

# Deciphering Oracle Bone Language with Diffusion Models

Haisu Guan<sup>1</sup>, Huanxin Yang<sup>1</sup>, Xinyu Wang<sup>2,\*</sup>, Shengwei Han<sup>3</sup>, Yongge Liu<sup>3</sup>,  
Lianwen Jin<sup>4</sup>, Xiang Bai<sup>1</sup>, Yuliang Liu<sup>1,\*</sup>

<sup>1</sup>Huazhong University of Science and Technology <sup>2</sup>The University of Adelaide

<sup>3</sup>Anyang Normal University <sup>4</sup>South China University of Technology

<sup>1</sup>{haisuguan, ylliu}@hust.edu.cn

\*Corresponding authors

## Abstract

Originating from China’s Shang Dynasty approximately 3,000 years ago, the Oracle Bone Script (OBS) is a cornerstone in the annals of linguistic history, predating many established writing systems. Despite the discovery of thousands of inscriptions, a vast expanse of OBS remains undeciphered, casting a veil of mystery over this ancient language. The emergence of modern AI technologies presents a novel frontier for OBS decipherment, challenging traditional NLP methods that rely heavily on large textual corpora, a luxury not afforded by historical languages. This paper introduces a novel approach by adopting image generation techniques, specifically through the development of Oracle Bone Script Decipherer (OBSD). Utilizing a conditional diffusion-based strategy, OBSD generates vital clues for decipherment, charting a new course for AI-assisted analysis of ancient languages. To validate its efficacy, extensive experiments were conducted on an oracle bone script dataset, with quantitative results demonstrating the effectiveness of OBSD. *Code and decipherment results will be made available at <https://github.com/guanhaisu/OBSD>.*

## 1 Introduction

Oracle Bone Script (OBS) represents an ancient language inscribed on turtle shells and animal bones, extensively utilized during China’s Shang Dynasty, a feudal dynasty dating back 3,000 years. The script not only chronicled the human geography and daily activities of that period but also encapsulates invaluable historical significance, offering a unique window into the linguistic and cultural practices of early Chinese civilization. However, despite the discovery of tens of thousands of fragments of oracle bones, a significant portion of the characters remain undeciphered (Wang and Deng, 2024), leaving the rest shrouded in mystery. To date, more than 4,500 Oracle Bone Script (OBS)

characters have been discovered, but only about 1,600 of these have been deciphered and linked to their modern Chinese counterparts. In modern Chinese, Unicode includes more than 90,000 Chinese characters, though only approximately 3,500 characters are commonly used in contemporary Chinese society. This challenge of understanding the remaining undeciphered OBS characters and linking them to modern Chinese has attracted significant research interest, with attempts being made to leverage modern AI technologies for the understanding of such an ancient language (Zhang et al., 2022; Jiang et al., 2023; Wang and Deng, 2024; Guan et al., 2024).

However, the majority of existing methodologies primarily focus on the recognition and understanding of already deciphered OBS (Guo et al., 2015; Meng et al., 2018; Zhang et al., 2019; Hu, 2023), with the utilization of AI to assist in the decipherment of unknown inscriptions remaining an underexplored area. This is partly because, unlike modern languages that can be digitized and stored as text due to established encoding systems, OBS lacks a standard input method or encoding scheme, resulting in its preservation predominantly in the form of images rather than digital text usually used in NLP methods. Additionally, since OBS was inscribed on turtle shells and animal bones, many of which have been damaged or fragmented upon discovery, there is essentially no complete corpus available. This absence of a comprehensive corpus severely limits the applicability of language models that require extensive datasets for training, such as BERT (Devlin et al., 2018), RoBERTa (Liu et al., 2019), and GPT (Brown et al., 2020).

To address the challenges inherent in the decipherment of OBS using conventional NLP methodologies, this paper introduces a novel approach by employing image-based generative techniques for auxiliary decipherment of OBS. Specifically, we train a conditional diffusion model that utilizes un-

seen categories of OBS as a conditional input to generate corresponding images of its modern counterpart. This direct provision of modern representations or potential decipherment clues leverages the model’s learned evolution from ancient scripts to contemporary fonts, circumventing the corpus construction and other challenges that traditional NLP methods face with ancient languages. Notably, while our experiments focus on OBS, this training paradigm holds the potential for extension to other ancient languages, such as Cuneiform and Hieroglyphics. In summary, this paper makes three key contributions:

- We introduce a novel approach to the task of ancient script decipherment by utilizing image generation techniques, offering a novel solution to challenges that conventional NLP methods struggle to address.
- We propose Oracle Bone Script Decipher (OBSD), a conditional diffusion model optimized for OBS decipherment. Our Localized Structural Sampling technique enhances the model’s ability to discern and interpret the intricate patterns of characters.
- OBSD demonstrates its effectiveness in decipherment through comprehensive ablation studies and benchmark comparisons. It offers a pioneering approach for AI-assisted ancient language decipherment, potentially laying a foundation for future research.

## 2 Related Works

Applying machine learning to the study of ancient languages represents a notable shift in linguistics and epigraphy. This area, distinct from the NLP tasks typically associated with modern languages, involves digitization, linguistic analysis, textual criticism, translation, and decipherment (Jin et al., 2023; Nuhn et al., 2012; Ravi and Knight, 2011). For a comprehensive overview of this field, we direct interested readers to the survey by Sommerschild et al. (Sommerschild et al., 2023; Li et al., 2020; Huang et al., 2019; Yang and Fu, 2020; Guo et al., 2015). Due to space constraints, our review is limited to literature most pertinent to oracle bone language decipherment.

The oracle bone language is considered a form of hieroglyphic that uses pictorial symbols to represent specific meanings. It originated around 1500 BC and has evolved over thousands of years into

modern Chinese characters. The evolution timeline can be summarized into seven periods as follows: Oracle Bone Script (1500 BC), Bronze Inscriptions (1300 BC - 221 BC), Seal Script (1100 BC - 221 BC), Spring & Autumn Characters (770 BC - 476 BC), Warring States Characters (475 BC - 221 BC), Clerical Script (221 BC - 220 AD) and Regular Script (around 3rd century AD). The continuous evolutionary path makes OBS a unique presence among ancient scripts. Many of its character forms have been preserved in modern standard Chinese characters. While these are significant overlaps in the forms and meanings of characters between adjacent periods, greater differences can be found between more distant periods. Some characters disappeared and later reappeared across different periods, highlighting the dynamic nature of this ancient writing system.

