

Radical-Conditioned Diffusion Model for Oracle Bone Character Generation and Analysis

Zengmao Ding¹, Xiaoping He², Xiao Li², Qi Li³, Xin Yan⁴, Xia Zhang⁴ and Bang Li^{1,*}

¹Key Laboratory of Oracle Bone Inscriptions Information Processing, Anyang Normal University, Anyang, China

²School of computer & information engineering, Anyang Normal University, Anyang, China

³College of Science and Engineering, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga, 525-8577, Japan

⁴State Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract

Oracle bone inscriptions (OBI), the earliest form of Chinese writing, are composed of complex characters built from recurring radical components. Understanding how these components form full characters is critical to studying the semantics and structure of early writing systems. In this paper, we propose a radical-conditioned diffusion model that synthesizes plausible OBI characters given a set of radicals and their counts. Our method encodes radical identity and positional context into structured embeddings, which condition the generation process via cross-attention in a U-Net backbone. To better preserve radical morphology and visual coherence, we introduce a perceptual loss that adapts dynamically during denoising. Experiments show that our model not only generates visually consistent and structurally valid characters, but also improves multi-label classification when used for data augmentation. These results demonstrate the potential of component-level generation as a tool for character reconstruction and structural analysis in ancient scripts.

Keywords

Oracle Bone inscriptions, Multi-instance Image Generation, DDPM Deep learning

Oracle bone inscriptions (OBI), known as the earliest systematic writing in China dating back to around 1600 BC, constitute a fundamental pillar for understanding ancient Chinese civilization, language, and the evolution of writing [1]. Crucially, the ideographic characters that form OBI are the direct progenitor of the modern Chinese character system. Within this early writing system, a relatively unified radical structure had already evolved. These recurring graphic elements serve as the building blocks of characters, are used for dictionary organization, and provide clues to a character's meaning and pronunciation [2]. Therefore, in-depth research into OBI radicals is essential for analyzing the reuse of components among oracle bone characters and understanding the fundamental principles of character composition within this ancient script.

The 7th International Symposium on Advanced Technologies and Applications in the Internet of Things (ATAIT 2025), September 10-11, 2025, Kusatsu, Japan

*Corresponding author.

✉ 200912004@stu.aynu.edu.cn (Z. Ding); hexp@stu.aynu.edu.cn (X. He); 01266@aynu.edu.cn (X. Li); liqi24@fc.ritsumeiki.ac.jp (Q. Li); xyan@bupt.edu.cn (X. Yan); xzhang@bupt.edu.cn (X. Zhang); libang@aynu.edu.cn (B. Li)

🆔 0009-0003-8258-2461 (Z. Ding); 0009-0009-1942-5862 (X. He); 0000-0003-2427-0630 (X. Li); 0000-0002-1963-5263 (Q. Li); 0000-0002-7719-0792 (X. Yan); 0000-0003-0937-2014 (X. Zhang); 0000-0002-7254-6569 (B. Li)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

However, it is important to emphasize that as an early stage in the development of Chinese characters, OBI exhibits a certain regularity in its radical structure, but the forms of its components are not absolutely fixed [2]. A key observation is that the same radical or component often displays subtle yet significant morphological variations across different OBI character forms. These differences are not arbitrary scribbles [3]; they frequently represent adaptive adjustments made by the scribe. These adjustments aimed to better integrate the component into the specific meaning or overall structure of the character it was part of, or were constrained by factors like the writing implement and spatial layout [4]. These nuanced morphological variations contain rich information about character construction, semantic associations, and writing conventions. They constitute a valuable window for researching the character-creation mindset and evolutionary patterns of OBI, holding extremely high value for systematic study.

OBC	Radical Annotations			OBC	Radical Annotations		

Figure 1: Illustration of the annotation of radicals in OBI within HWOBC [5]. Overview of radical annotations for OBI characters. The **OBI** column lists characters from the HWOBC dataset, while the **Radical Annotations** column shows corresponding radical information annotated in the Yinqueshuwenyan [6] corpus. For each character, the annotation specifies each radical instance r_i along with its occurrence count c_i , i.e., how many times r_i appears in the character.

