https://doi.org/10.1038/s40494-026-02417-7

OBI designer: zero-shot oracle bone inscription artistic characters generation with multimodal style transfer

Jiahao Zhang[1,3], Fei Deng[2,3], Jiang Yuan[1], Chenjun Xu[1], Guang Long[1], Ruiyuan Li[1] & Shanxiong Chen[1] ✉

Oracle Bone Inscriptions (OBI) artistic character generation combines ancient characters with artificial intelligence to create artworks with cultural heritage and visual impact. Current creation relies on specialized, time-consuming manual skills. We propose OBI-Designer, a zero-shot method to automatically create OBI art characters that relies on pre-trained Vision-Language Models. OBI-Designer uses a two-stage pipeline: glyph synthesis and texture synthesis. The glyph synthesis stage introduces a semantic-driven image vectorization method based on Interval Score Matching (ISM) to optimize character outlines, using additional loss functions to ensure readability. The texture synthesis stage stylizes the glyphs using ControlNet and LoRA fine-tuning. Extensive experiments demonstrate our method's effectiveness, achieving a Top-1 recognition accuracy of 93.8% and ranking highest in user studies (4.20 out of 5). OBI-Designer further promotes the combination of oracle art synthesis and artificial intelligence.

Oracle Bone Inscriptions(OBI), also known as "Yinxu Script, were unearthed in Xiaotun Village, Anyang, Henan Province. They are cultural relics from the late Shang Dynasty, named for their practice of being inscribed on turtle shells and animal bones. These inscriptions represent the oldest discovered form of Chinese writing[1]. As the origin of Chinese characters, they not only play a significant role in etymological research but also contribute to the understanding of the culture and history of the Shang Dynasty, ancient China, and even the world. Known for their ideographic nature, each character in OBI conveys a unique meaning and image[2]. With simple yet vivid strokes, they exhibit strong artistic beauty and symbolic significance. The characters in OBI are highly variable in form, with the same character often appearing in different shapes across various texts. Although no longer used for everyday writing, OBI continues to thrive in contemporary art through its inheritance and innovation in calligraphy and seal carving[3]. Many calligraphers and seal engravers draw inspiration from OBI to create numerous remarkable artworks[4].

The artisticization of OBI refers to the re-creation and expression of this ancient writing form through artistic creation, so that it not only retains the original historical and cultural connotations but also gives it new artistic vitality. The artisticization of OBI is a fusion of traditional culture and modern art. It not only continues the historical value and cultural connotations of OBI but also gives it new vitality through innovative forms of expression. This artistic process not only enriches people's aesthetic experience and enhances their ability to appreciate art, but also serves as an educational tool to help the public understand and appreciate ancient civilizations more intuitively. At the same time, it promotes cultural exchanges, enhances national cultural identity, contributes cultural products to economic development, and raises society's awareness of the protection of traditional cultural heritage.

At present, there are many challenges in the artisticization of OBI: 1) Digital data of OBI is relatively scarce, and the annotation cost is high, which makes it difficult to meet the needs of deep learning model training. 2) For the stylization of character, the stylized OBI must still retain the original characteristics of the OBI to ensure their historical and cultural authenticity. Finding a balance between artistry and readability is one of the core difficulties of the stylization task. 3) Artistic feeling is a subjective feeling, and it is difficult to use quantitative indicators to describe the artistic content of an image.

In the broader field of artistic text generation, methods have evolved from GAN-based approaches[5–8] to vector domain optimization[9,10]. Recent works like Word-As-Image[9] leverage differentiable vector renderers DIFFVG[11] and guidance from pre-trained diffusion models[12–20]. However, their reliance on guidance techniques like Score Distillation Sampling (SDS) often produces overly smooth or blurry results, which are unsuitable for the detailed, pictographic nature of OBI. This highlights a critical need for a method that can harness the power of modern generative models while preserving the sharp fidelity of glyphs.

[1]College of Computer and Information Science, Southwest University, Chongqing, China. [2]School of humanities, HaiNan University, Hainan, China. [3]These authors contributed equally: Jiahao Zhang, Fei Deng. ✉e-mail: csxpml@163.com

Due to the lack of OBI datasets, this paper proposes a zero-shot learning method to achieve the creation of OBI (OBI-Designer), which gets rid of the dilemma of needing a large number of training samples. OBI-Designer mainly consists of two parts, namely glyph synthesis and texture synthesis. In glyph synthesis, this paper introduces a semantic-driven image vectorization method based on ISM[21], which vividly expresses text concepts by optimizing character contours, and uses additional loss functions to ensure the readability and consistency of characters. As a result, the optimized glyphs retain the basic structure of OBI while incorporating semantic features, achieving a balance between artistry and readability. Texture synthesis is based on ControlNet[22] and LoRA[23] technologies to stylize the results of glyph synthesis, and by incorporating richer textures, the final more artistic Oracle font art can be obtained. Figure 1 shows the optimization effect of the proposed method on the OBI twelve zodiac characters.

The main contributions of this paper are as follows:

1. We employ a cubic Bezier curve generation method based on adaptive control points to construct a high-quality dataset of OBI glyphs with more refined control point distribution, providing a solid data foundation for the research on OBI artistic characters synthesis.
2. We propose a zero-shot learning framework, OBI-Designer, for the generation of OBI artistic characters. This framework integrates glyph synthesis and texture synthesis, enabling high-quality generation of OBI artistic characters without relying on paired training datasets. It further promotes the integration of artificial intelligence with traditional cultural art.
3. We introduce a glyph synthesis method for OBI artistic characters that combines a Differentiable Vector Graphics Renderer (DIFFVG) with ISM. By optimizing character outlines and driving the process with semantics, this method ensures that the generated glyphs are both artistically appealing and legible.
4. Leveraging the powerful generative capabilities of pre-trained diffusion models, we propose a texture synthesis method based on ControlNet and LoRA fine-tuning techniques. This method allows for fine-grained stylization of the glyph generation results, producing OBI artistic characters with rich textures and artistic expressiveness.

Methods

Technical preliminaries

Our approach builds upon several key technologies in vector graphics, diffusion models, and text-to-image synthesis.

Differentiable Vector Graphics Rendering refers to the process of converting vector graphics, which are composed of control points and paths, into raster images. Traditional rasterization methods (such as OpenGL[24]) can efficiently render vector graphics; however, their discretized sampling process leads to the loss of gradient information, making it difficult to support gradient-based optimization. To address this issue, Li et al. proposed DIFFVG in 2020, which for the first time achieved end-to-end differentiable rendering from vector path parameters to pixel space. This technology parameterizes vector primitives such as Bezier curves as optimizable variables, replacing traditional discrete sampling with analytical gradient computation. This enables tasks such as font design and image vectorization to directly optimize vector control points using neural networks.

Diffusion Models are a class of deep generative models inspired by thermodynamics. Classic diffusion models, represented by DDPM[25], consist of two key stages: a forward process where noise is gradually added to data, and a reverse process where a neural network is trained to progressively remove that noise, thereby reconstructing data. While capable of producing high-quality results, this process is computationally intensive. Denoising Diffusion Implicit Models (DDIM[26]) significantly improved efficiency by employing a non-Markovian forward process and a deterministic reverse sampling strategy that intelligently skips steps. DDIM can achieve high-quality generation in fewer steps, laying the foundation for practical models such as Stable Diffusion.

Text-to-Image Generation[27–31] has been revolutionized by these models. Rombach et al. introduced the Stable Diffusion model[32], which innovatively enabled intelligent generation of high-quality images from text descriptions. This model uses a text encoder to transform prompts into semantic conditions that guide the diffusion process in a compressed latent space. The core mechanism involves a U-Net[33]-structured diffusion model that, guided by text semantics, iteratively denoises random noise to match the input descriptions. However, training such models is costly. To address this, Poole et al. proposed Score Distillation Sampling (SDS), which leverages a pre-trained diffusion model (like Stable Diffusion) to guide an optimization process (e.g., generating a 3D model or vector graphic). The core idea is to optimize the parameters of the generated image by backpropagating gradients from the diffusion model to align with a given text prompt. However, this method can lead to overly smooth generated images. To mitigate this issue, LucidDreamer introduced Interval Score Matching (ISM), which inverts the diffusion process using DDIM to generate reversible diffusion trajectories. By matching scores between two interval steps, ISM avoids the high reconstruction errors of single-step reconstruction. This approach enables ISM to provide consistent and high-quality pseudo-ground truths, generating outputs with rich details.