While the majority of work related to OBS has focused on employing CV or NLP techniques to recognize (Zhang et al., 2021a; Fu et al., 2022; Wang et al., 2022) or understand (Han et al., 2020; Qi et al., 2023; Hu, 2023) already deciphered characters, the use of AI to assist in deciphering characters with unknown meanings remains a largely unexplored and challenging task. Among these, the case-based reasoning strategy developed by Zhang et al. (Zhang et al., 2021b) stands out in its method of drawing parallels to already interpreted characters to decipher OBS. While effective to a degree, this approach is inherently constrained by its dependence on the corpus of previously deciphered characters, potentially stymieing the discovery of novel meanings. On another front, Chang et al.’s cascade generative adversarial networks framework (Chang et al., 2022) presents an innovative attempt at deciphering, yet it faces challenges due to evolutionary gaps in OBS and the completeness of training data. These challenges arise because when a character disappears for a specific period, the evolutionary path relied upon by such methods no longer remains intact, significantly impacting the success rate of deciphering and restricting their effectiveness to small datasets with clear evolutionary paths.

## 3 Method

### 3.1 Preliminary

In this study, we focus on the task of OBS decipherment, aiming to predict the corresponding modern Chinese character forms for the oracle bone language. This endeavor not only seeks to

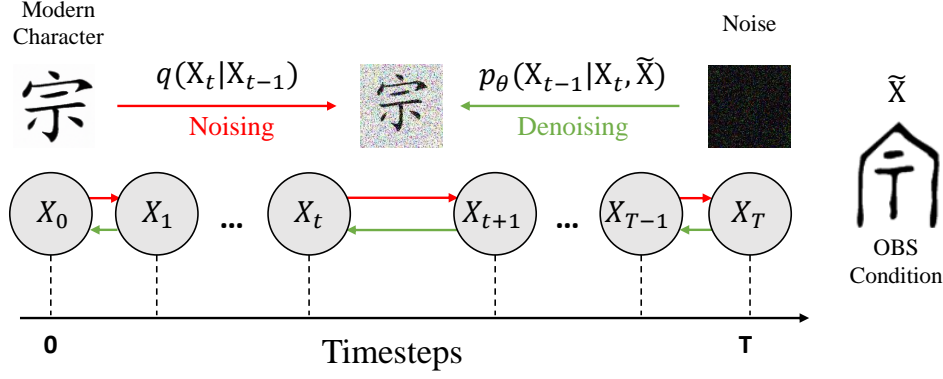


Figure 1: Conditional diffusion model for OBS decipherment.

match known characters but also to uncover new forms that could elucidate the meanings of these ancient scripts. Formally, the training set denoted as  $S = \{(s_i, c_i) \mid s_i \text{ is an OBS instance and } c_i \in C\}$ , pairs OBS instances with their modern Chinese counterparts from a set of known Categories  $C$ . The model is designed to extend beyond the training set  $S$ , identifying modern equivalents for OBS instances  $s'$ , and proposing new character forms where existing matches are absent.

To achieve this, our approach utilizes a diffusion-based (Ho et al., 2020) model, for transforming OBS character images  $\tilde{X}$  into their modern Chinese equivalents, as illustrated in Figure 1. The model operates in two phases: the forward process, the forward phase introduces noise to the modern Chinese character images  $X_0$ , transitioning them towards a state resembling pure noise via a controlled Markov chain process, ultimately conforming to a Gaussian distribution  $\mathcal{N}(0, I)$ . This is mathematically articulated as follows:

$$q(X_{1:T} \mid X_0) = \prod_{t=1}^T q(X_t \mid X_{t-1}) \quad (1)$$

where  $T$  denotes the total number of steps. For each step  $t$ , noise is added according to the following equation:

$$q(X_t \mid X_{t-1}) = \mathcal{N}(X_t \mid \sqrt{\alpha_t}X_{t-1}, (1 - \alpha_t)I) \quad (2)$$

where  $\alpha_t$  is a hyperparameter controlling the noise intensity, and  $I$  represents the identity matrix. The transition from  $X_0$  to a noisy state  $X_t$  over  $t$  step is captured by the equation:

$$X_t = \sqrt{\gamma_t}X_0 + \sqrt{1 - \gamma_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \quad (3)$$

with  $\gamma_t$  being the cumulative product of  $\alpha$  values up to  $t$ .

The denoising phase employs a U-Net architecture (Ronneberger et al., 2015) for the model  $f_\theta$ , trained to predict the noise  $\epsilon$  and restore the image. The training objective minimizes the loss function:

$$\mathcal{L} = \mathbb{E}_{\epsilon, \gamma} \left\| \epsilon - f_\theta(\tilde{X}, X_t, \gamma) \right\|^2 \quad (4)$$

which measures the discrepancy between the actual noise  $\epsilon$  and its estimation by the  $f_\theta$ . In the inference stage  $p_\theta(X_t \mid X_t, \tilde{X})$ , we reverse the noise addition process, starting from the noisiest state  $X_T$  and iteratively denoising down to  $t = 1$ .

$$X_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left( X_t - \frac{1 - \alpha_t}{\sqrt{1 - \gamma_t}} f_\theta(\tilde{X}, X_t, \gamma_t) \right) + \sqrt{1 - \alpha_t} \epsilon_t \quad (5)$$

where  $\epsilon_t \sim \mathcal{N}(0, I)$  introduces randomness to enhance the diversity of model generated results. The outcome is the denoised image  $\hat{X}_0$ , representing the deciphered results.

Building on this, our OBSD model integrates an Initial Decipherment phase with a Zero-shot Refinement stage to improve the decipherment accuracy. As shown in Figure 2, initially, an OBS image  $\tilde{X}$  undergoes conditional diffusion to approximate an initial decipherment  $X_0$ , which is then refined using a zero-shot learning approach, leveraging a reference style image  $X_{ref}$  to correct and enhance the structure. with a distinct style to enhance  $X_0$ , learning from the structure of modern Chinese characters. The final result  $X_F$  emerges as a refined representation of the intended modern Chinese character, benefiting from the refinement process's structural insights.