To delve deeper into these characteristics of OBI radicals and their role in character formation, we model the relationship between a radical in an oracle bone character and its instances in different characters. As shown in Figure 1, the relationships between different radicals are not simply linear combinations, but involve more complex structural compositions, this task presents certain challenges, manifested primarily in data scarcity and in designing relational modeling specifically tailored to the aforementioned characteristics. To address these challenges, we utilize radical data from the YinQiWenYuan [6] database and leverage the powerful associative generative capabilities of diffusion models. We train a radical-guided diffusion model for OBI based on this data. The core idea is to input specific radical set information as a condition to guide the model in generating target oracle bone characters containing that radical set. During this process, we observe that the model spontaneously evolves a bias for different radicals under this training paradigm. This bias is manifested as the model spontaneously capturing and preserving the subtle morphological differences of radicals when they participate in forming

different characters (as mentioned above), as well as summarizing the positional regularity of radicals within OBI characters. This bias exhibited during the generation process provides us with a novel perspective to analyze the morphological characteristics, variation patterns of OBI radicals, and their deep associations with character meaning and structure.

In addition to structural analysis, radical-conditioned character generation provides a practical benefit in low-resource settings. Given the scarcity and imbalance of radical annotations in existing OBI datasets, we hypothesize that our model can generate structurally valid yet diverse character forms to augment the training data. These synthetic samples may help improve downstream classification tasks, particularly in the presence of rare radical combinations. We validate this hypothesis through a controlled data augmentation experiment in Section 3.3.

Overall, our main contributions can be summarized below:

- We propose a new generative task that maps oracle radicals to full characters, framing character construction as a component-conditioned generation process.
- We show that the generation behavior of a diffusion model inherently reflects structural biases, allowing us to analyze radical similarity, spatial regularity, and semantic stability through its learned latent space.

The remainder of this article is organized as follows. In Section 1, we review related work on oracle bone recognition and multi-instance image generation. Section 2 details our proposed radical-conditioned diffusion framework, including the embedding design and loss functions. Section 3 presents experimental settings, evaluation metrics, and both quantitative and qualitative analyses. Finally, Section 4 concludes the paper and discusses future directions.

1. Related Work

1.1. OBI Recognition

The recognition of OBI aims to classify characters in hand-written or authentic OBI images. Recent advancements [7, 8, 9, 10] in Oracle Bone Inscription (OBI) recognition, particularly to address challenges in recognizing complete characters, have highlighted the potential of component-level analysis. For instance, frameworks like the Oracle Bone Inscription Component Analysis proposed by Zhao et al. [1] utilize image similarity metrics to extract and compare radical components, revealing their structural roles across different OBI characters. Similarly, studies on character evolution, such as those employing few-shot learning to trace morphological changes from OBI to modern scripts, demonstrate that radicals undergo simplification, merging, and stroke variations to adapt to diverse character compositions [11]. These findings underscore the necessity of modeling radical-specific variations to achieve robust character generation. To support such modeling, large-scale datasets like HUST-OBC [12] provide a rich resource, containing 77,064 deciphered and 62,989 undeciphered character images, many of which exhibit significant radical variations due to evolving writing styles.

1.2. Multi-instance Image Generation

Multi-instance image generation (MIIG) focuses on synthesizing complex scenes containing multiple objects with precise spatial relationships and instance-specific attributes. Early text-to-

