiv:2309.07597*, 2023. 4

[35] Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1191–1200, 2022. 6

[36] Zongyuan Yang, Baolin Liu, Yongping Xxiong, Lan Yi, Guibin Wu, Xiaojun Tang, Ziqi Liu, Junjie Zhou, and Xing Zhang. Docdiff: Document enhancement via residual diffusion models. In *Proceedings of the ACM International Conference on Multimedia*, pages 2795–2806, 2023. 3, 5, 6

[37] Zhenhua Yang, Dezhi Peng, Yongxin Shi, Yuyi Zhang, Chongyu Liu, and Lianwen Jin. Predicting the original appearance of damaged historical documents. In *Proceedings of the Annual AAAI Conference on Artificial Intelligence*, pages 9382–9390, 2025. 3, 6

[38] Eduard Zamfir, Zongwei Wu, Nancy Mehta, Yuedong Tan, Danda Pani Paudel, Yulun Zhang, and Radu Timofte. Complexity experts are task-discriminative learners for any image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12753–12763, 2025. 3, 5, 6

[39] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022. 3, 5, 6

[40] Jiaxin Zhang, Dezhi Peng, Chongyu Liu, Peirong Zhang, and Lianwen Jin. Docres: A generalist model toward unifying document image restoration tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15654–15664, 2024. 3

[41] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018. 6

[42] Rui Zhang, Yongsheng Zhou, Qianyi Jiang, Qi Song, Nan Li, Kai Zhou, Lei Wang, Dong Wang, Minghui Liao, Mingkun Yang, et al. Icdar 2019 robust reading challenge on reading chinese text on signboard. In *Proceedings of the International Conference on Dcument Analysis and Recognition*, pages 1577–1581. IEEE, 2019. 6

[43] Yifan Zhang, Bingyi Kang, Bryan Hooi, Shuicheng Yan, and Jiashi Feng. Deep long-tailed learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10795–10816, 2023. 6

[44] Yuyi Zhang, Peirong Zhang, Zhenhua Yang, Pengyu Yan, Yongxin Shi, Pengwei Liu, Fengjun Guo, and Lianwen Jin. Reviving cultural heritage: A novel approach for comprehensive historical document restoration. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 28876–28892, 2025. 2, 3, 6

[45] Wenjun Zheng, Benpeng Su, Ruiqi Feng, Xihua Peng, and Shanxiong Chen. Ea-gan: Restoration of text in ancient chinese books based on an example attention generative adversarial network. *Heritage Science*, 2023. 2, 3

[46] Kaiwen Zhu, Jinjin Gu, Zhiyuan You, Yu Qiao, and Chao Dong. An intelligent agentic system for complex image restoration problems. In *Proceedings of the International Conference on Learning Representations*, 2025. 2, 3, 6

[47] Shipeng Zhu, Pengfei Fang, Chenjie Zhu, Zuoyan Zhao, Qiang Xu, and Hui Xue. Text image inpainting via global structure-guided diffusion models. In *Proceedings of the Annual AAAI Conference on Artificial Intelligence*, pages 7775–7783, 2024. 3, 5, 6

[48] Shipeng Zhu, Hui Xue, Na Nie, Chenjie Zhu, Haiyue Liu, and Pengfei Fang. Reproducing the past: A dataset for benchmarking inscription restoration. In *Proceedings of the ACM International Conference on Multimedia*, pages 7714—-7723, 2024. 2, 3, 5, 6, 7