Construction of the oracle font glyph dataset (OFGD)

The Oracle Bone Font Glyph Dataset (OFGD) constructed in this paper is based on the "HanYi Chen Style Oracle Bone Inscriptions" font library. This font library contains 3665 commonly used OBI characters and provides vector contour information in the TrueType format. TrueType fonts precisely describe glyph shapes through control points and curves, ensuring consistent display across different devices. However, the limited number of control points in the original font files may affect the precision of glyph transformations. Therefore, preprocessing of the glyphs is necessary to convert them into cubic Bezier curves and save them in SVG format. This process not only increases the number of control points but also enhances the smoothness of the curves by subdividing long curve segments, thereby providing a higher-quality geometric representation for subsequent font synthesis.

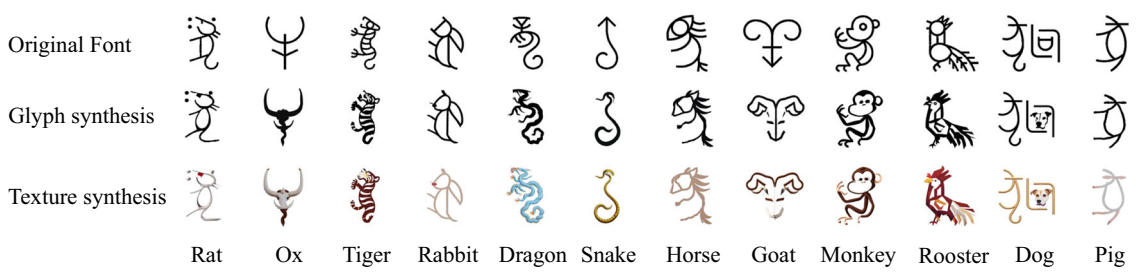

Fig. 1 | The optimization results of our method on the twelve zodiac characters of the OBI. The first row is the original oracle character, the second row is the glyph synthesis result, and the third row is the texture synthesis result.

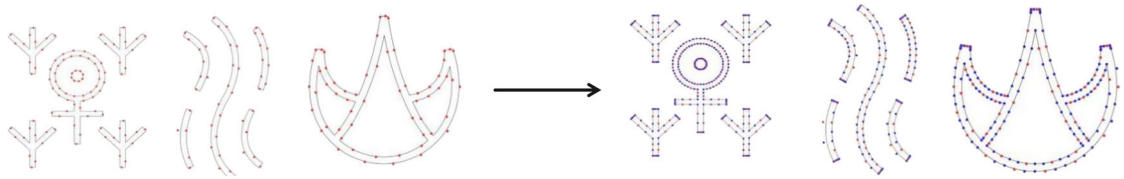

Fig. 2 | The control point diagrams of OBI glyphs before(left) and after(right) preprocessing. The left shows the original glyph control points, and the right shows the optimized control points.

Algorithm 1. proposes a method for generating cubic Bezier curves based on adaptive control points, aiming to balance glyph fidelity with computational efficiency. The algorithm initially loads the vector contour data of the target character through a font rendering engine (such as FreeType) and parses it into an initial set of cubic Bezier curves. It then initializes an empty list of Bezier curves and calculates the current total number of control points. By dynamically setting a subdivision threshold, the algorithm recursively subdivides curve segments that exceed the threshold in length (based on the De Casteljau[34] algorithm), gradually increasing the density of control points. During the subdivision process, the algorithm strictly maintains the topological structure of the original glyph and ensures refined modeling in critical areas (such as regions of high curvature) through a closed-loop feedback mechanism, while avoiding global oversampling. Ultimately, the algorithm outputs a set of Bezier curves that meet the control point constraints and saves each OBI character glyph in SVG format. In Algorithm 1, the input *num* is set as the dynamic stopping criterion. We first calculate the number of control points in the original TrueType glyph, denoted as P_{init}. The iterative subdivision process of Algorithm 1 (line 5) continues until the current total number of control points $n_{control}$ reaches twice the original number. We find that this threshold of $P_0 = 2P_{init}$ provides sufficient flexibility for subsequent glyph transformations while avoiding redundant calculations caused by oversampling. Figure 2 illustrates the change in control points of OBI character glyphs before and after preprocessing; the control points of the preprocessed glyphs are more densely distributed, providing greater flexibility and accuracy for subsequent glyph transformations and design.

Algorithm 1. Generation of cubic Bezier curves based on adaptive control points

Input: Font Files *font* ; character *c* ; Number of control points $num = 2P_{init}$

Output: Bezier curve collection *beziers*(Satisfy control point constraints)

1: Load the font file and extract the glyph outline data of character *c*

2: Initialize an empty list of Bezier curves *beziers*

3: Decompose the glyph outline into a set of initial cubic Bezier curves *beziers*

4: Calculate the current total number of control points:

$n_{control} = \sum(\text{len(curve)}|\text{curve} \in beziers)$

5: while $n_{control} < num$ do:

6: longest = max(curve, length) // Calculate the maximum length of a curve segment)

7: thresh = longest × 0.5 // Set dynamic segmentation threshold

8: Subdivide line segments longer than *thresh* to generate denser control points

9: Update the number of control points $n_{control}$

10: end while

11: return *beziers*

To generate artistic OBI characters that balance both readability and aesthetic quality, we propose OBI-Designer, a two-stage font generation framework. As illustrated in Fig. 3, our method consists of:

Glyph Synthesis (Step 1): Producing a preliminary artistic OBI font image.

Texture Synthesis (Step 2): Refining the output from Step 1 and applying texture synthesis.

The technical details of each stage are elaborated in the following sections.

Glyph synthesis

Pipeline: Glyph synthesis is based on Stable Diffusion Model and DIFFVG, as shown in step 1 in Fig. 3. This framework combines vector graphics rendering, latent diffusion model and conditional generation technology, and can efficiently generate high-quality oracle bone script art characters. We define the control point set of a single oracle bone script glyph as *P*. The goal of this method is to generate a new set of control points $\hat{P}$ given the control point set *P*. and the corresponding text prompt *y* (such as "head of a dog"), and generate optimized oracle bone script glyphs based on this. The optimized glyphs must meet the following two conditions: first, they can effectively convey the given text concept *y*; second, the overall structural characteristics and visual recognizability of the original oracle bone script glyphs are maintained during the optimization process.

We first initializes a learnable control point set $\hat{P}$ from the control point set *P* and inputs it into DIFFVG. The renderer converts the control point set $\hat{P}$ into a rasterized oracle bone image $\hat{I}$ Subsequently, the selected oracle bone glyph area is randomly enhanced, and the enhanced image is input into the pre-trained Stable Diffusion model. The encoder in the model is used to encode the input image x_0 into a latent feature z_0. At the same time, the text encoder of CLIP is used to encode the given text concept y (such as "head of a dog") as a guiding condition. The ISM Loss is used to guide the rasterized image ˆI to semantically align with the text concept *y*, thereby achieving glyph optimization.

In order to ensure that the optimized OBI does not undergo significant morphological distortion and maintains its recognizability, this paper introduces two additional loss functions to constrain the optimization process: ACAP loss and Tone loss. ACAP loss constraints the angle of shape deformation in the local area of the optimized glyph to remain unchanged as much as possible, so that the deformed shape still maintains a similar geometric structure, thereby retaining the overall morphological characteristics and style consistency of the font. Tone loss mainly constrains the $\mathscr{L}_2$ loss of the font image under low-pass filtering before and after optimization to retain the overall structure and visual consistency of the font.

ISM Loss: We introduce the ISM Loss for oracle bone glyph optimization, which incorporates DDIM inversion to generate high-quality pseudo-labels. The DDIM inversion constructs stable trajectories to address pseudolabel inconsistency, avoiding feature blurring and averaging. Its fast sampling and deterministic properties enable the inversion process to generate precise and stable pseudo-labels, enhancing the robustness and convergence accuracy of the ISM loss function, thereby providing reliable gradient guidance for glyph optimization.