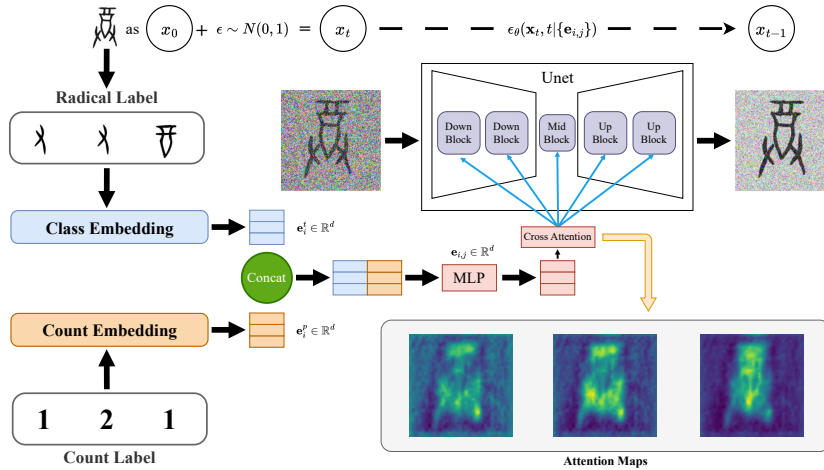


Figure 2: Given a target character and its associated radical labels with counts, we encode both the class and count information into structured embeddings. These are combined and projected into conditioning vectors, which are injected into the U-Net denoising process via cross-attention. Attention maps visualize how the model dynamically focuses on radical-relevant regions during generation.

image models struggled with compositional consistency, leading to innovations in instance-level control [13]. InstanceDiffusion [14] pioneered unified instance conditioning via its UniFusion module, supporting flexible location inputs (points, masks, boxes) and per-instance textual descriptions. Large-scale text-to-image diffusion models like Stable Diffusion [15], GLIDE [16], Imagen [17], and DALL-E 2 [18] Generate rich instances by using rich text combinations.

These works collectively address core MIIG challenges: unifying diverse instance conditions, mitigating inter-instance interference, and scaling relational reasoning. However, handling overlapping instances and abstract spatial instructions remains challenging. To Radical-Conditioned diffusion task is inherently challenging due to the non-fixed morphology of radicals across different characters and the need to model complex spatial and compositional relationships.

2. Methodology

Our core objective is to develop a generative model capable of synthesizing OBI characters under the guidance of specified radical components. Formally, given a set of radicals $\mathcal{R} = \{r_1, r_2, \dots, r_k\}$ and their corresponding frequency within the target character $\mathcal{C} = \{c_1, c_2, \dots, c_k\}$, where the r_i denotes the type of radical and the radical frequency r_i appears, the frequency r_i of each radical is included because repetition is semantically and structurally meaningful in oracle characters. For example, repeating a radical may imply intensity or plurality, and some characters are explicitly formed by duplicating a component. Capturing such information enables the model to generate character structures that more accurately reflect historical compositional rules. Our model aims to generate a plausible OBI character image x_0 that incorporates the specified radicals with their respective frequencies. The overview of the Radical-Conditioned Diffusion framework is shown in Figure 2.

2.1. Conditioned Generation via Radical-Guided Diffusion

We employ a Denoising Diffusion Probabilistic Model [19] as the backbone generative framework. The key innovation lies in how we effectively condition the diffusion process on the specified radical set $(\mathcal{R}, \mathcal{C})$.

Radical Representation with Positional Context: To encode both the radical type and its positional information within the target character’s composition, we design a specialized embedding module. For each radical instance r_i appearing c_i times in the character, we generate two embeddings: A type embedding $\mathbf{e}_i^t \in \mathbb{R}^d$ representing the semantic category of radical r_i . A positional embedding $\mathbf{e}_j^p \in \mathbb{R}^d$ representing the sequential order j (where $j = 1, \dots, c_i$) of occurrence for that radical type within the character. These embeddings are concatenated $[\mathbf{e}_i^t; \mathbf{e}_j^p]$ and projected into a unified conditioning vector $\mathbf{e}_{i,j} \in \mathbb{R}^d$ via a non-linear transformation:

$$\mathbf{e}_{i,j} = \text{Proj}([\mathbf{e}_i^t; \mathbf{e}_j^p])$$

where $\text{Proj}(\cdot)$ denotes a projection function implemented by a multi-layer perceptron (MLP) with a GELU activation. The complete conditioning signal for the diffusion model is the set of all such vectors $\{\mathbf{e}_{i,j}\}$ for all radical types $r_i \in \mathcal{R}$ and their $j = 1, \dots, c_i$ occurrences. This structured embedding explicitly informs the model what radicals are needed and in what sequence they are expected to appear, capturing potential positional biases observed in oracle bone script composition.