### 3.2 Initial Decipherment

After revisiting the fundamentals in Section 3.1, a preliminary and somewhat naive idea was to directly utilize OBS images as the condition  $\tilde{X}$  and modern Chinese characters as the target images  $X_0$

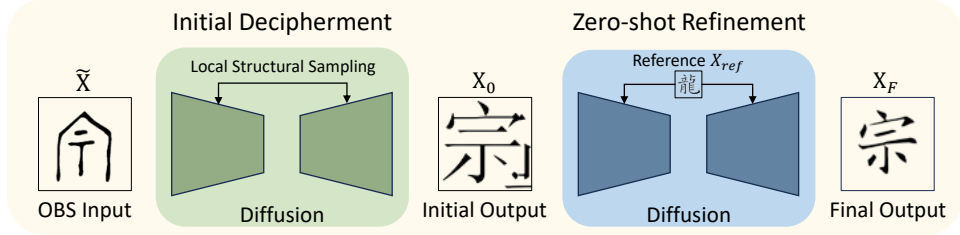


Figure 2: Overview pipeline of the proposed OBSD. The input OBS  $\tilde{X}$  undergoes a diffusion model to generate initial decipherment result  $X_0$ , which is then refined with a style-specific reference to produce the final output  $X_F$ .

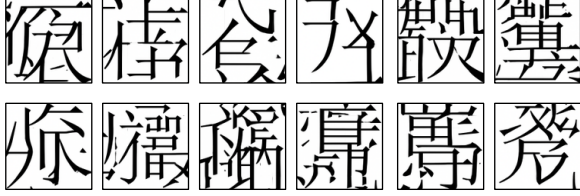


Figure 3: Directly training a conditional diffusion model results in **failure** decipherment.

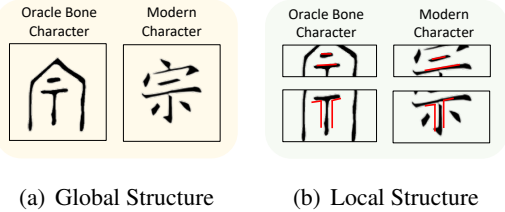


Figure 4: Comparative analysis of the Chinese character 宗 (zōng): (a) Depicts the evolution of global structure from OBS to its modern form. (b) Highlights the retention of specific local structures amidst the evolution.

to train a conditional diffusion model for decipherment. However, as shown in Figure 3, we observed that directly training such a model did not result in the accurate generation of the corresponding photos of modern Chinese characters. Instead, the model produces images comprised of a multitude of random stroke fragments, resembling gibberish. We speculate that this discrepancy arises because diffusion models are primarily designed for generating natural images, where the input conditions, such as edges and sketches, provide structural information to guide the generation of target images. However, in the context of deciphering OBS, the structural disparity between the input OBS images and the expected modern Chinese character outcomes is significant (see Figure 4(a)), rendering the standard conditional diffusion model ineffective for accurate reconstruction of the target modern characters. To address this challenge, we introduce the concept of Localized Structural Sampling (LSS) as a means to aid the diffusion model in learning how to map local radical structures of OBS to the corresponding

modern Chinese character space (see Figure 4(b) red marks), thereby enhancing the model’s capability to bridge the structural gap between ancient inscriptions and contemporary linguistic forms.

Figure 4 has demonstrated that despite the considerable structural evolution from OBS to modern Chinese characters, certain local structures have been preserved. As shown in Figure 5, to enable the diffusion model to learn these localized radical features, the LSS module employs a sliding window approach to segment the target modern Chinese character images  $X_0 \in R^{H \times W \times 3}$  and corresponding OBS images  $\tilde{X} \in R^{H \times W \times 3}$  into  $D$  patches of size  $\tilde{p} \times \tilde{p}$ , denoted as  $\tilde{X}^{(d)}$  and  $X_t^{(d)} \in R^{\tilde{p} \times \tilde{p} \times 3}$ ,  $d = 1, 2, \dots, D$ ,  $\tilde{p} = 64$ . Here,  $X_t$  represents the modern text image with added Gaussian noise  $\epsilon_t$  at timestep  $t$ . Consequently, we focus on learning the conditional reverse process as follows:

$$p_\theta(X_{0:T}^{(i)} | \tilde{X}^{(i)}) = p(X_T^{(i)}) \prod_{t=1}^T p_\theta(X_{t-1}^{(i)} | X_t^{(i)}, \tilde{X}^{(i)}) \quad (6)$$

By adopting this approach, the model iteratively refines each patch by learning the nuanced mappings from the localized structures of OBS to their modern counterparts. The loss function in Equation 4 can then be rewritten as follows:

$$\hat{\epsilon}_t^{(d)} = f_\theta(X_t^{(d)}, \tilde{X}^{(d)}, t) \quad (7)$$

$$\mathcal{L}' = \mathbb{E}_{t,d} \|\hat{\epsilon}_t^{(d)} - \epsilon_t^{(d)}\|^2$$

Here, the model’s goal is to minimize the difference between the estimated noise  $\hat{\epsilon}_t^{(d)}$ , and the actual noise,  $\epsilon_t^{(d)}$ , within each patch.

In the inference phase, our approach involves dissecting the OBS image  $\tilde{X}$  into  $\tilde{p} \times \tilde{p}$  patches, with  $p$  set at 64, through a structured grid layout, utilizing a sliding window for systematic extraction. The grid is arranged such that each cell hosts  $r \times r$  patches, with  $r$  set at 16, allowing for a finer subdivision than the patch size  $\tilde{p}$ . Patches are extracted by navigating the grid in both horizontal



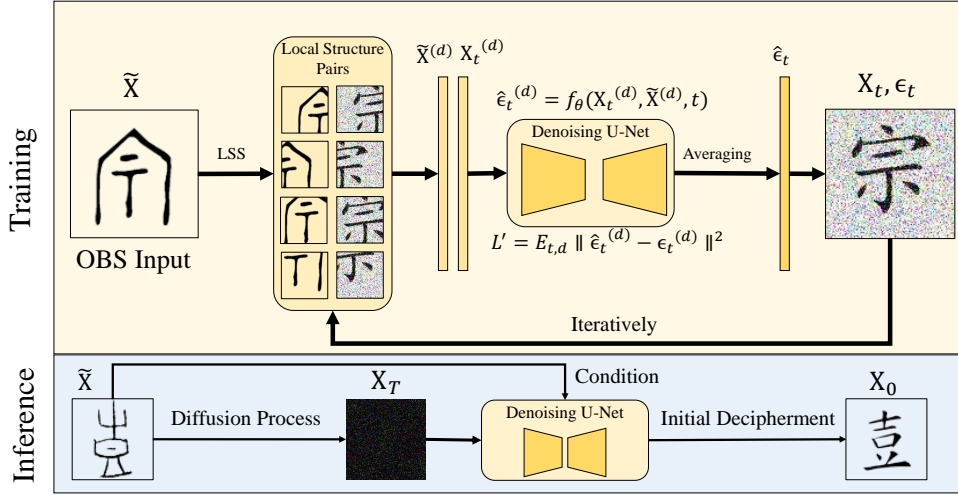


Figure 5: The overview pipeline of initial decipherment of OBSD.

and vertical directions with a step size of  $r$ . The initial decipherment model then progressively refines each patch by denoising and sampling.