Specifically, given an initial OBI glyph image z_0,we encode it into a latent vector z_0. The DDIM inversion iteratively generates a series of noisy latent states $\{z_{\delta_T}, z_{2\delta_T}, \ldots, z_t\}$, where t represents the timestep and δ_T denotes the timestep interval. Let ϵ_ϕ be a pre-trained denoising model (e.g., Stable Diffusion) that takes the current noisy state z_s,timestep s,and conditional vector y (encoded from text using CLIP) as input, and outputs the

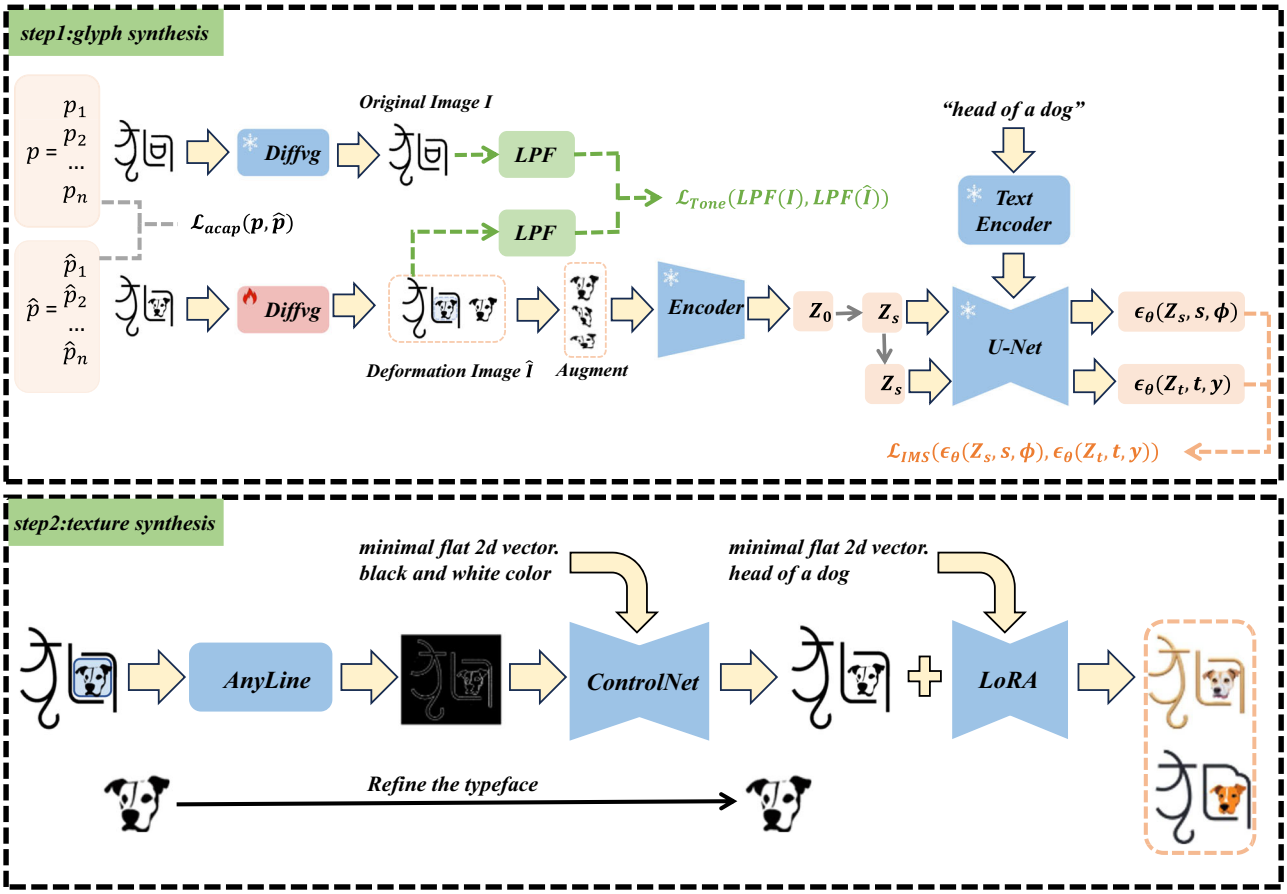

Fig. 3 | Overview of OBI-Designer. The entire framework consists of two stages. In the first stage, glyph generation is performed, using DIFFVG for differentiable rasterization combined with a diffusion model for semantic supervision. In the second stage, texture synthesis is carried out, applying the Control and LoRA methods for contour refinement and texture rendering.

predicted noise vector. The core DDIM inversion formula is:

$$\hat{z}_t = \sqrt{\alpha_t}\,\hat{z}_0^s + \sqrt{1-\alpha_t}\epsilon_\phi(z_s, s, y) \tag{1}$$

where $s = t - \delta_T$ represents the previous timestep, α_t is a predefined noise schedule coefficient, and $\hat{z}_0^s$ denotes the estimated initial image from the current state z_s. The initial image estimate can be reconstructed via the denoising model's predicted noise $\epsilon_\phi(z_s, s, y)$:

$$\hat{z}_0^s = \frac{1}{\sqrt{\alpha_s}} z_s - \gamma(s)\epsilon_\phi(z_s, s, y) \tag{2}$$

where $\gamma(s) = \sqrt{(1-\alpha_s)/\alpha_s}$ balances the weights between noise and reconstructed signal.

One of the core improvements of ISM is the replacement of single-step pseudo-label estimation with a multi-step DDIM denoising process, thereby generating $\hat{z}_0^s$ that can be decoded into higher-quality pseudo-labels. Additionally, an improved gradient update formula is derived based on the SDS Loss.

$$\nabla_\theta L(\theta) = \frac{\omega(t)}{\gamma(t)}(z_0 - \hat{z}_0^t)\frac{\partial g(\theta, y)}{\partial \theta} \tag{3}$$

Here, z_0 denotes the latent vector obtained by encoding the real initial image, $\hat{z}_0^t$ represents the pseudo-label generated through the multi-step denoising process, $\omega(t)$ is the weighting function used to adjust the loss contributions at different time steps, and $g(\theta, c)$ is the generative model with parameters θ and input condition y.

However, directly estimating $\hat{z}_0^s$ requires multiple calls to the model ϵ, which makes the calculation very time-consuming. Therefore, ISM Loss does not directly match the pseudo-reference $\hat{z}_0^s$ with the true initial latent vector, but matches it in two interval steps of the diffusion trajectory. It generates noisy latent states z_s and z_t through DDIM inversion, and uses the conditional denoising model ϵ_ϕ to calculate the noise prediction difference between the two as the optimization target. When given a text concept y, its loss function is defined as:

$$\mathcal{L}_{ISM} = E_t, y[\omega(t)\| \epsilon_\phi(z_t, t, y) - \epsilon_\phi(z_s, s, \varnothing) \|^2] \tag{4}$$

Here, $\epsilon_\phi(z_t, t, y)$ is the conditional denoising model, with the input being the noise latent state z_t, time step t and text concept y. $\epsilon_\phi(z_s, s,)$ is the unconditional denoising model, with the input being the noise latent state z_s and time step s. $\omega(t)$ is the associated weight function used to adjust the loss contribution of different time steps. In order to ensure that the generated artistic characters still maintain the characteristics of readability, in addition to $\mathcal{L}_{ISM}$, we also introduce two loss functions to limit the degree of font deformation.

ACAP Loss: The core principle of the As-Conformal-As-Possible Loss[35] (ACAP Loss) is to ensure that the geometric deformation of a font remains as conformal as possible to its original shape by comparing the triangulation results of the original and deformed fonts. Specifically, both the original font and its deformed counterpart undergo Delaunay[36] triangulation based on their respective control point sets P and $\hat{P}$, as illustrated in Fig. 4. This process decomposes the font contours into a set of non-overlapping triangles that collectively cover the entire glyph.