Integration into Diffusion: These conditioning vectors $\{\mathbf{e}_{i,j}\}$ are integrated into the diffusion model’s U-Net backbone using cross-attention layers. At each denoising step t , the intermediate features of the U-Net decoder attend to the conditioning embeddings, allowing the generation process to be dynamically guided by the specified radical composition throughout the diffusion trajectory.

The standard training objective for the diffusion model is to minimize the noise prediction error. Specifically, during the forward process of diffusion, the clean OBI character image \mathbf{x}_0 is progressively corrupted by adding Gaussian noise at timestep t , resulting in \mathbf{x}_t . The model ϵ_θ aims to predict the noise ϵ added to \mathbf{x}_0 . This standard diffusion loss is defined as:

$$\mathcal{L}_{diff} = \mathbb{E}_{\mathbf{x}_0, \epsilon \sim \mathcal{N}(0, \mathbf{I}), t} [\|\epsilon - \epsilon_\theta(\mathbf{x}_t, t | \{\mathbf{e}_{i,j}\})\|_2^2]$$

Here, $\epsilon_\theta(\mathbf{x}_t, t | \{\mathbf{e}_{i,j}\})$ denotes our conditional diffusion model, which takes the noisy image \mathbf{x}_t , the timestep t , and the radical conditioning embeddings $\{\mathbf{e}_{i,j}\}$ as input to predict the noise ϵ .

2.2. Content-Aware Perceptual Loss

Standard diffusion training optimizes the prediction of the noise ϵ added at step t . To enhance the semantic fidelity and structural coherence of generated characters, particularly respecting the subtle morphological variations of radicals, we introduce a supplementary **Content-Aware Perceptual Loss** $\mathcal{L}_{content}$ [20].

This loss operates on the predicted clean image $\hat{\mathbf{x}}_0$ at an intermediate denoising step t , derived from the current noisy state \mathbf{x}_t and the predicted noise $\hat{\epsilon}_\theta$:

$$\hat{\mathbf{x}}_0 = (\mathbf{x}_t - \sqrt{1 - \bar{\alpha}_t} \hat{\epsilon}_\theta) / \sqrt{\bar{\alpha}_t}$$

where $\bar{\alpha}_t$ is the cumulative product of the variance schedule. We extract multi-scale features $\phi_l(\cdot)$ from different layers l of a pre-trained feature extractor (VGG16 network) for both the true clean image \mathbf{x}_0 and the predicted $\hat{\mathbf{x}}_0$.

Crucially, $\mathcal{L}_{content}$ weights the contribution of different feature levels based on the current timestep t :

$$\mathcal{L}_{content} = \frac{1}{N} \sum_l w_l(t) \cdot \|\phi_l(\mathbf{x}_0) - \phi_l(\hat{\mathbf{x}}_0)\|_2^2$$

Here, $w_l(t)$ is a time-dependent weighting function. **Low-level features** (capturing edges, textures) are emphasized during **high-noise stages** (large t), as they are crucial for establishing the fundamental radical shapes and layout early in denoising. Conversely, **high-level features** (capturing semantic structures) are emphasized during **low-noise stages** (small t), refining the semantic coherence and fine details of the radicals and their integration as the image nears completion. N is a normalization factor accounting for the number of active feature elements at each timestep. This dynamic weighting ensures the loss focuses on the most relevant visual aspects at each denoising phase, significantly improving the preservation of radical morphology and overall character integrity.

2.3. Overall Training Objective

The complete loss function for training our radical-guided diffusion model combines the standard DDPM noise prediction loss \mathcal{L}_{diff} , and the Content-Aware Perceptual Loss $\mathcal{L}_{content}$:

$$\mathcal{L}_{total} = \mathcal{L}_{diff} + \lambda \mathcal{L}_{content}$$

where λ is hyperparameter balancing the contribution of each loss term. The model is trained end-to-end, learning to generate semantically coherent and structurally accurate oracle bone characters conditioned on the specified radical set and counts, while simultaneously developing rich internal representations of radical morphology and compositional rules.

