OracleFusion: Assisting the Decipherment of Oracle Bone Script with Structurally Constrained Semantic Typography

Supplementary Material

A. Case Study for OBS Decipherment

To evaluate the decipherment capabilities of OracleFusion, we applied it to both undeciphered and deciphered OBS, as shown in Figure 1. In the first two rows, OracleFusion generates location, semantic, and visual results for an undeciphered character, providing plausible interpretations, such as "drooping crops" and "thorny seeds," that could offer valuable clues for expert analysis. In the last two rows, OracleFusion accurately replicates the structure and meaning of a deciphered character, closely matching the ground truth. These results demonstrate its effectiveness in capturing both the semantic essence and structural details of OBS, supporting its potential as a valuable tool in OBS decipherment.

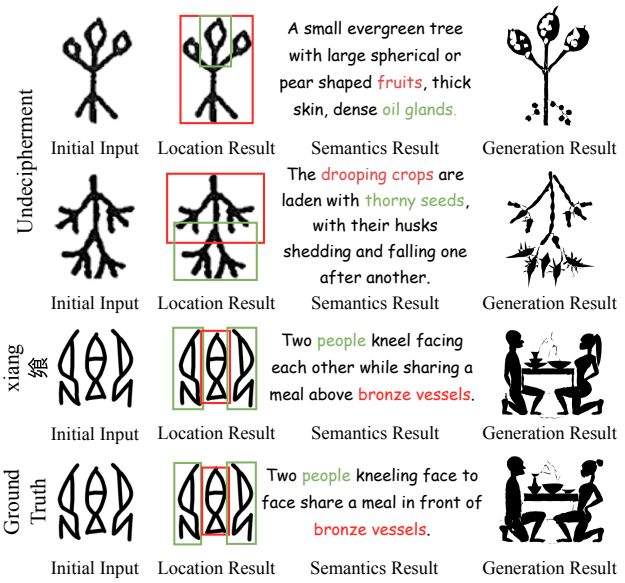

Figure 1. Results of unseen undeciphered and deciphered OBS.

B. Applications of OracleFusion

In this section, we explore additional applications of OracleFusion, emphasizing its potential to advance the decipherment of oracle bone script. The black-and-white glyphs produced by our model, enriched with semantic information, can be effectively employed as inputs for state-of-the-art frameworks such as the Depth-to-Image in Stable Diffusion 2 [3] and the Scribble-to-Image generation in ControlNet-SDXL 1.0 [4]. This integration highlights OracleFusion's ability to generate structured and interpretable outputs, enabling downstream models to leverage these representations for image-based analysis and reconstruction tasks. These results further illustrate the versatility of OracleFusion as a foundational tool for bridging symbolic representation and computational decipherment methodologies.

B.1. Depth-to-Image Post-Processing

In this subsection, we employ the Depth-to-Image in Stable Diffusion 2 [3] as a post-processing step to enhance the outputs of our method by incorporating color and texture. The results, presented in Figure 2, highlight the effectiveness of integrating depth-to-image processing for improving the visual fidelity and interpretability of the generated illustrations. This approach demonstrates the potential for combining generative and post-processing techniques to further refine and contextualize the outputs of computational decipherment systems.

B.2. Scribbles-to-Image Post-Processing

In this subsection, we utilize the results from our method as conditional inputs to guide the Scribbles-to-Image generation in ControlNet-SDXL 1.0 [4] to produce natural images. The results, presented in Figure 3, demonstrate the potential of integrating this framework to significantly enhance the visual quality and realism of the generated illustrations. This integration highlights the power of combining structured semantic information with advanced generative models to improve the fidelity and interpretability of computationally generated content.

C. Additional Decipherment Results

Figure 4 presents the decipherment results generated by OracleFusion for previously undeciphered oracle characters. For each character, the figure provides both grounding result and semantic typography result. Note that the different colors of bounding boxes in the grounding results indicate distinct concepts of key structural components. The concepts are predicted by our method but omitted from the figure for clarity. This contribution aims to support scholars and researchers in advancing the computational study and linguistic analysis of ancient scripts.