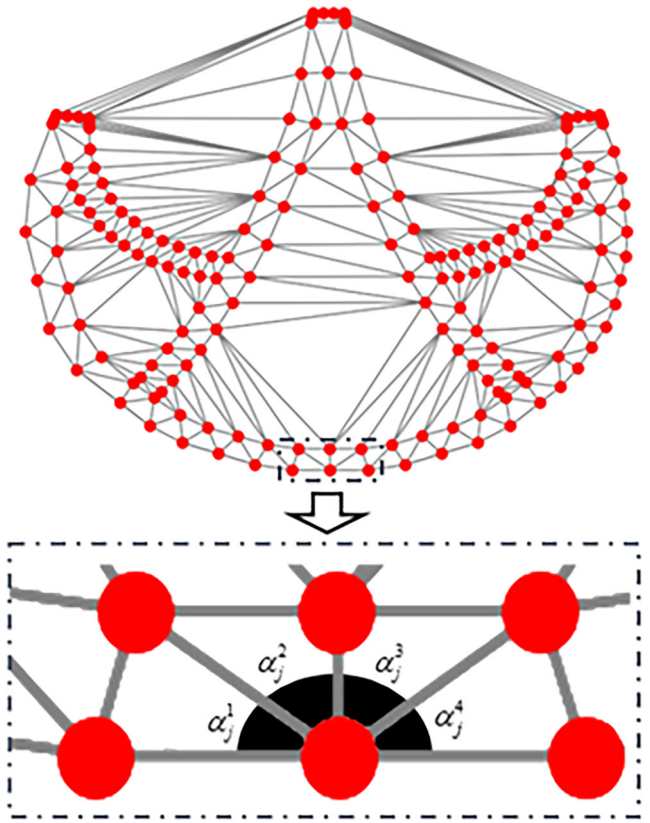

Fig. 4 | Delaunay Triangulation Schematic of the OBI Character "fire". The red points indicate the control points.

Fig. 5 | Low-pass filtering on the original font image (2 on the left) and low-pass filtering on the deformed font image(2 on the right). The two images on the left show the original oracle bone character and the low-pass filtered version, respectively. The two images on the right show the synthesized glyph and its low-pass filtered version, respectively.

The goal of ACAP Loss is to constrain the geometric differences of corresponding triangles in the font before and after deformation. It measures the geometric changes before and after deformation by comparing the internal angles of the triangles around each control point. For each control point P_j (where j = 1,2,…,k, k is the total number of vertices after triangulation), we calculate the set of internal angles of all triangles with it as a vertex $\{\alpha_j^i\}_{i=1}^{m_j}$), where m_j is the number of triangles with P_j as a vertex. Similarly, the corresponding set of internal angles of the deformed font is $\{\alpha_j^i\}_{i=1}^{m_j}$. The loss function of ACAP Loss is defined as:

$$\mathscr{L}_{acap}(P,\hat{P}) = \frac{1}{k}\sum_{j=1}^{k}\left(\sum_{i=1}^{m_j}(\alpha_j^i - \hat{a}_j^i)^2\right) \tag{5}$$

Here, k is the total number of vertices after segmentation, a_j^i and $\hat{a}_j^i$ represent the internal angles of the i-th triangle with control points P_j and $P_{j_i}^i$ as vertices in the original font and the deformed font, respectively. This loss function ensures that the font maintains the similarity of geometric structure during the deformation process by minimizing the difference in the internal angles of the corresponding triangles before and after deformation, thereby preventing readability from being lost due to excessive deformation.

Tone Loss: In order to better preserve the overall structure and visual consistency of the font, we use a loss function based on low-pass filtering and mean square error, Tone Loss. First, the original font and the artistic font image are low-pass filtered to extract their low-frequency information, that is, the overall structure and contour of the image, as shown in Fig. 5, and then the mean square error between the two is calculated to constrain the global geometric consistency in the style transfer process. The definition of Tone Loss is as follows:

$$\mathscr{L}_{tone}(I,\hat{I}) = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\left(LPF(I)_{(i,j)} - LPF(\hat{I})_{(i,j)}\right)^2 \tag{6}$$

Among them, I and $\hat{I}$ are the raster images obtained by passing DIFFVG to the vector font, H and W represent the height and width of the raster image, and LPF is the low-pass filter. By minimizing this loss function, it can be ensured that the overall structural information of the original font is retained during the stylization process of the artistic font, thereby avoiding font distortion or loss of readability due to excessive stylization. In addition, the use of low-pass filtering makes the loss function insensitive to high-frequency noise and detail changes, and focuses more on the constraints of global geometric consistency.

Dynamic weight strategy: Since $\mathscr{L}_{ISM}$ drives the font to deform into an artistic style, while $\mathscr{L}_{acap}$ and $\mathscr{L}_{tone}$ maintain the readability and structural stability of the font through geometric constraints and structural consistency, respectively. Therefore, assigning different weights to these three loss functions will directly affect the effect of generating artistic words: when the weight of $\mathscr{L}_{ISM}$ is large, the deformation of the font will be more significant and the artistic effect will be stronger; when the weights of $\mathscr{L}_{\text{tone}}$ and $\mathscr{L}_{\text{acap}}$ are large, the font is more inclined to maintain the original structure and readability. In order to balance this adversarial relationship, we designed an overall loss function that combines the three weights:

$$L = \lambda_1 L_{ISM} + \lambda_2 L_{acap} + \lambda_3 L_{tone} \tag{7}$$

λ_1, λ_2 and λ_3 are the weight coefficients of $\mathscr{L}_{ISM}$, $\mathscr{L}_{acap}$ and $\mathscr{L}_{tone}$ respectively. Based on experience, we set $\lambda_1 = 1$ to ensure the driving force of the artistic deformation; for λ_2, we referenced the setting of the conformal constraint in Word-As-Image and fixed it to $\lambda_2 = 0.5$ to ensure that the strength of the geometric constraint is moderate. We found that if λ_3 is too small, the effect of the tone consistency constraint will become less obvious; but if λ_3 is too large, especially in the early stages of training, it may over-restrict the deformation strength of the font. To solve this problem, we adopted a strategy of dynamically adjusting λ_3 so that it changes with the training time t. The change of λ_3 is described by a function:

$$\lambda_3(t) = a * exp(-\frac{(t-b)^2}{2c^2}) \tag{8}$$

Among them, $a = 100$, $b = 200$, $c = 30$. The design of this function makes the value of λ_3 smaller in the early and late stages of training, thus allowing the font to be deformed to a greater extent in the early stages of training. In the middle stages of training (t is close to b) λ_3 reaches its maximum value, at which time the effect of $\mathscr{L}_{tone}$ constraint is most significant, which helps to further optimize the quality of generated art fonts

Texture synthesis

Algorithm 2. Oracle Art Word Texture Synthesis
Input: Initial Art Font Image I_{art}; Style Tips S
Output: The final generated artistic font imageI_{final}

Fig. 6 | Different stylized renderings. The first column on the left is the original OBI font, the second column is the OBI art font after step 1 in Fig. 3, and the remaining columns are the results generated by different style prompts. From left to right, the style prompts are paper-cutting, Chinese calligraphy, three-dimensional wood carving, pixel style, street graffiti, futuristic metallic glossy fluid lines, and neon lights.

1: Use the △AnyLine algorithm to extract line features from I_{art} and generate an edge map E

2: Structure retention phase Input

(1) E_{latent} = ControlNet Encoder(E) // Encoding edge maps into latent Character space.

(2) I_{struct} = ControlNet_Optimize(E_{latent}) // Optimizing the latent representation to generate structure-enhanced images)

3: Style transfer phase

(1) $\theta_{adapted} = \theta_{original} + \Delta\theta_{LoRA}$ // LoRA model parameter adaptation

(2) Cross-Attention Layer Fine-tuning:

• $W_{original} = U \cdot V^T$ // Decomposition weight matrix, among them U $\in R^{d_{model} \times r}$, V $\in R^{d_{ff} \times r}$, r $\ll \min(d_{model}, d_{ff})$

• $\min ||(U \cdot V^T)(X) - Y_{style}||^2$ // Optimization goals

(3) $I_{texture}$ = LoRA Generate(I_{struct}, S) // Generate a textured synthetic image

4: return $I_{texture}$

In order to generate more ornamental OBI art characters, we proposed a fusion method based on ControlNet and LoRA models, which achieved a high degree of preservation of the OBI structure and efficient migration of texture style. As a structured generation framework based on the conditional diffusion model, ControlNet effectively constrains the structural consistency of the generation process by introducing edge detection maps as conditional inputs. In the OBI scene, we use the AnyLine algorithm(Inspired by[37]) to extract the glyph contour information, and embed the edge map into the latent space of the diffusion model through the encoder of ControlNet to ensure that the geometric structure of the generated image is strictly aligned with the original OBI glyphs to avoid semantic distortion caused by deformation. At the same time, LoRA, as an efficient parameter fine-tuning technology, can adapt to specific tasks by low-rank decomposition of the weight matrix of Stable Diffusion, and only a small number of parameters need to be adjusted. In OBI texture migration, LoRA acts on the cross-attention layer of the diffusion model and uses a small amount of texture. data to fine-tune the model so that the generated texture not only conforms to the target style, but also retains the style characteristics of the original glyphs.