---

**Algorithm 1** LSS Algorithm

---

**Require:** OBS image  $\tilde{X}$ , conditional diffusion model  $f_\theta(X_t, \tilde{X}, t)$ , dictionary of  $D$  overlapping patch locations.

- 1:  $X_T \sim \mathcal{N}(0, I)$
  - 2: **for**  $t = T, \dots, 1$  **do**
  - 3:    $\Omega_t = 0$  and  $M = 0$
  - 4:   **for**  $d = 1, \dots, D$  **do**
  - 5:      $X_t^{(d)} = \text{Crop}(P_d \circ X_t)$  and  $\tilde{X}^{(d)} = \text{Crop}(P_d \circ \tilde{X})$  //  $P_d$  represents the mask of the  $d$ th patch in the image.
  - 6:      $\Omega_t = \Omega_t + P_d \cdot f_\theta(X_t^{(d)}, \tilde{X}^{(d)}, t)$
  - 7:      $M = M + P_d$
  - 8:   **end for**
  - 9:    $\Omega_t = \Omega_t \odot M$  //  $\odot$ : element-wise division
  - 10:  $X_{t-1} = \frac{1}{\sqrt{\alpha_t}}(X_t - \frac{1-\alpha_t}{\sqrt{1-\gamma_t}}\Omega_t) + \sqrt{1-\alpha_t}\epsilon_t$   
//  $\epsilon_t \sim \mathcal{N}(0, I)$
  - 11: **end for**
  - 12: **return**  $X_0$
- 

Unique to our method is the handling of overlaps between patches. Instead of waiting until the denoising is complete, we average the overlapped sections at every timestep  $t$ , ensuring a uniform effect across the shared areas. This continuous averaging at each timestep prevents the formation of merging artifacts that typically occur when patches are processed independently. By smoothing transitions between patches during the sampling, we avoid edge discrepancies, maintaining the visual coherence of the reconstructed image. The sampling

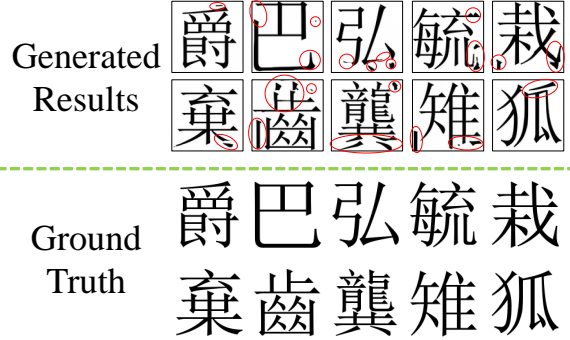


Figure 6: Modern Chinese characters generated by the initial decipherment stage, showing numerous artifacts and deformations as identified by the red circles.

dynamics at each step are defined by Equation 5, which guides the process toward a seamless and artifact-free image assembly. Algorithm 1 shows the pseudocode of LSS. Figure 5 demonstrates the overview pipeline of initial decipherment.

### 3.3 Zero-shot Refinement

Despite advancements in generating modern Chinese characters with Localized Structural Sampling, initial decipherment efforts encounter notable obstacles, such as structural deformities and artifacts, highlighted in Figure 6. These issues stem from the many-to-one training approach used, where multiple OBS instances are mapped to a single modern Chinese character image (see Figure 8), leading to confusion and inaccuracies in capturing character evolution, and resulting in artifacts or incomplete structures due to a limited variety of modern Chinese character samples.

To overcome these challenges, we propose a zero-shot refinement strategy that involves training a model on a diverse collection of modern Chi-

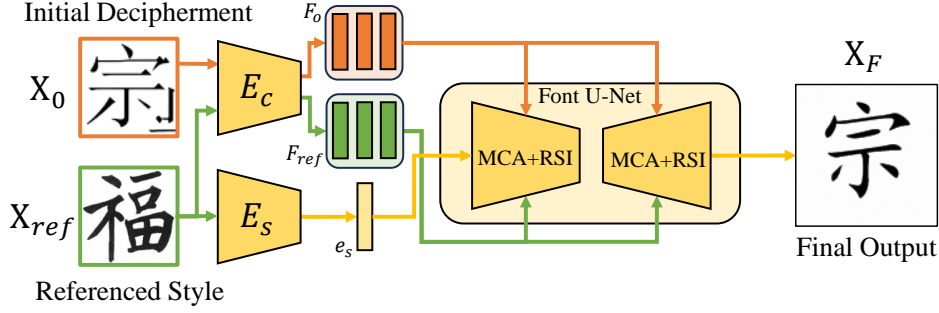


Figure 7: Overview pipeline of the zero-shot refiner.

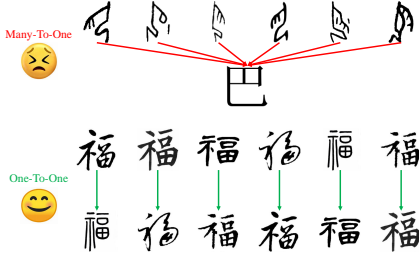


Figure 8: Comparison of many-to-one and one-to-one training paradigms. In the many-to-one approach, multiple OBC images, despite their large structural variances, are mapped to a single modern Chinese character. Conversely, the one-to-one paradigm ensures each image is individually paired.

nese characters. Considering the multiple writing styles for modern Chinese characters, we aim to improve the model’s understanding of their structure by employing a transformation task between different styles. We trained the module on 20 different modern Chinese character fonts to learn structural transformations between different modern Chinese character writing styles. As shown in Figure 8, this training process is one-to-one. This method simplifies data collection by leveraging readily available font variations, thereby enhancing the model’s understanding of character structures and enabling the application of this knowledge to improve initial decipherment results without direct training on OBS-to-modern character mappings.