D. User Study Details

To evaluate the effectiveness of our OracleFusion, we conducted a user study on comparisons between Word-As-Image [2], ClipDraw [1] and our method. Figure 5 shows the description that participants read before answering the

Figure 2. Examples of utilizing Depth-to-Image in Stable Diffusion 2 [3] as a post-processing step.

questions. As shown in Figure 6, for each oracle character in our study, we present the initial glyph and three illustrations generated by the aforementioned three methods. Participants were asked to rate scores on semantics, visual appeal, and glyph maintenance, respectively. In Figure 7, we display some of the qualitative comparison results in our user study.

Figure 3. Examples of utilizing Scribbles-to-Image with ControlNet-SDXL 1.0 [4] as a post-processing step.

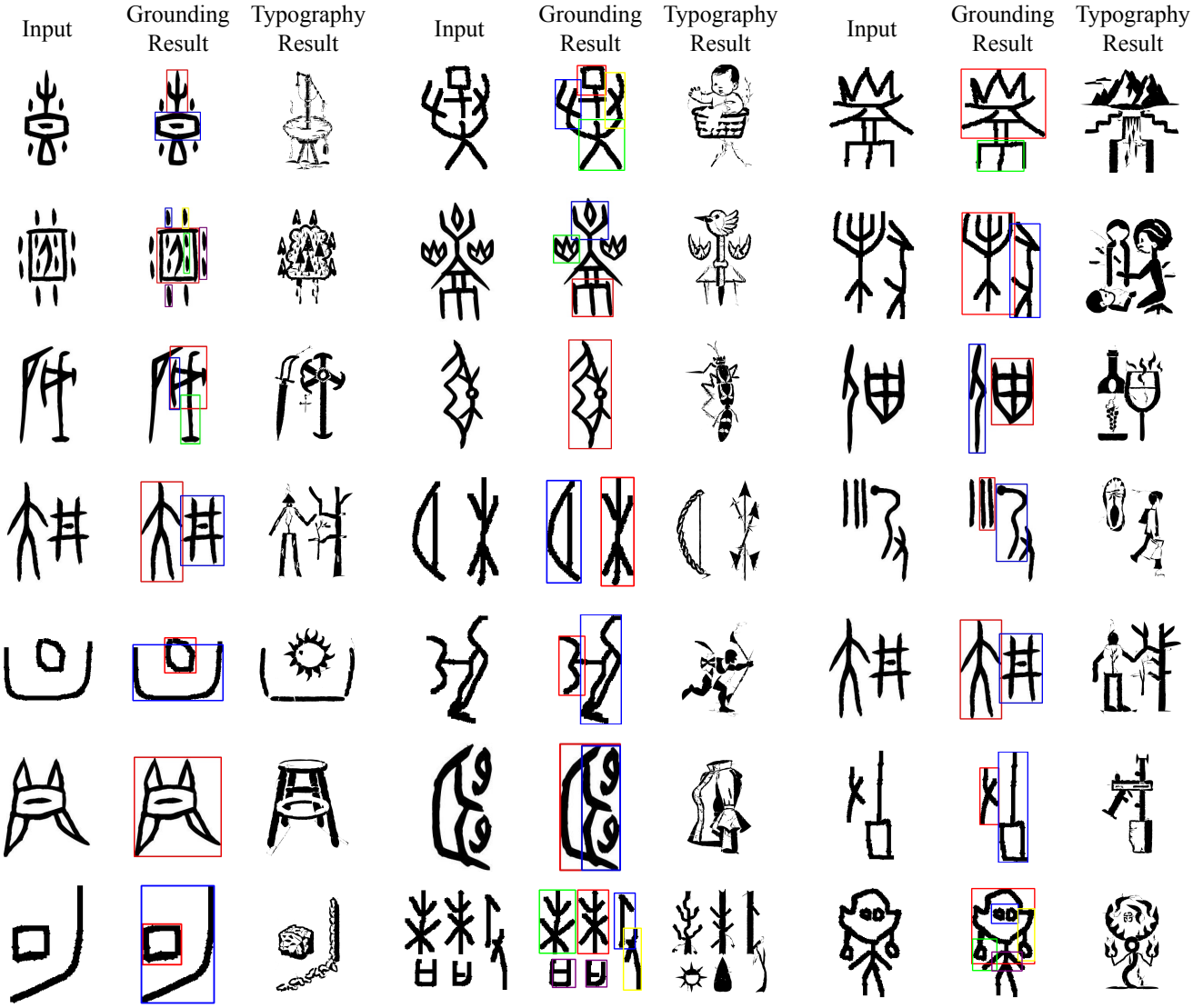

Figure 4. More grounding and semantic typography results for undeciphered oracle bone scripts. Note that the different colors of bounding boxes in the grounding results indicate distinct concepts of key structural components. The concepts are predicted by our method but omitted from the figure for clarity.

Oracle Semantic Typography Survery

Thank you for participating in our Oracle Semantic Typography survey. Each page displays an initial oracle glyph with its semantic concept and three results from different methods. Please rate each result on a scale of 1 to 5 based on the following three criteria, with a higher score indicating a better result.