As shown in step 2 of Fig. 3, first, the initial word art image is input into AnyLine for preprocessing. This tool can accurately extract the lines and detail features in the image. Subsequently, the processed image is input into ControlNet, which uses its fine adjustment capabilities to optimize the clarity and detail of the lines and generate a font image with precise structure. Finally, the image output by ControlNet is input into the LoRA model, and its texture generation capability is used to add artistic texture to the image, so that the final word art can maintain both fine line features and rich artistic expression. This multi-stage processing framework ensures the dual improvement of word art in terms of glyph accuracy and artistic expression through the organic combination of structure preservation and texture migration. Algorithm 2 gives the detailed process of texture synthesis.

As shown in Fig. 1, the OBI-Designer proposed in this paper shows a multilevel generation effect in the design of the OBI twelve zodiac characters. The first row in the figure shows the original OBI glyph samples. These characters are used as a reference and reflect the original stroke structure and pictographic features of the OBI; the second row shows the results generated by the step1 glyph synthesis module. This stage is based on the semantic driven vectorization method of ISM. On the basis of strictly following the OBI word formation rules, it realizes the intelligent reconstruction of the glyph structure, so that the generated characters not only retain the original skeleton of the OBI, but also incorporate the semantic features of the text prompts (such as "head of a dog", "dragon"); the third row shows the final effect after processing by the step2 texture rendering module. Through the texture migration technology coordinated by ControlNet and LoRA, a richer texture style is superimposed on the glyph structure, while keeping the strokes clear and discernible. This hierarchical and progressive generation process of "structure-semantics-texture" not only fully reproduces the artistic characteristics of OBI, but also realizes the adjustability of artistic expression through parametric control - users can balance semantic expression and glyph fidelity by adjusting different loss weights, or choose different styles through the LoRA adapter, as shown in Fig. 6.

Results

Experimental setup

Dataset. We conducted OBI stylization experiments on the OFGD dataset constructed in *Construction of the Oracle Font Glyph Dataset (OFGD)* to verify the effectiveness of the proposed OBI-Designer framework. It should be noted that the "Hanyi Chenti Oracle Bone Script" font on which this dataset is based is a publicly released commercial font, and its use is subject to specific licenses. This study only uses the font within the "fair use" framework for academic research and algorithm verification. To avoid any potential copyright and distribution licensing issues, we will not currently publicly release OFGD datasets derived from this font.

Implementation. This paper uses the PyTorch deep learning framework to build the network. In the model optimization stage, the rasterized OBI glyph images are uniformly scaled to 600×600 size, and data enhancement methods such as random cropping and perspective

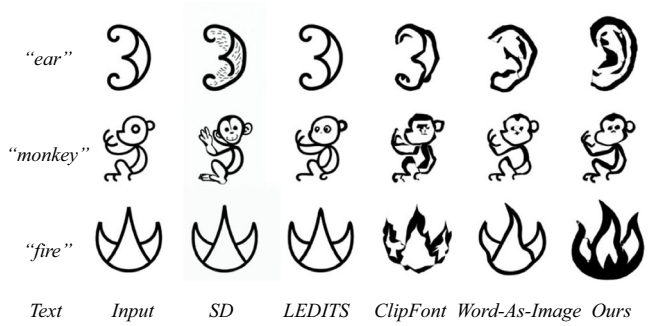

Fig. 7 | Schematic diagram of the experimental results compared with the text-to-image model. The first column shows the prompts, the second column shows the original oracle bone character images, and the remaining columns show the glyph synthesis results using different methods.

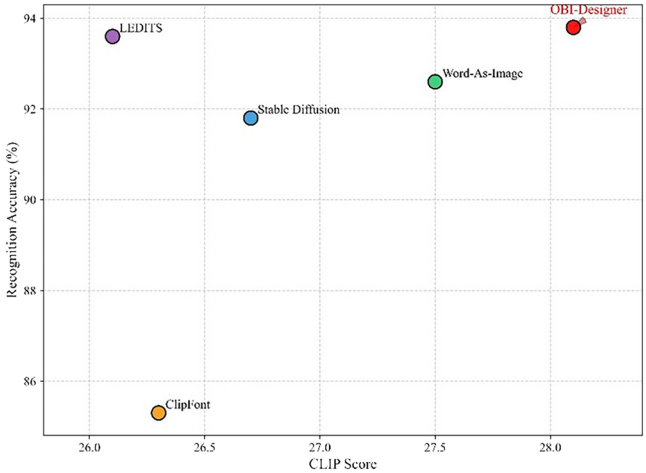

Fig. 8 | Comparative Analysis on the Readability-Semantic Alignment Frontier. Different points represent different methods.

transformation are used, and the Adam optimizer is used for parameter optimization. The initial learning rate is set to 0.002 and the minimum learning rate is set to 0.0008. The hardware configuration of the operating platform in this paper is an Nvidia GeForce RTX3090 GPU with 24GB video memory. In the texture synthesis stage, we fine-tuned the Stable Diffusion model using LoRA to adapt it to a specific texture style (as shown in Fig. 6). We adopted a common LoRA configuration: the rank r was set to 8, and the learning rate was set to 1×10^{-4}. During training, we only fine-tuned the cross-attention layer in U-Net. The model was trained for approximately 800 steps for each reference style image.

Evaluation metrics. As an important carrier of visual communication, artistic characters must strike a delicate balance between readability and artistry. To verify the feasibility and superiority of the proposed OBI-Designer, we utilize two critical metrics for quantitative analysis: Readability, assessed by Recognition Accuracy (Top-1/Top-3) measured using the YOLOv8[38] recognition model (which was trained on 40,000 original OBI images), and Semantic Alignment, assessed by CLIP Score[39], which measures the semantic relevance between the generated image and the textual prompt.

Comparison with different models

Due to the lack of baseline models specifically for OBI stylization tasks, this paper selects several mainstream text-guided text or image generation models as comparison methods, including Stable Diffusion (SD), LEDITS[40], ClipFont[41], and Word-As-Image. These models represent the latest progress in the fields of text-to-image generation (SD and LEDITS) and font synthesis (ClipFont and Word-As-Image). This paper applies these methods to three typical OBI—"ear", "monkey", and "fire", and conducts glyph optimization comparison experiments to evaluate their performance in the OBI stylization task.

To ensure fairness in the comparison, all baseline models used the same text prompts ("ear", "monkey", "fire") as our method (Fig. 7). For SD and LEDITS, we used only text prompts for guidance without providing any additional structural constraints. For ClipFont and Word-As-Image, we followed the settings in their original papers and used the same original OBI glyphs as our method as the starting point.The experimental results are shown in Fig. 7. The experimental results show that different models show significant differences in processing OBI stylization tasks.

For the character "monkey" (the second row of the experimental results), SD and LEDITS as text-to-image generation models can convey semantic information more accurately, but the generated character skeletons have a large deviation from the original OBI characters, resulting in distortion of the character structure. In contrast, ClipFont and Word-As-Image as font synthesis models can better preserve the basic skeleton of the original OBI characters, but they are insufficient in semantic communication and character quality. Specifically, the character contours generated by ClipFont are not smooth enough and the semantic expression is not clear enough; while Word-As-Image retains the skeleton, the accuracy of semantic communication still has room for improvement. In contrast, the method proposed in this paper achieves higher semantic accuracy and character quality on the basis of retaining the original character skeleton.

