Diffusion-based Image Generation of Oracle Bone Inscription Style Characters

2310069 Xiaoxuan Xie

Oracle bone inscriptions, a corpus of ancient Chinese script carved onto animal bones and turtle shells, constitute one of the most valuable cultural assets in understanding the early formation of Chinese civilization. Dating back more than three millennia (circa the Shang Dynasty, c. 1600–1046 Before the Common Era), these inscriptions embody proto-forms of Chinese characters and reflect the beliefs, rituals, historical records, and sociopolitical structures of the period.

The distinct pictographic nature of oracle bone inscriptions, where glyphs represent objects or concepts, makes them an invaluable resource for archaeologists, epigraphers, historians, and artists. However, the comprehensive digital analysis, stylistic rendering, and generation of oracle bone inscription-style images present profound challenges. These challenges stem from limited datasets, intricate visual features, and the need to translate modern concepts into the archaic and stylistically rich oracle bone inscription's visual language.

However, applying state-of-the-art generative artificial intelligence (AI) to oracle bone inscriptions is nontrivial. Transformation of a modern object image into an oracle bone inscription-inspired glyph requires modeling a unique aesthetic that is neither purely symbolic nor entirely representational. oracle bone inscription glyphs often combine ideographic and pictographic elements, implying that their stylistic formation depends on both the object's semantic meaning and visual representation. Unlike conventional style transfer tasks that apply superficial filters or patterns, oracle bone inscription style transformation demands fidelity to ancient carving techniques, line thickness variations, spatial composition rules, and subtle textural cues that reflect inscription on hard surfaces rather than ink on paper. Moreover, the historical erosion of numerous oracle bone inscription samples introduces additional noise and uncertainty to the visual features.

To solve these issues, this thesis introduces a generation pipeline based on a diffusion model specifically tailored to generate images in the style of oracle bone inscriptions. The proposed approach builds upon three key components: (1) constructing a domain-specific dataset aligning ancient oracle bone inscription references, textual descriptions, and contemporary object images, it contains 44 categories of oracle bone inscription and 180 sets of data pairs; (2) fine-tuning a diffusion model enhanced by ControlNet to achieve controllable oracle bone inscription-style image generation aligned with both shape and semantic intent; and (3) refining generative outputs to better adhere to

the structural norms, carving patterns, and stylistic conventions of authentic oracle bone inscriptions. Evaluations using IP-Adapter, pix2pix, and CycleGAN demonstrate that the proposed method achieves superior results in generating semantically consistent oracle bone inscription-style images.

Moreover, the proposed method is evaluated with the IP-Adapter, pix2pix, and CycleGAN. In qualitative evaluation, this work shows excellent performance in reconstructing existing oracle bone inscriptions, generating new characters, and generating diverse stylistic variants. In quantitative evaluation this work achieves optimal scores in Fréchet Inception Distance, CLIP Image-Image Similarity, and Neural Image Assessment. In the user preference study, 44% of the users preferred the results generated by this work and also obtained the highest scores for original image similarity. All the evaluations show that this method generates semantically consistent oracle bone inscription-style images.