Scoring Criteria:

1. **Semantics**:

➢ Please evaluate whether the typography result accurately reflects the meaning of the character. Consider whether the layout of the character aligns with its meaning and the significance of its radical, and whether it clearly conveys the original semantics of the Oracle bone character.

➢ **Score Range:** 1-5 (1 = does not reflect the meaning at all, 5 = fully reflects the meaning).

2. **Visual Appeal**:

➢ Please evaluate the artistic quality of the typography result. Consider whether the layout is visually appealing, smooth, and harmonious, and whether the overall aesthetic is pleasing.

➢ **Score Range:** 1-5 (1 = poor visual appeal, 5 = excellent visual appeal).

3. **Glyph Maintenance**:

➢ Please evaluate whether the typography result maintains the original shape of the Oracle bone character.

➢ **Score Range:** 1-5 (1 = significant deviation from original shape, 5 = minimal deviation from original shape).

Please independently rate each Oracle bone character typography result based on the above three criteria.

We appreciate your time and effort. Thank you!

Figure 5. The description of our user study.

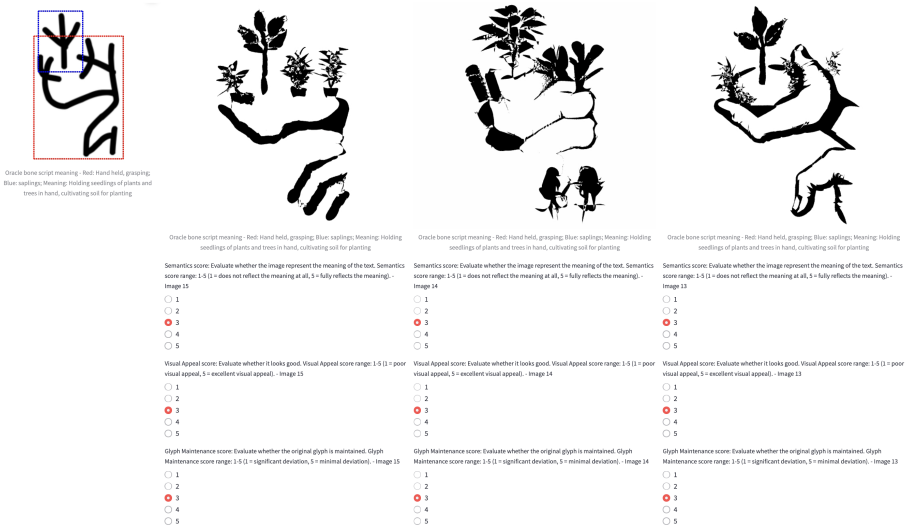

Figure 6. The description of our user study.