For the character "fire" (the third row of the experimental results), due to the weak correlation between its character structure and semantics, SD and LEDITS tend to maintain the stability of the original character during the generation process and abandon deformation, resulting in limited semantic communication effect. ClipFont and Word-As-Image try to make the glyph closer to the semantics through deformation, but the text contour generated by ClipFont is still not smooth and the semantic expression is not accurate enough; although Word-As-Image performs well in deformation, the matching degree between the generated text and the semantics is still insufficient. In the stylization task of the character "fire", the method in this paper successfully retains the original glyph skeleton while achieving an effective combination of semantics and glyph through reasonable deformation and texture optimization.

Figure 8 shows the comparison results between OBI-Designer and existing baseline methods (Stable Diffusion, LEDITS, ClipFont, Word-As-Image). The results clearly indicate that OBI-Designer achieved the highest CLIP Score (28.1), demonstrating its significant advantage in accurately integrating text semantics into glyph deformation. In particular, Word-As-Image (CLIP Score 27.5) also performed well in the semantic dimension, but was slightly inferior to OBI-Designer in readability (92.60% vs. 93.80%). In contrast, methods such as ClipFont sacrificed semantic alignment (CLIP Score 26.3) in pursuit of higher readability (85.30%). Overall, OBI-Designer achieved optimal semantic expression while maintaining acceptable readability (93.80%), effectively occupying the optimal frontier position on the trade-off curve, and strongly supporting the core proposition of this method to balance artistry and readability in zero-shot OBI character generation tasks.

User study

This study evaluated the performance of the above five models through user experiments. The experiment recruited 30 subjects with basic knowledge of ancient characters and conducted a multi-dimensional evaluation on the optimization results of 15 typical oracle bone characters. Each character corresponds to the optimized version of 5 models. The subjects were required to score from low to high (1-5 points) based on the integrity of the glyph structure, the accuracy of semantic representation, and visual recognizability.

Before the experiment, all participants received the same Rater Instructions, which clearly defined structural integrity (preserving the original oracle bone script skeleton), semantic accuracy (matching with the

Table 1 | Average user rankings of different method

Method	Average User Ranking
Stable Diffusion	1.02 ± 0.01
LEDITS	2.30 ± 0.57
ClipFont	3.29 ± 0.72
Word-As-Image	3.79 ± 0.23
Ours	4.20 ± 0.45

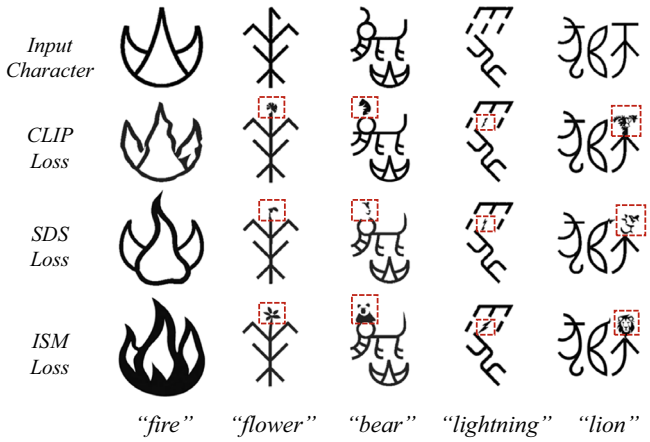

Fig. 9 | Comparison of the effects of OBI generated by different deformation loss functions. Red boxes indicate the optimized parts, while images without boxes indicate overall optimization.

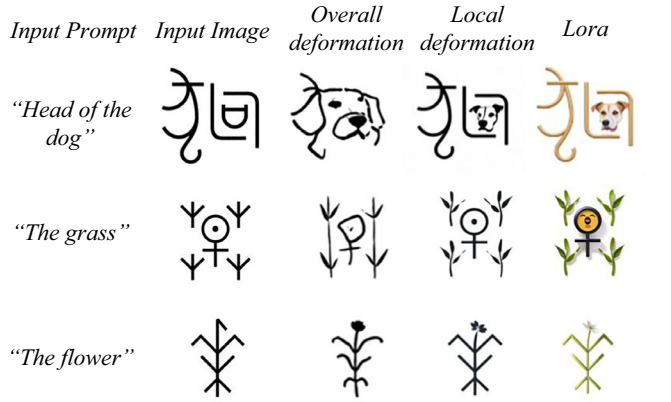

Fig. 10 | Comparison results of different deformation strategies. The first column shows the input prompts, the second column shows the original glyphs, the third and fourth columns show the results of different deformation strategies, and the last column shows the texture rendering results.

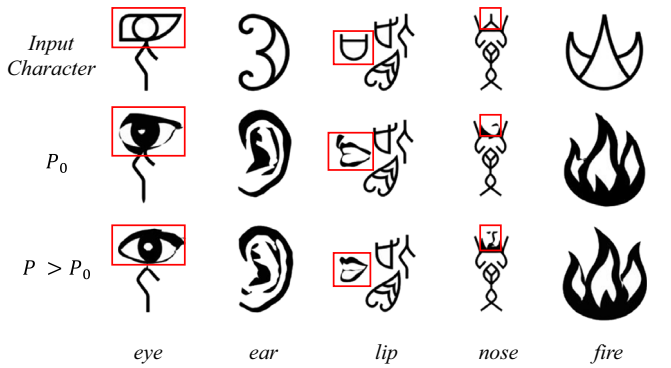

Fig. 11 | The generation effect of oracle art characters under different numbers of control points. Red boxes indicate the optimized parts, while images without boxes indicate overall optimization.

prompt words), and visual recognizability. To reduce bias, all model-generated images were randomized and anonymized, so participants did not know which model each image came from during the evaluation. To test the consistency of the evaluation, we calculated the inter-rater reliability and obtained Cronbach's $\alpha = 0.87$ which indicates good consistency in the scoring criteria among different raters.

The experimental data are shown in Table 1. The experimental results show that the method proposed in this paper has an average score of 4.20 in the comprehensive score, which achieved the highest average score among the models, verifying the effectiveness of the method in this paper in balancing the fidelity of glyphs and the expression of semantic concepts.

Comparison with different loss

This paper further compares three loss functions based on pre-trained large model distillation, including CLIP Loss, SDS Loss and ISM Loss, to evaluate their performance differences in the OBI stylization task. These loss functions represent different semantic guidance and optimization strategies, which can affect the quality of the generated results from different angles. This paper applies the above method to five typical oracle bone characters, namely "fire", "flower", "bear", "electricity" and "lion".

To ensure fairness in the comparison, all comparison experiments for loss functions were conducted under the same weight constraints: the semantic loss weight λ_1 was fixed at 1, the ACAP loss weightλ_2 was fixed at 0.5, and the Tone loss weight λ_3 used a dynamic $\lambda_3(t)$. The experimental results are shown in Fig. 9, where the red box indicates the pre-set optimization area. Fire is the overall structure optimization. The experimental results show that ISM Loss performs best in semantic expression and glyph coherence. Specifically, ISM Loss can accurately capture the semantic information of the text while ensuring the structural coherence and smoothness of the generated glyphs. For example, in the generated result of the word "lion, ISM Loss successfully combines the morphological characteristics of the lion with the structure of the OBI, which not only retains the overall skeleton of the glyph, but also clearly conveys the semantic information. In contrast, CLIP Loss has certain problems in semantic expression and glyph contour smoothness. For example, in the generated result of the character "lion", the outline of the lion's head appears to be messy and lacks coherence, resulting in a decrease in the quality of the glyph. This shows that CLIP Loss has limitations in dealing with the balance between complex semantics and glyph structure. On the other hand, SDS Loss is unable to provide accurate pseudo-labels during the generation process, resulting in incomplete semantic expression of some oracle bone characters. For example, in the generated result of the character "flower", SDS Loss failed to fully capture the morphological features of the flower, resulting in unclear semantic expression of the generated text. This result shows that the performance of SDS Loss is limited when dealing with oracle bone characters with richer semantic details.