Our zero-shot refinement approach is grounded in a generic font style transformation framework, as depicted in Figure 7 and based on (Yang et al., 2024). The process involves a dual-encoder system to adapt the style of a source font image  $X_0$  to a target style  $X_{ref}$ , preserving content integrity. The style encoder  $E_s$  extracts style features  $e_s$  from  $X_{ref}$ , while the content encoder  $E_c$  processes  $X_0$  and  $X_{ref}$  to obtain multi-scale content features  $F_0 = \{f_o^1, f_o^2, f_o^3\}$  and  $F_{ref} = \{f_{ref}^1, f_{ref}^2, f_{ref}^3\}$ , refined by a specialized UNet with Multi-scale Content Aggregation (MCA) and Reference-Structure

Interaction (RSI) blocks for enhanced feature integration. The model employs cross-attention mechanisms to align features and address structural differences, formalized as:

$$\begin{aligned} S_{ref} &\in \mathbb{R}^{C_{ref}^i \times H_i W_i} = \text{flatten}(f_{ref}^i) \\ S_s &\in \mathbb{R}^{C_s^i \times H_i W_i} = \text{flatten}(o_i) \\ Q &= \Phi_q(S_{ref}), K = \Phi_k(S_s), V = \Phi_v(S_s) \end{aligned} \quad (8)$$

where  $o_i$  represents the UNet feature derived from  $f_o^i$  and  $e_s$ , and  $\Phi_q, \Phi_k, \Phi_v$  denote linear projections. The deformation offset  $\delta_{\text{offset}}$  is calculated as follows:

$$\begin{aligned} F_{\text{attn}} &= \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \\ \delta_{\text{offset}} &= \text{FFN}(F_{\text{attn}}) \end{aligned} \quad (9)$$

The output  $I_f$  is the result of rendering the source image with DCN (Dai et al., 2017), considering the calculated deformation offset:

$$I_f = \text{DCN}(o_i, \delta_{\text{offset}}) \quad (10)$$

In adapting the framework for OBS decipherment, we streamline the model by focusing on a singular font style, thereby omitting the style contrastive refinement module and its contrastive loss, simplifying the training process. The encoders are trained using the offset loss  $\mathcal{L}_{\text{offset}}$ , which measures the mean magnitude of deformation offsets:

$$\mathcal{L}_{\text{offset}} = \text{mean}(\|\delta_{\text{offset}}\|) \quad (11)$$

where  $\delta_{\text{offset}}$  signifies the deformation offset, encapsulating structural information gleaned from the reference features, and the mean operation computes the average magnitude of these offsets.

After training, the zero-shot refinement module was directly employed to refine the results generated by the diffusion model.

## 4 Experiments

### 4.1 Dataset and Evaluation Metric

To train and evaluate the proposed OBSD model, we selected the HUST-OBS dataset (Wang et al., 2024) and EVOBC dataset (Guan et al., 2024), which stands as one of the largest repositories of OBS, with 1,590 distinct characters depicted in 71,698 images. Recognizing the complexities involved in deciphering unknown OBS, which usually require comprehensive expert validation, we opted for already deciphered inscriptions in our testing set to streamline the evaluation process. Importantly, the categories of characters in the testing set were specifically chosen to be absent from the training set, ensuring that the model faces the genuine challenge of deciphering **unseen** and novel categories. The dataset was partitioned into training and test sets with a 9:1 ratio, providing a robust framework for assessment.

While the proposed OBSD model approaches OBS decipherment from an image generation perspective, it is crucial to acknowledge that traditional image generation metrics, such as SSIM (Nilsson and Akenine-Möller, 2020), are not suitable for this distinct challenge. Instead, we adopted OCR technology as a more objective measure of decipherment success. Our custom-built OCR tool, OBS-OCR, is a simple classifier using ResNet-101 backbone specifically trained on a large dataset of 88,899 categories modern Chinese characters to evaluate the model’s output. The custom-built OCR tool achieved a recognition accuracy of 99.87% on 88,899 categories of Chinese characters, which demonstrates reliable performance to evaluate the decipherment results. Its aim is to automatically recognize the results generated by the diffusion models and compare these results with the ground truth in order to evaluate the model’s deciphering performance. By comparing the OCR-recognized characters against their ground truth labels, we simulate a quantifiable form of expert validation. To make a more reliable and objective evaluation, we also incorporated the widely-used, open-source Chinese OCR tool PaddleOCR<sup>1</sup> as an additional OCR tool to support further evaluations. This dual-OCR method provides a robust framework for assessing the model’s efficacy in accurately deciphering oracle bone languages.

<sup>1</sup><https://github.com/PaddlePaddle/PaddleOCR>

### 4.2 Quantitative Results

In quantitatively evaluating the performance of our proposed OBSD, we employ two distinct assessment criteria: single-round decipherment and multi-round decipherment. The single-round decipherment evaluation aims to gauge the method’s capability to decipher individual samples accurately, providing insight into its immediate effectiveness. On the other hand, the multi-round decipherment assessment offers a more practical appraisal of the method’s performance, where multiple attempts at deciphering a single image are permitted. This approach mirrors the iterative nature of real-world decipherment tasks, allowing for a comprehensive assessment of the method’s resilience and adaptability over successive trials.

Given the absence of dedicated tools for oracle bone language decipherment, we employ a comparative framework that adapts leading image-to-image translation methods to this specialized task. This set includes GAN-based approaches such as Pix2Pix (Isola et al., 2017), CycleGAN (Zhu et al., 2017), DRIT++ (Lee et al., 2020), and diffusion-based methods like CDE (Saharia et al., 2022b), Palette (Saharia et al., 2022a), BBDM (Li et al., 2023). This setting not only mirrors the core mechanism of our OBSD method but also allows for a comprehensive evaluation against the backdrop of the latest advancements in image translation. Each method was carefully adapted to the OBS context, ensuring consistent training and testing conditions for a fair evaluation.

In the single-round decipherment evaluation, as shown in Table 1, our OBSD demonstrates a significant advantage over the adapted image-to-image translation methods in deciphering oracle bone language. Notably, the top-1 accuracy for OBS-OCR and PaddleOCR achieved by OBSD stand at 41.0% and 30.0%, respectively, surpassing the performance of other methods. As the rank increases, there is a clear trend of improving accuracy, at Top-500 accuracy, OBSD reaches a 64.5% OBS-OCR recognition accuracy. It is noteworthy that all GAN-based approaches, such as Pix2Pix, Palette, DRIT++, and CycleGAN, exhibit minimal effectiveness in this context, with top-1 accuracies at 0%. This could be attributed to the GANs’ inherent challenge in capturing the complex and nuanced mappings required for accurately deciphering the oracle bone language into modern Chinese. Surprisingly, the adapted diffusion models, despite their