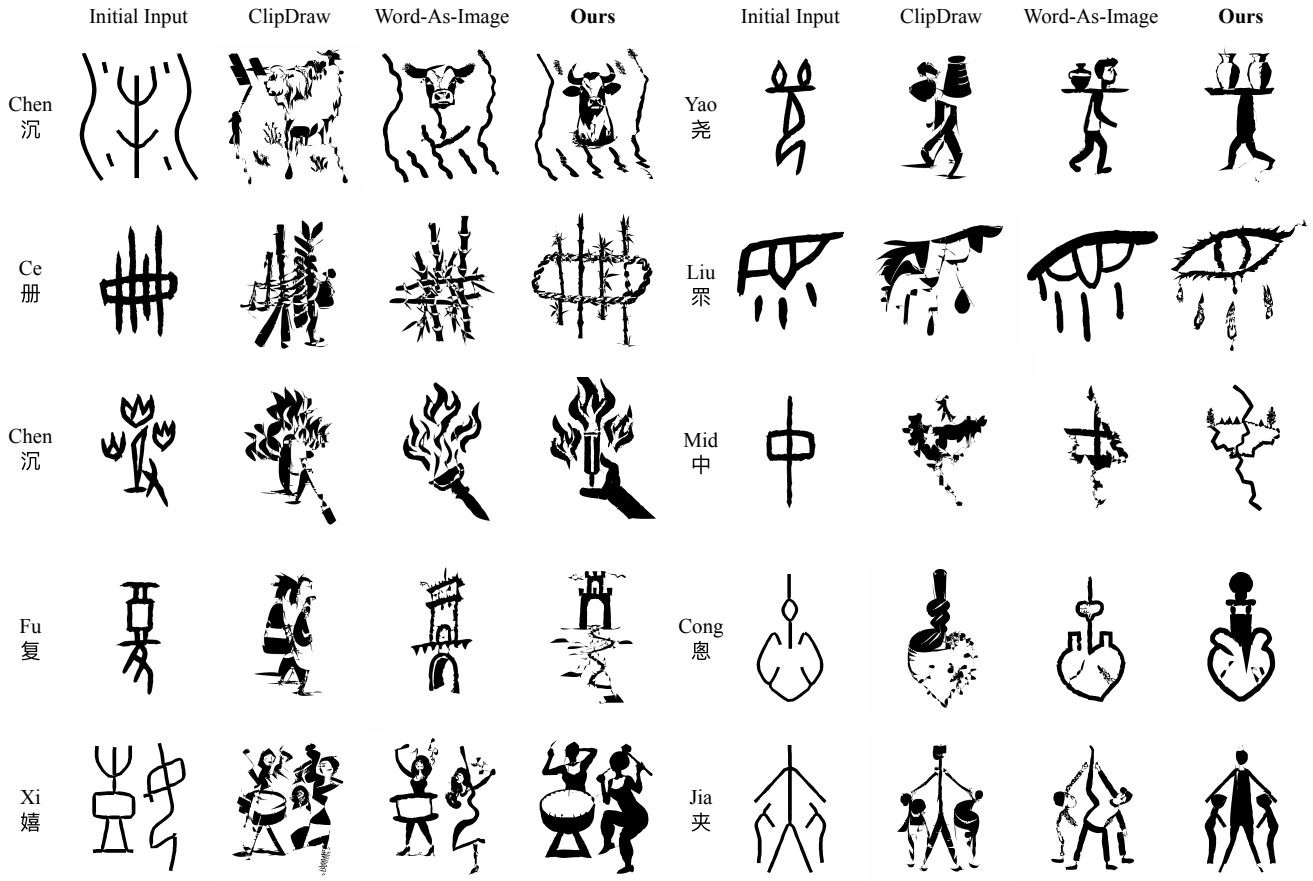

Figure 7. More qualitative comparison results between ClipDraw, Word-As-Image, and our OracleFusion in our user study.

References

[1] Kevin Frans, L.B. Soros, and Olaf Witkowski. Clipdraw: Exploring text-to-drawing synthesis through language-image encoders. *Cornell University - arXiv,Cornell University - arXiv*, 2021. 1

[2] Shir Iluz, Yael Vinker, Amir Hertz, Daniel Berio, Daniel Cohen-Or, and Ariel Shamir. Word-as-image for semantic typography. *ACM Trans. Graph.*, 42(4), 2023. 1

[3] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1, 2

[4] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models, 2023. 1, 3