Comparison with different deformation strategies

Figure 10 shows the results achieved by selecting different deformation regions. As can be seen from the three examples in the figure, globally deforming the entire font image destroys the balance between readability and artistry, resulting in poor generation results. On the contrary, the correct selection of deformation regions can generate ideal images while retaining the readability and artistic beauty of the font. However, for some specific characters (such as "pig" and "monkey" in Fig. 1), global deformation can achieve ideal results. This shows that the selection of deformation regions needs to be flexibly adjusted according to the shape and semantic characteristics of the source font. Therefore, the reasonable selection of deformation regions is the key to achieving high-quality artistic word generation,

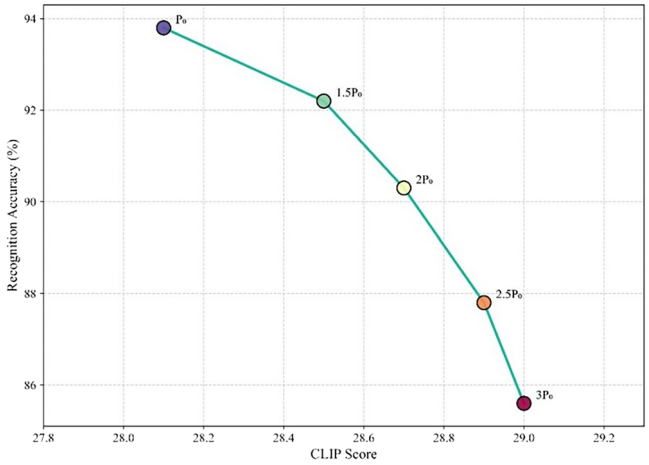

Fig. 12 | Ablation analysis of control point density. Different points represent the recognition accuracy of characters after optimization with different numbers of control points.

Table 2 | Performance comparison of OBI recognition models on different methods

Dataset	Top-1 ACC	Top-3 ACC
Original OBI	96.4%	99.1%
SD	91.8%	96.3%
LEDITS	93.6%	97.8%
ClipFont	85.3%	92.9%
Word-As-Image	92.6%	95.5%
Ours	93.8%	97.5%

which can significantly improve the satisfaction of the final generation effect. In addition, we show the effect of rendering the local deformation results by applying the LoRA method in the last column of Fig. 10. The generated image is more semantically in line with the requirements of the prompt word, further improving the visual expression and semantic consistency of the artistic word.

Ablation experiment

We experimentally compare the effects of different numbers of control points on the optimization results of OBI glyphs, including our target (baseline) number of control points P_0 and the increased number of control points of P. The experiment selected five typical OBI glyphs as research objects, namely "eye", "ear", "lip", "nose" and "fire". During the experiment, the area marked with a red box indicates that local optimization is required, while the area without a red box is optimized for the whole picture. The experimental results are shown in Fig. 11.

The experimental results show that in the local area optimization task, increasing the number of control points can significantly improve the flexibility and accuracy of optimization. Specifically, more control points make the glyphs more detailed in the local area, so as to better capture and convey the semantic concepts of the text. For example, in the local optimization of the word "nose", increasing the number of control points makes the outline and details of the eye clearer, significantly improving the accuracy of semantic expression. However, in the whole picture optimization task, the increase in the number of control points does not always bring positive effects. The experimental results show that too many control points may lead to the generation of redundant details, thus affecting the coordination and aesthetics of the overall glyph. For example, in the full-image optimization of the character "ear", after the number of control points increased, unnecessary details appeared in the glyph, making the overall structure redundant and not concise enough. This shows that the choice of

Table 3 | Performance comparison of OBI recognition models on different number of control points. CP means Control Points

Number of CP	Top-1 ACC	Top-3 ACC
P_0	93.8%	97.5%
$2P_0$	90.3%	93.6%
$3P_0$	85.6%	90.5%

the number of control points needs to be weighed according to the specific optimization task: local optimization tasks usually benefit from more control points, while full-image optimization may require a more cautious number of control points to avoid overfitting and the generation of redundant details.

To quantify the dynamic balance between the artistry (semantic alignment) and functionality (Top-1 readability accuracy) of OBI-Designer, we constructed a two-dimensional Pareto front analysis plot, where the X-axis represents the CLIP score and the Y-axis represents the recognition accuracy (Read Acc). Figure 12 shows the Pareto curves at different control point densities (from P_0 to $3P_0$), visually reflecting the relationship between control point density and model performance. The results reveal a clear trade-off: when the number of control points is at its maximum ($3P_0$), the model achieves the highest semantic alignment (CLIP score 29.0), but readability is relatively low (85.60%). As the control point density decreases to P_0, readability significantly improves to 93.80%, but at the cost of a decrease in semantic alignment (CLIP score 28.1). This inverse relationship provides valuable design guidance: for applications that prioritize semantic expression and artistic variation, the more control points the better; while for tasks that require maximum readability and structural preservation, a lower density of control points should be used.

Quantitative readability analysis

To quantitatively evaluate whether OBI-Designer maintains the readability of OBI after artistic processing, we designed a comparative recognition experiment. We tested on multiple datasets, including the original OBI dataset (Original OBI), artistic characters generated by our method (Ours), and output from several other mainstream generative models (such as SD and LEDITS). Recognition accuracy was used as the core metric for measuring readability.

The recognizer (YOLOv8) was trained on 40,000 raw OBI images (as described in the Evaluation metrics paragraph), with the dataset divided into an 80% training set, a 10% validation set, and a 10% test set. During the readability tests in this section (Table 2), the glyphs generated by all methods (Ours, SD, LEDITS, etc.) were uniformly processed into black-and-white rasterized images of the same size (256×256 pixels) to ensure fair matching of test conditions. As shown in Table 2, the model achieved a top-1 accuracy of 96.4% on the original OBI dataset, setting a performance benchmark for the experiment.

On our artistic character dataset (Ours), our model achieved a Top-1 accuracy of 93.8%, a slight difference of 2.6 percentage points from the original characters, which initially demonstrates the effectiveness of our method in maintaining readability. More notably, our method achieved a Top-3 accuracy of 97.5%, further narrowing the gap with the original oracle bone script (99.1%). Furthermore, compared to other state-of-the-art generative models in Table 2, our method exhibits excellent overall performance, particularly in maintaining structural integrity, outperforming models such as SD (91.8%) and ClipFont (85.3%).

To further explore the key factors influencing readability, we also conducted an internal comparative experiment on the number of control points, as shown in Table 3. The results reveal a clear trend: when the number of control points is at our target baseline $P_0(P_0 = 2P_{init})$, recognition accuracy is highest (93.8%), which is consistent with the performance of our method in Table 2. However, as the number of control points increases to $2P_0$ and $3P_0$, accuracy drops sharply to 90.3% and 85.6%,

respectively. This strongly demonstrates that while OBI-Designer achieves significant artistic enhancement, its ability to maintain high readability lies in precise control of the number of control points. The P0 setting successfully strikes a delicate balance between artistic expression and structural fidelity, ensuring the preservation of core features.

These results reveal an important trade-off between artistic expressiveness and glyph readability. As demonstrated in the ablation study, having more control points ($P > P_0$) provides greater transformational freedom and detail expressiveness for local regions of the glyph, which is crucial for realizing complex semantic concepts. However, this high degree of flexibility can also introduce redundant or excessive details to the global structure, potentially degrading the core features that a recognition model relies on. Therefore, while our method supports finer artistic creation by increasing control points, the experiments prove that for tasks requiring maximum readability, our target baseline number of control points (P_0) strikes the optimal balance between artistic expression and structural fidelity. This indicates that the choice of control point density is a task-dependent decision, validating both the robustness of our method in maintaining high readability and the potential of our constructed OFGD dataset to support diverse artistic creations.

To verify the statistical significance of the accuracy differences in Tables 2 and 3, we performed t-tests on the key results. The results showed that the improvement of our method (93.8%) compared to ClipFont (85.3%) was statistically significant ($p < 0.01$). Similarly, in Table 3, the accuracy difference between P_0 (93.8%) and $3P_0$ (85.6%) was also statistically significant ($p < 0.005$), providing strong statistical support for our conclusions.

Discussion

This paper constructs a more sophisticated OBI font glyph dataset, OFGD, which provides a data foundation for the artisticization of OBI. At the same time, a method for generating OBI artistic characters based on zero-shot learning, OBI-Designer, is proposed. Through the organic combination of glyph synthesis and texture synthesis, the creation of OBI artistic characters without the supervision of artistic character samples is realized. In the glyph synthesis stage, the semantic-driven image vectorization method based on ISM can effectively optimize the character contour and vividly express the text concept. At the same time, the readability and consistency of the generated font are ensured through an additional loss function. In the texture synthesis stage, the glyphs are stylized by combining the ControlNet and LoRA models, which further enhances the visual expression and artistry of the artistic characters. Experimental results show that OBI-Designer performs well in semantic expression, glyph detail retention, and artistic style generation, all of which are better than existing methods.