3. Experiment

In this section, we evaluate our radical-conditioned diffusion model for oracle character generation. Conventional metrics like FID or Inception Score are unsuitable, since oracle characters admit multiple valid forms for the same radical set. Instead, we adopt structure-aware evaluation: multi-label classification (Section 3.3), case studies (Section 3.4), and semantic embedding visualization (Figure 5) to assess whether generated characters contain the intended radicals while maintaining structural diversity and positional patterns observed in real OBI samples.

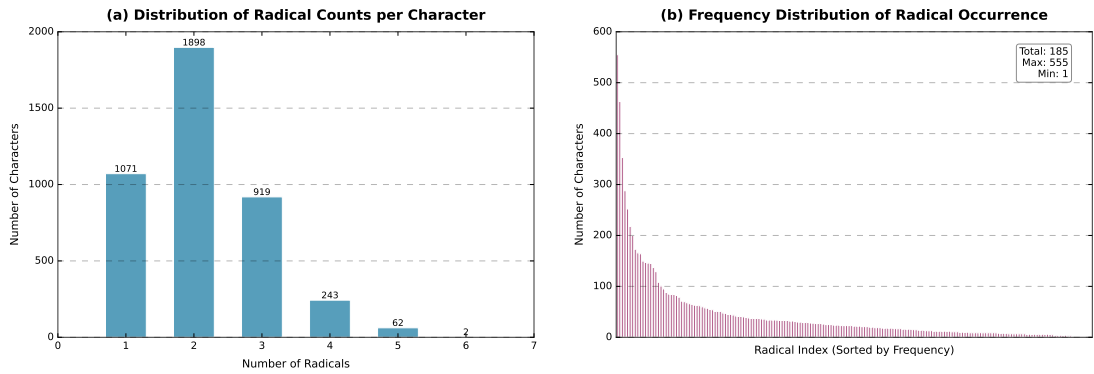


Figure 3: Radical statistics. (a) Radicals per character. (b) Radical frequency distribution (sorted). Shows long-tail pattern: few frequent, many rare radicals.

3.1. Dataset

We construct our dataset based on the HWOBC [5] collection, which contains handwritten oracle character images covering 3,881 distinct character categories. A large portion of these categories have been structurally annotated with radical information in the Yinqi Wen Yuan project. Using these annotations, we associate 185 distinct radicals with 3,767 oracle character categories, resulting in 80,823 annotated character images. Each annotated character is associated with one or more radicals that reflect its structural components. The distribution of radical annotations per category is visualized in Figure 1 and the distribution of radicals shown in Figure 3.

3.2. Implementation details

All experiments have been performed on a platform with an NVIDIA GeForce RTX 4090 graphics card. The deep learning framework used was PyTorch 2.6.1 with CUDA 12.4. The epochs and batch size have been set to 400 and 64, All models have been trained using the AdamW optimizer, with an initial and final learning rate of 0.0001, momentum of 0.937, and weight decay of 0.0005. The learning rate has adopted the warm-up cosine annealing algorithm. The learning rate of the model gradually increases within the first 3 epochs. In the experiment, nearly all models triggered early stopping around epoch 400, which means sufficient training has been conducted on the whole dataset.

To evaluate the effectiveness of our radical-guided diffusion model and assess the utility of the generated samples in downstream multi-label classification tasks, we construct a controlled experimental setup based on the YinQiWenYuan [6] oracle bone character dataset. Each character in the dataset is annotated with one or more radicals, and the task is formulated as multi-label classification over these radicals.

Due to the highly imbalanced nature of radical distributions and the limited number of samples for certain rare radicals, it is crucial to ensure that all radicals are observed during training while avoiding trivial memorization of character samples. To this end, we adopt a coverage-based greedy selection strategy to construct the training set. Specifically, we iteratively select samples that contribute the most previously unseen radicals until the entire radical set

is covered. This ensures that the model is exposed to all radical types during training while leaving a subset of unseen character forms for evaluation.