Evaluation Tool	Rank	Pix2Pix	Palette	DRIT++	CycleGAN	BBDM	CDE	OBSD (ours)
OBS-OCR	Top-1@Acc	0.0%	0.0%	0.0%	0.0%	19.5%	31.0%	<b>41.0%</b>
OBS-OCR	Top-10@Acc	0.0%	0.0%	0.0%	0.0%	29.5%	47.5%	<b>50.5%</b>
OBS-OCR	Top-20@Acc	0.0%	0.0%	0.0%	0.0%	34.5%	50.0%	<b>54.5%</b>
OBS-OCR	Top-50@Acc	0.0%	0.0%	4.5%	8.5%	39.0%	52.5%	<b>58.0%</b>
OBS-OCR	Top-100@Acc	0.0%	3.0%	13.0%	19.0%	42.0%	56.0%	<b>61.0%</b>
OBS-OCR	Top-200@Acc	14.5%	8.5%	20.0%	37.5%	46.0%	59.5%	<b>62.5%</b>
OBS-OCR	Top-500@Acc	17.5%	19.5%	21.5%	60.0%	58.0%	64.0%	<b>64.5%</b>
PaddleOCR	Top-1@Acc	0.0%	0.0%	0.0%	0.0%	7.0%	19.0%	<b>30.0%</b>

Table 1: Comparison of single-round decipherment success rate between the proposed OBSD and state-of-the-art image-to-image translation methods.

Number of Trial		1	2	3	4	5
Evaluation Tool	OBS-OCR	41.0%	56.0%	67.5%	75.5%	76.5%
	PaddleOCR	30.0%	40.0%	46.0%	50.5%	53.0%
Number of Trial		6	7	8	9	10
Evaluation Tool	OBS-OCR	77.0%	78.0%	79.5%	80.0%	80.0%
	PaddleOCR	55.5%	57.0%	57.5%	58.5%	58.5%

Table 2: Top-1 accuracy of the multi-round decipherment success rate of the proposed OBSD.

general-purpose nature, have shown commendable performance, underscoring the viability of leveraging image generation techniques in addressing the challenges traditional NLP algorithms encounter in decipherment tasks. This aligns with our methodological premise, validating the novel approach of integrating image-based generative models into the domain of linguistic decipherment.

In addition, Table 2 presents the multi-round decipherment results, where a progressive increase in decipherment success rates can be witnessed across multiple trials. The OBS-OCR metric starts at a success rate of 41.0%, and levels out at 80.0% by the 10th trial, showcasing the cumulative benefit of iterative testing. Similarly, the PaddleOCR metric exhibits a consistent upward trend, commencing at 30.0% and culminating at 58.5% in the final trial. These results validate the incremental improvements achievable through successive attempts.

### 4.3 Ablation Study

To further examine the impact of individual components in our proposed method, we conducted an ablation study focusing on the LSS module and zero-shot refinement. The results, presented in Table 3, highlight the limitations of employing only the basic conditional diffusion model for OBS decipherment, which resulted in notably low accuracy rates. Specifically, training the diffusion model without any enhancements led to outputs that were essentially nonsensical, characterized by random and uninterpretable stroke combinations (see Figure 3). The introduction of the LSS module marked a significant improvement, enabling the generation

Metric	Rank	Diffusion	+LSS	+Refinement
OBS-OCR	Top-1	0.5%	37.5%	41.0%
OBS-OCR	Top-10	2.5%	49.0%	50.5%
OBS-OCR	Top-20	4.5%	52.0%	54.5%
OBS-OCR	Top-50	6.5%	55.0%	58.0%
OBS-OCR	Top-100	9.0%	58.0%	61.0%
OBS-OCR	Top-200	10.5%	60.5%	62.5%
OBS-OCR	Top-500	16.5%	64.0%	64.5%
PaddleOCR	Top-1	0.0%	24.0%	30.0%

Table 3: Ablation Study of OBSD.

of decipherment outcomes with a Top-1 recognition rate of 37.5% for OBS-OCR and 24% for PaddleOCR. The addition of the zero-shot refinement module, in conjunction with the LSS, further increased the Top-1 accuracy for both OBS-OCR and PaddleOCR by an additional 3.5% and 6%, respectively.

### 4.4 Qualitative Results

OBS								
Ground Truth	百	沓	康	保	亥	鱼	竹	陟
Pix2Pix								
Palette	甫		權	婁	犀	婚	姨	弄
DRIT++								
CycleGAN								
BBDM	遺	告	之	執	梃	警	貞	復
CDE	伯	舌	尙	保	成	蟻	竹	韻
OBSD(ours no refinement)	百	沓	康	保	亥	魚	竹	陟
OBSD(ours)	百	沓	康	保	亥	魚	竹	陟

Figure 9: Comparison of qualitative results between the proposed OBSD and other state-of-the-art image-to-image translation frameworks, including Pix2Pix (Isola et al., 2017), Palette (Saharia et al., 2022a), DRIT++ (Lee et al., 2020), CycleGAN (Zhu et al., 2017), BBDM (Li et al., 2023), and CDE (Saharia et al., 2022b).



Figure 9 showcases the qualitative results of various image-to-image translation models, with our method, OBSD, standing out by producing the most accurate reconstructions of modern Chinese characters from OBS inputs. Pix2Pix (Isola et al., 2017), for example, generates outputs that are highly uniform across different inputs, demonstrating a lack of differentiation in character decipherment. On the other hand, DRIT++ (Lee et al., 2020) struggles to produce complete characters, often resulting in fragmented and unrecognizable forms. In stark contrast, OBSD demonstrates a robust capability to discern and reconstruct the intricate details of each OBS, leading to coherent and precise character forms that closely align with the ground truth. These results not only highlight the efficacy of OBSD but also its potential as a tool for experts in the field of oracle bone language decipherment.

To demonstrate the performance of OBSD on authentic, undeciphered OBS, we present an extensive evaluation in the appendix, depicted in Figure 10, 11, 12 and 13. This evaluation showcases a range of decipherment outcomes, from partial reconstructions that shed light on the structural elements of OBS characters, such as radicals and strokes, to complete character forms that exhibit a high resemblance to modern Chinese script. While the bulk of these results provide structural clues, the fully reconstructed characters hold particular promise, indicating the potential of OBSD to contribute meaningfully to the field of oracle bone language decipherment.

#### 4.5 Discussion

**Experiment Results:** We compared the proposed OBSD with other generic image generation models for the OBS deciphering task. As shown in Figure 9, most generic image generation models fail to produce structurally complete Chinese characters. This is because these methods, based on conditional generation, attempt to directly map the input OBS image to modern characters, neglecting the structural and writing conventions of the characters. In contrast, the proposed OBSD addresses these issues by incorporating local radical structure information into the training process, resulting in more accurate outputs.