At the same time, we recognize the ethical responsibility involved in handling precious cultural heritage such as oracle bone inscriptions. While this method aims for artistic innovation, the generated results carry the potential risk of misuse or confusion with authentic historical inscriptions. Therefore, we advocate adding explicit watermarks or metadata to publicly release images generated by OBI-Designer to indicate their "AI-generated" attribute, thereby mitigating potential academic misuse. In the future, we will further explore the application of OBI-Designer in the generation of other ancient characters artistic characters, opening up more possibilities for the combination of traditional culture and modern technology.

Data availability

The datasets generated and/or analyzed during the current study are available in the (https://www.hanyi.com.cn/productdetail.php?id=2638).

Received: 10 October 2025; Accepted: 27 February 2026;
Published online: 13 March 2026

References

1. Gao, F. et al. Image translation for oracle bone character interpretation. *Symmetry* **14**(4), 743 (2022).
2. Zhang, C., et al. Data-driven oracle bone rejoining: a dataset and practical self-supervised learning scheme//Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining. 4482-4492 (2022).
3. Gao, W. et al. OBM-CNN: a new double-stream convolutional neural network for shield pattern segmentation in ancient oracle bones. *Appl. Intell.* **52**(11), 12241–12257 (2022).
4. Guan. H. et al. Deciphering oracle bone language with diffusion models//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics 1, 15554-15567 (2024).
5. Yang, S. et al. TET-GAN: Text effects transfer via stylization and destylization. *Proc. AAAI Conf. Artif. Intell.* **33**(01), 1238–1245 (2019).
6. Yang, S., et al. Controllable artistic text style transfer via shape-matching gan//Proceedings of the IEEE/CVF International Conference on Computer Vision. 4442-4451 (IEEE, 2019).
7. Goodfellow, I. et al. Generative adversarial networks. *Commun. ACM* **63**(11), 139–144 (2020).
8. Chen, F. et al. Style transfer network for complex multi-stroke text. *Multimed. Syst.* **29**(3), 1291–1300 (2023).
9. Iluz, S. et al. Word-as-image for semantic typography. *ACM Trans. Graph. (TOG)* **42**(4), 1–11 (2023).
10. Ma, X., et al. Towards layer-wise image vectorization//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 16314-16323 (IEEE, 2022).
11. Li, T. M. et al. Differentiable vector graphics rasterization for editing and learning. *ACM Trans. Graph. (TOG)* **39**(6), 1–15 (2020).
12. Poole, B. et al. Dreamfusion: Text-to-3D using 2d diffusion. arXiv preprint arXiv:2209.14988 (2022).
13. Zhang, L., Rao, A., Agrawala, M. Adding conditional control to text-to-image diffusion models//Proceedings of the IEEE/CVF international conference on computer vision. 3836-3847 (IEEE, 2023).
14. Zhang, F. et al. Improving font effect generation based on pyramid style feature. *Int. J. Perform. Eng.* **16**(8), 1271 (2020).
15. Huang, Q. et al. Gentext: Unsupervised artistic text generation via decoupled font and texture manipulation. arXiv preprint arXiv:2207.09649 (2022).
16. Xue, M., Ito, Y. & Nakano, K. An art font generation technique using Pix2Pix-based networks. *Bull. Netw., Comput., Syst., Softw.* **12**(1), 6–12 (2023).
17. Tanveer, M., et al. Ds-fusion: artistic typography via discriminated and stylized diffusion//Proceedings of the IEEE/CVF International Conference on Computer Vision. 374-384 (IEEE, 2023).
18. Amit, T. et al. Segdiff: Image segmentation with diffusion probabilistic models. arXiv preprint arXiv:2112.00390 (2021).
19. Avrahami, O., Lischinski, D., Fried, O. Blended diffusion for text-driven editing of natural images//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 18208-18218 (IEEE, 2022).
20. Xiao, S., et al. TypeDance: Creating semantic typographic logos from image through personalized generation//Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. 1-18 (2024).
21. Liang, Y., et al. Luciddreamer: Towards high-fidelity text-to-3D generation via interval score matching//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 6517-6526 (2024).
22. Nichol, A. et al. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. arXiv preprint arXiv:2112.10741 (2021).
23. Hu, E. J. et al. Lora: Low-rank adaptation of large language models. *ICLR* **1**(2), 3 (2022).
24. Angel, E. Open GL Primer[M]. (Addison-Wesley Longman Publishing Co., Inc., 2001).
25. Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. *Adv. neural Inf. Process. Syst.* **33**, 6840–6851 (2020).

26. Song, J, Meng, C. & Ermon, S. Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020).
27. Radford, A. et al. Learning transferable visual models from natural language supervision//International conference on machine learning. 8748-8763 (PmLR, 2021).
28. Gal, R. et al. An image is worth one word: Personalizing text-to-image generation using textual inversion. arXiv preprint arXiv:2208.01618 (2022).
29. Frans, K., Soros, L. & Witkowski, O. Clipdraw: exploring text-to-drawing synthesis through language-image encoders. *Adv. Neural Inf. Process. Syst.* **35**, 5207–5218 (2022).
30. Vinker, Y. et al. Clipasso: Semantically-aware object sketching. *ACM Trans. Graph. (TOG)* **41**(4), 1–11 (2022).
31. Hussein, A., et al. Khattat: Enhancing Readability and Concept Representation of Semantic Typography//European Conference on Computer Vision. 278-295 (Springer Nature Switzerland, 2024).
32. Ramesh, A. et al. Hierarchical text-conditional image generation with clip latents. arXiv preprint arXiv:2204.065, 2022, 1(2): 3.
33. Ronneberger, O., Fischer, P., Brox, T. U-net: Convolutional networks for biomedical image segmentation//International Conference on Medical Image Computing and Computer-assisted Intervention. 234-241 (Springer International Publishing, 2015).
34. Crouch, P., Kun, G. & Leite, F. S. The De Casteljau algorithm on Lie groups and spheres. *J. Dynamical Control Syst.* **5**(3), 397–429 (1999).
35. Hormann, K., Greiner, G. MIPS: An efficient global parametrization method. 2000.
36. Delaunay, B. Sur la sphère vide. *A la m.émoire de. Georges Voronoï. Известия Российской академии наук. Серия математическая* **6**, 793–800 (1934).
37. Soria, X., et al. Tiny and efficient model for the edge detection generalization//Proceedings of the IEEE/CVF International Conference on Computer Vision. 1364-1373 (IEEE, 2023).
38. Varghese, R., Sambath M. Yolov8: A novel object detection algorithm with enhanced performance and robustness//2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). 1-6 (IEEE, 2024).
39. Hessel, J., et al. Clipscore: A reference-free evaluation metric for image captioning//Proceedings of the 2021 conference on empirical methods in natural language processing. 7514-7528 (2021).
40. Tsaban, L. Passos, A. Ledits: Real image editing with ddpm inversion and semantic guidance. arXiv preprint arXiv:2307.00522 (2023).
41. Xiong, J., Wang, Y., Zeng, J. Clip-font: Sementic self-supervised few-shot font generation with clip//ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 3620-3624 (IEEE, 2024).

Acknowledgements

This work was supported by the Major Bidding Project of the National Social Science Fund, “Compilation of a General History of Chinese Grammar Based on Chronological Research” (Grant No. 24&ZD246). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

Jiahao Zhang and Fei Deng contributed equally to this work. Jiahao Zhang conceived the idea, designed and performed the experiments, and wrote the manuscript. Fei Deng provided critical guidance on the historical authenticity and cultural fidelity of the OBI glyphs. He co-designed the evaluation metrics for the user study, guided the selection of culturally significant OBI characters for the experiments, and validated the artistic stylization choices to ensure they maintained readability.Jiang Yuan and Chenjun Xu contributed to the implementation of the software and data curation. Guang Long and Ruiyuan Li assisted with the validation and visualization of the results. Shanxiong Chen supervised the project and reviewed & edited the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Shanxiong Chen.