The remaining samples are filtered to construct a test set such that each test sample contains only radicals already present in the training set. This constraint ensures that the evaluation focuses on compositional generalization rather than extrapolation to unseen radical types. Finally, we randomly sample 10% of the total dataset to form the test set, with the remaining 90% used for training. Both the radical-guided diffusion model and the baseline multi-label classifier are trained on this split.

3.3. Performance analysis

To evaluate diffusion-based augmentation, we generate 8,000 synthetic oracle bone character images for each radical set size (1–5) using our radical-conditioned diffusion model. The synthetic data are combined with the original training set to retrain the multi-label classifier.

Importantly, the classification model used to evaluate augmentation effects is trained from scratch under two conditions: with and without the inclusion of diffusion-generated data. This enables a direct comparison of the impact of synthetic data on the classifier’s performance across multiple metrics.

Table 1 summarizes the performance of the classifier before and after introducing diffusion-generated samples into the training set. Overall, the inclusion of generated samples leads to consistent improvements across all evaluation metrics. Notably, the Average Precision improves from 0.671 to 0.702, and the Subset Accuracy increases from 0.286 to 0.307. The Hamming Loss is also slightly reduced, indicating better precision in multi-label predictions.

Table 1

Comparison of multi-label classification metrics before and after data augmentation. Arrows (↑/↓) indicate that higher/lower values are better, respectively.

Metric	Before Augmentation	After Augmentation
Average Precision (AP) ↑	0.671	0.702
Coverage Error ↓	15.734	14.730
Hamming Loss ↓	0.00766	0.00755
Subset Accuracy ↑	0.286	0.307
F1 Score ↑	0.570	0.601
Jaccard Index ↑	0.418	0.440
Recall ↑	0.549	0.592
Precision ↑	0.640	0.658

These results demonstrate that the diffusion model not only generates plausible character forms conditioned on radical sets but also introduces useful variance into the training data that benefits generalization. The improvement in recall and F1 score further suggests that the model becomes more capable of identifying rare or co-occurring radicals after exposure to the synthetic examples. This validates our hypothesis that the morphological diversity captured by the diffusion model can enhance downstream classification performance.

3.4. Case study

	Input Radical	Generation Sample	Real Sample
Easy Instance			
Complex Instance			

Figure 4: Visual comparison of generated and real oracle bone characters conditioned on specified radicals. Each row shows: (1) the input radical(s), (2) multiple generated samples guided by the same radical(s), and (3) real samples containing those radicals. The upper half illustrates easy instances with simpler structures, while the lower half showcases complex instances with intricate radical combinations.

To further evaluate the fidelity and diversity of generated samples, we conduct a qualitative case study comparing generated characters with real ones sharing the same radical components, as shown in Figure 4. For each example, we provide the input radical(s), multiple generated results, and real oracle bone characters containing those radicals.

In the Easy Instance setting, radicals are typically standalone or structurally dominant. The generated characters preserve visual similarity to real samples while capturing stylistic nuances—stroke thickness, spatial balance, and subtle morphological variations typical of handwritten oracle scripts.

In the Complex Instance setting, inputs include multiple radicals with diverse layouts and structural entanglement. The model produces plausible characters that retain radical identity and arrangement, often applying adaptive transformations—stretching, rotation, or compression—to emulate real OBI spatial strategies. These results demonstrate the model’s ability to internalize compositional flexibility and radical-level variation.

Overall, the visual results support our claim that the radical-conditioned diffusion model effectively captures both the identity and adaptive behavior of radicals in context, enabling realistic and semantically meaningful character generation.

To investigate how the model organizes radical information, As shown in Figure 5, we extract the embedding vectors $e_{i,j}$ and visualize them with t-distributed Stochastic Neighbor