**Analysis of Proposed Modules:** According to the experimental results, we found that the proposed LSS module effectively directs the diffusion model’s focus towards the local structures of both

OBS and modern Chinese characters. This results in clearer character strokes and more reasonable character structures. Additionally, the Zero-shot Refinement module refines the initial decipherment results by learning the structural characteristics of modern Chinese characters, ensuring a more precise and coherent structure.

**Generalizability to Other Languages:** The proposed method was initially designed for ideographic or pictographic languages, such as Chinese characters or Mayan script, where a single character represents a word or morpheme. This design enables the adaptation of the method to similar languages. For alphabetic scripts, which typically have a small number of letters, decipherment is rarely an issue. The applicability of these methods to other languages presents an interesting research question, which we will explore in our future work.

## 5 Conclusion

In this work, we presented OBSD, an innovative approach leveraging conditional image generation for the decipherment of OBS. Our novel Local Structure Sampling technique addresses the inherent challenges in learning modern Chinese characters’ structures from limited samples, enabling effective structural correspondence learning between OBS and modern Chinese characters. Furthermore, the integration of a zero-shot refinement module significantly enhances the decipherment accuracy, a claim substantiated by promising results on the HUST-OBS dataset and EVOBC dataset. The potential of OBSD extends beyond OBS, offering prospects for deciphering other ancient scripts, such as hieroglyphs and Maya glyphs. Looking ahead, we aim to collaborate with epigraphy experts to further validate and refine the OBSD, aspiring to advance AI’s role in the decipherment of ancient languages.

## 6 Limitations

In this study, we employed OCR technology, including a custom-built tool and the off-the-shelf package PaddleOCR, to evaluate the success of our OBSD in deciphering oracle bone language. While this approach offers a novel and objective metric, it is important to recognize its inherent limitations. However, these methods cannot be directly applied to evaluate truly undeciphered OBS, where the absence of ground truth necessitates expert validation.

Evaluating the decipherment results of entirely

unknown OBS characters presents a unique challenge that goes beyond the capabilities of OCR technology. This task involves interpreting historical, cultural, and linguistic contexts that are deeply embedded within the languages. Therefore, the ultimate validation of our model’s decipherment for such inscriptions requires the involvement of scholars and experts in oracle bone studies. We acknowledge the importance of this expert validation and are exploring collaborations with specialists in the field to assess the relevance and accuracy of our model’s outputs for genuinely undeciphered texts.

## 7 Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 61936003, No.62225603, No.62206103, No.62441604).

## References

- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Proc. Advances in Neural Inf. Process. Syst.*, 33:1877–1901.
- Xiang Chang, Fei Chao, Changjing Shang, and Qiang Shen. 2022. Sundial-gan: A cascade generative adversarial networks framework for deciphering oracle bone inscriptions. In *Proc. ACM Int. Conf. Multimedia*, pages 1195–1203.
- Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. 2017. Deformable convolutional networks. In *Proc. IEEE Int. Conf. Comp. Vis.*, pages 764–773.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *Proc. Annu. Conf. North Am. Chapter Assoc. Comput. Linguist.*
- Xuanming Fu, Zhengfeng Yang, Zhenbing Zeng, Yidan Zhang, and Qianting Zhou. 2022. Improvement of oracle bone inscription recognition accuracy: A deep learning perspective. *ISPRS International Journal of Geo-Information*, 11(1):45.
- Haisu Guan, Jinpeng Wan, Yuliang Liu, Pengjie Wang, Kaile Zhang, Zhebin Kuang, Xinyu Wang, Xiang Bai, and Lianwen Jin. 2024. An open dataset for the evolution of oracle bone characters: Evobc. *arXiv preprint arXiv:2401.12467*.
- Jun Guo, Changhu Wang, Edgar Roman-Rangel, Hongyang Chao, and Yong Rui. 2015. Building hierarchical representations for oracle character and sketch recognition. *IEEE Trans. Image Process.*, 25(1):104–118.
- Xu Han, Yuzhuo Bai, Keyue Qiu, Zhiyuan Liu, and Maosong Sun. 2020. Isobs: An information system for oracle bone script. In *Proc. Conf. Empir. Methods in Natural Language Process.*, pages 227–233.
- Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Proc. Advances in Neural Inf. Process. Syst.*, 33:6840–6851.
- Dongxin Hu. 2023. Coding design of oracle bone inscriptions input method based on “zhonghuaziku” database. In *Proc. Annu. Meet. Assoc. Comput. Linguist. Workshop*, pages 138–147.
- Shuangping Huang, Haobin Wang, Yongge Liu, Xiaosong Shi, and Lianwen Jin. 2019. Obs306: A large-scale oracle bone character recognition dataset. In *Proc. Int. Conf. Doc. Anal. and Recognit.*, pages 681–688. IEEE.
- Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, pages 1125–1134.
- Runhua Jiang, Yongge Liu, Boyuan Zhang, Xu Chen, Deng Li, and Yahong Han. 2023. Oraclepoints: A hybrid neural representation for oracle character. In *Proc. ACM Int. Conf. Multimedia*, pages 7901–7911.
- Kai Jin, Dan Zhao, and Wuying Liu. 2023. Morphological and semantic evaluation of ancient chinese machine translation. In *Proc. Annu. Meet. Assoc. Comput. Linguist. Workshop*, pages 96–102.
- Hsin-Ying Lee, Hung-Yu Tseng, Qi Mao, Jia-Bin Huang, Yu-Ding Lu, Maneesh Singh, and Ming-Hsuan Yang. 2020. Drit++: Diverse image-to-image translation via disentangled representations. *Int. J. Comput. Vis.*, 128:2402–2417.
- Bang Li, Qianwen Dai, Feng Gao, Weiye Zhu, Qiang Li, and Yongge Liu. 2020. Hwobc-a handwriting oracle bone character recognition database. In *Journal of Physics: Conference Series*.
- Bo Li, Kaitao Xue, Bin Liu, and Yu-Kun Lai. 2023. Bdbm: Image-to-image translation with brownian bridge diffusion models. In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, pages 1952–1961.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Lin Meng, Naoki Kamitoku, and Katsuhiko Yamazaki. 2018. Recognition of oracle bone inscriptions using deep learning based on data augmentation. In *Metrolology for Archaeology and Cultural Heritage*, pages 33–38. IEEE.
- Jim Nilsson and Tomas Akenine-Möller. 2020. Understanding ssim. *arXiv preprint arXiv:2006.13846*.