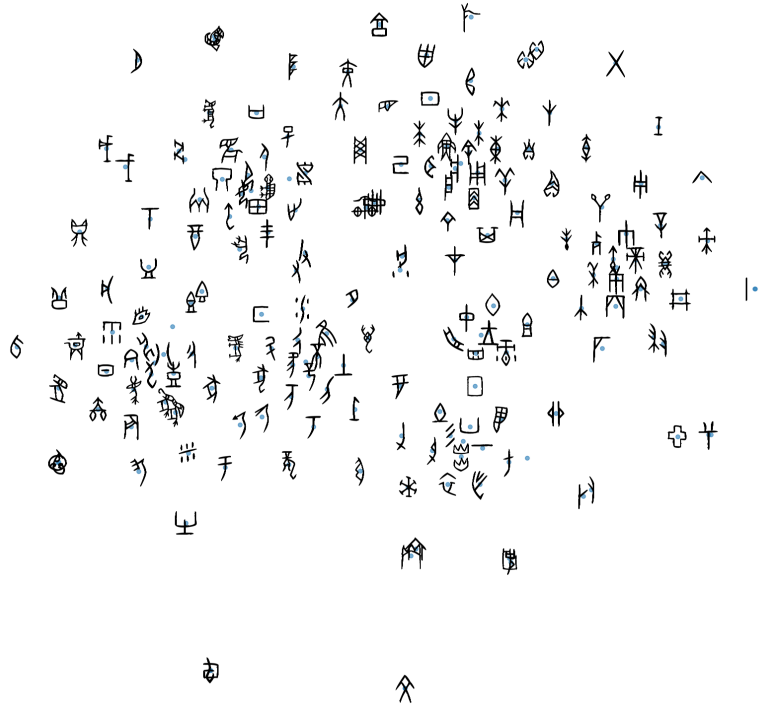


Figure 5: t-SNE visualization of radical embeddings extracted from the conditioning encoder. Each point corresponds to a specific radical instance. Similar radicals are clustered together, revealing potential confusions, while isolated radicals form distinct clusters.

Embedding(t-SNE) [21]. We find that radicals with similar morphological structures cluster closely in the embedding space, suggesting the model captures structural similarity. However, this can cause confusion during generation, as the model may struggle to distinguish between similar radicals, especially those sharing stroke patterns or symmetry. In contrast, radicals with larger shape differences are easier to cluster and preserve stably during generation, with less distortion or substitution. This implies a trade-off between embedding expressiveness and discriminability, which may be improved via contrastive regularization or additional supervision.

4. Conclusion

In this work, we propose a radical-conditioned diffusion model for oracle bone character generation, which effectively captures the morphological variations and combinatorial patterns of radicals. By incorporating embeddings that encode both radical identity and positional information, the model preserves subtle visual features of radicals across different character contexts. Our experiments demonstrate that such a generative approach facilitates structural understanding of oracle bone script and models the compositional relationships among specific radicals. Future work may integrate structural priors or contrastive learning to further enhance radical disambiguation, and extend the task to generate corresponding glyphs from different historical stages using radical-based priors derived from oracle bone inscriptions.

Acknowledgments

This research was supported by the Natural Science Foundation of China (Grant No. 62506007), the Natural Science Foundation of Henan Province (Grant No. 242300420680), the Paleography and Chinese Civilization Inheritance and Development Program (Grant Nos. G1807, G1806, G2821), the Henan Province Science and Technology Research Project (Grant Nos. 242102210116, 252102321071), Major Science and Technology Project of Anyang (Grant No. 2025A02SF007) and the Henan Province High-Level Talents International Training Program (Grant No. GCC2025028). National Natural Science Foundation of China (Grant No. 62506007). Major Science and Technology Project of Anyang (Grant No. 2025A02SF007).

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

- [1] Z. Zhao, Q. Jiao, et al., Oracle bone inscriptions components analysis based on image similarity, in: 2020 IEEE 9th joint international information technology and artificial intelligence conference (ITAIC), volume 9, IEEE, 2020, pp. 1666–1670.
- [2] J. Zhang, On the improvement of the oracle radical system, *Lexicographical Studies* (2013) 27–33. doi:10.16134/j.cnki.cn31-1997/g2.2013.05.004.
- [3] Q. Zhu, On the radicals in oracle bone inscriptions: The earliest set of pictographs in china, *Journal of Ancient Books Collation and Studies* (2002) 32–35. In Chinese.
- [4] Y. Wang, A Study on the Evolution of the Radical System in Oracle Bone Inscriptions, Master's thesis, Zhengzhou University, 2020.
- [5] B. Li, Q. Dai, F. Gao, W. Zhu, Q. Li, Y. Liu, Hwobc-a handwriting oracle bone character recognition database, in: *Journal of Physics: Conference Series*, volume 1651, IOP Publishing, 2020, p. 012050.
- [6] Yinqi wen yuan: Oracle bone inscriptions information platform, <https://jgw.aynu.edu.cn/>, 2025. Maintained by Anyang Normal University in collaboration with the Chinese Academy of Social Sciences, updated in 2025.
- [7] R. Jiang, Y. Liu, B. Zhang, X. Chen, D. Li, Y. Han, Oraclepoints: A hybrid neural representation for oracle character, in: *Proceedings of the 31st ACM International Conference on Multimedia, MM '23*, Association for Computing Machinery, New York, NY, USA, 2023, p. 7901–7911. URL: <https://doi.org/10.1145/3581783.3612534>. doi:10.1145/3581783.3612534.
- [8] J. Li, Q.-F. Wang, R. Zhang, K. Huang, Mix-up augmentation for oracle character recognition with imbalanced data distribution, in: *Document Analysis and Recognition – ICDAR 2021: 16th International Conference, Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part I*, Springer-Verlag, Berlin, Heidelberg, 2021, p. 237–251. URL: https://doi.org/10.1007/978-3-030-86549-8_16. doi:10.1007/978-3-030-86549-8_16.

- [9] M. Wang, W. Deng, C.-L. Liu, Unsupervised structure-texture separation network for oracle character recognition, *IEEE Transactions on Image Processing* 31 (2022) 3137–3150. doi:10.1109/TIP.2022.3165989.
- [10] F. Gao, J. Zhang, Y. ge Liu, Y. Han, Image translation for oracle bone character interpretation, *Symmetry* 14 (2022) 743. URL: <https://api.semanticscholar.org/CorpusID:247959908>.
- [11] M. Wang, Y. Cai, L. Gao, R. Feng, Q. Jiao, X. Ma, Y. Jia, Study on the evolution of chinese characters based on few-shot learning: From oracle bone inscriptions to regular script, *Plos one* 17 (2022) e0272974.
- [12] P. Wang, K. Zhang, X. Wang, S. Han, Y. Liu, J. Wan, H. Guan, Z. Kuang, L. Jin, X. Bai, et al., An open dataset for oracle bone character recognition and decipherment, *Scientific Data* 11 (2024) 976.
- [13] X. Li, J. Thickstun, I. Gulrajani, P. S. Liang, T. B. Hashimoto, Diffusion-lm improves controllable text generation, *Advances in neural information processing systems* 35 (2022) 4328–4343.
- [14] X. Wang, T. Darrell, S. S. Rambhatla, R. Girdhar, I. Misra, Instancediffusion: Instance-level control for image generation, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 6232–6242.
- [15] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-resolution image synthesis with latent diffusion models, in: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. URL: <http://dx.doi.org/10.1109/cvpr52688.2022.01042>. doi:10.1109/cvpr52688.2022.01042.
- [16] A. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. McGrew, I. Sutskever, M. Chen, Glide: Towards photorealistic image generation and editing with text-guided diffusion models (2021).
- [17] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. L. Denton, K. Ghasemipour, R. Gontijo Lopes, B. Karagol Ayan, T. Salimans, et al., Photorealistic text-to-image diffusion models with deep language understanding, *Advances in neural information processing systems* 35 (2022) 36479–36494.
- [18] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, M. Chen, Hierarchical text-conditional image generation with clip latents, *arXiv preprint arXiv:2204.06125* 1 (2022) 3.
- [19] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, *Advances in neural information processing systems* 33 (2020) 6840–6851.
- [20] Z. Yang, D. Peng, Y. Kong, Y. Zhang, C. Yao, L. Jin, Fontdiffuser: One-shot font generation via denoising diffusion with multi-scale content aggregation and style contrastive learning, in: *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 2024, pp. 6603–6611.
- [21] L. v. d. Maaten, G. Hinton, Visualizing data using t-sne, *Journal of machine learning research* 9 (2008) 2579–2605.