- Malte Nuhn, Arne Mauser, and Hermann Ney. 2012. Deciphering foreign language by combining language models and context vectors. In *Proc. Annu. Meet. Assoc. Comput. Linguist.*, pages 156–164.
- Yue Qi, Liu Liu, Bin Li, and Dongbo Wang. 2023. Vector based stylistic analysis on ancient chinese books: Take the three commentaries on the spring and autumn annals as an example. In *Proc. Annu. Meet. Assoc. Comput. Linguist. Workshop*, pages 117–121.
- Sujith Ravi and Kevin Knight. 2011. Deciphering foreign language. In *Proc. Annu. Meet. Assoc. Comput. Linguist.*, pages 12–21.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, pages 234–241. Springer.
- Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. 2022a. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–10.
- Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. 2022b. Image super-resolution via iterative refinement. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(4):4713–4726.
- Thea Sommerschildt, Yannis Assael, John Pavlopoulos, Vanessa Stefanak, Andrew Senior, Chris Dyer, John Bodel, Jonathan Prag, Ion Androustopoulos, and Nando de Freitas. 2023. Machine learning for ancient languages: A survey. *Comput. Linguist.*, pages 1–44.
- Mei Wang and Weihong Deng. 2024. A dataset of oracle characters for benchmarking machine learning algorithms. *Scientific Data*, 11(1):87.
- Mei Wang, Weihong Deng, and Cheng-Lin Liu. 2022. Unsupervised structure-texture separation network for oracle character recognition. *IEEE Trans. Image Process.*, 31:3137–3150.
- Pengjie Wang, Kaile Zhang, Yuliang Liu, Jinpeng Wan, Haisu Guan, Zhebin Kuang, Xinyu Wang, Lianwen Jin, and Xiang Bai. 2024. An open dataset for oracle bone script recognition and decipherment. *arXiv preprint arXiv:2401.15365*.
- Zhen Yang and Ting Fu. 2020. Oracle detection and recognition based on improved tiny-yolov4. In *Proceedings of the 2020 4th International Conference on Video and Image Processing*, pages 128–133.
- Zhenhua Yang, Dezhi Peng, Yuxin Kong, Yuyi Zhang, Cong Yao, and Lianwen Jin. 2024. Fontdiffuser: One-shot font generation via denoising diffusion with multi-scale content aggregation and style contrastive learning. In *Proc. AAAI Conf. Artificial Intell.*
- Chongsheng Zhang, Bin Wang, Ke Chen, Ruixing Zong, Bo-feng Mo, Yi Men, George Almpandis, Shanxiong Chen, and Xiangliang Zhang. 2022. Data-driven oracle bone rejoining: A dataset and practical self-supervised learning scheme. In *Proc. ACM SIGKDD Int. Conf. Knowledge discovery & data mining*, pages 4482–4492.
- Chongsheng Zhang, Ruixing Zong, Shuang Cao, Yi Men, and Bofeng Mo. 2021a. Ai-powered oracle bone inscriptions recognition and fragments rejoining. In *Proc. Int. Joint Conf. Artificial Intell.*, pages 5309–5311.
- Gechuan Zhang, Dairui Liu, Barry Smyth, and Ruihai Dong. 2021b. Deciphering ancient chinese oracle bone inscriptions using case-based reasoning. In *International Conference on Case-Based Reasoning*, pages 309–324. Springer.
- Yi-Kang Zhang, Heng Zhang, Yong-Ge Liu, Qing Yang, and Cheng-Lin Liu. 2019. Oracle character recognition by nearest neighbor classification with deep metric learning. In *Proc. Int. Conf. Doc. Anal. and Recognit.*, pages 309–314. IEEE.
- Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proc. IEEE Int. Conf. Comp. Vis.*, pages 2223–2232.

## A Appendix

### A.1 Implementation Details

The proposed OBSD was trained using the Adam optimizer with a weight decay of  $10^{-4}$ ,  $\beta_1 = 0.9$ , and  $\beta_2 = 0.999$ . During training, the learning rate was set to  $2e^{-5}$ , and the batch size was 8. Each batch contained 8 patches of size  $64 \times 64$ , and the model was trained on an Nvidia RTX A6000 model for 300 epochs. The entire training process spanned over 2 weeks.

### A.2 Decipherment Results on Genuine Unknown OBS

Figure 10, 11, 12 and 13 showcase the OBSD model’s decipherment outputs for previously undeciphered characters from the HUST-OBS (Wang et al., 2024) dataset and EVOBC dataset (Guan et al., 2024). For each character, we present a set of 10 potential interpretations, generated using distinct random seeds to ensure diversity in the results. In our commitment to supporting ongoing research in this field, we plan to make the code, the pre-trained models, and a comprehensive collection encompassing all decipherment outcomes publicly available. We hope this contribution will assist scholars and researchers in advancing the study of ancient languages.



Figure 10: Deciphered results for genuine undeciphered OBS.





出	斥	翊	恣	兽	芒	善	恣	平	至	翊
崇	乘	毒	毒	所	屋	里	流	翳	毒	
刊	宿	年	剑	字	中	齿	互	售	芬	年
抄	鱼	谷	楔	危	乖	荀	管	苙	芸	谷
端	颍	汧	媪	媪	姑	癖	媪	窠	冢	汧
夕	豕	受	陵	夕	夕	仄	仔	棗	丑	受
俞	乖	似	翁	請	烈	温	楷	垂	慈	似
寅	盂	豕	齋	齋	齋	齋	厩	厩	厩	齋
史	年	庭	尖	叫	止	岳	兵	鎰	甲	庭
級	級	勿	效	柔	效	殺	殺	殺	殺	勿
澗	驢	償	焦	陵	痾	痾	瘳	瘳	瘳	償
晉	馮	旣	荷	辰	蔚	只	并	帝	璠	旣
楚	姪	姪	姪	姪	姪	姪	姪	姪	姪	姪
聘	姓	尔	蒸	惘	尔	陸	妃	隨	堯	尔
彖	虱	乘	崑	龍	崑	崑	堯	堯	龍	乘
示	示	示	示	示	示	示	示	示	示	示
叔	又	族	身	浚	德	汝	灰	光	繫	族
000	珣	冏	冏	冏	冏	冏	冏	冏	冏	冏
冏	冏	冏	冏	冏	冏	冏	冏	冏	冏	冏
冏	冏	冏	冏	冏	冏	冏	冏	冏	冏	冏

Figure 13: Deciphered results for genuine undeciphered OBS.