npj | heritage science

Article

https://doi.org/10.1038/s40494-025-01774-z

Oracle bone heritage data augmentation based on two-stage decomposition GANs

Check for updates

Jiaying Gao[1], Fausto Giunchiglia[2], Tongyu Zhao[3,4], Xin Zhao[5], Chuntao Li[6] ✉ & Hao Xu[4,7] ✉

Oracle bone inscriptions(OBI) constitute a significant part of Chinese civilization, with many mysterious characters remaining undeciphered. Recent advances in deep learning have introduced novel approaches for OBI decipherment. However, OBI samples, often derived from tortoise shells and bones, are prone to corrosion and weathering, degrading dataset quality. This paper proposes an OBI augmentation model based on two-stage decomposition GANs, learning an unidirectional mapping to transform low-quality samples into high-quality ones. Unlike traditional data augmentation approaches that require pre-defined standard samples, our model utilizes a two-stage decomposition framework to augment OBI samples exclusively from existing data, ensuring enhanced realism and diversity, as measured by the Inception Score. Compared to existing augmentation techniques, our approach improves the Inception Score by 6.7% and reduces the Brisque Score by 30.2% on the public dataset. The augmented dataset quality improves recognition accuracy by 2.56%, demonstrating our approach's effectiveness in real-world OBI decipherment.

Oracle bone inscriptions (OBI) represent the earliest known systematic and mature writing system in China and East Asia, dating back to 1500 BC[1] during the late Shang Dynasty. OBI is considered a crucial precursor to modern Chinese characters. These inscriptions were primarily engraved on animal bones and tortoise shells, though some were also found on stone tablets and other materials. After being buried underground for centuries, many OBI artifacts have suffered severe corrosion and weathering, as illustrated in Fig. 1. This degradation poses significant challenges to deciphering OBI after excavation. To date, archeologists have unearthed over 150,000 artifacts bearing OBI, encompassing 4672 distinct characters. Among these, only 1723 have been successfully deciphered into modern Chinese characters, leaving the majority of their meanings unresolved[2,3].

Artificial intelligence (AI) has introduced novel approaches to OBI decipherment, significantly improving efficiency and bridging the gap between readers and unearthed artifacts. Researchers have developed various OBI image datasets based on sample sources, including hand-printed datasets[4–7] and scanned datasets[8–11]. Hand-printed datasets offer exceptional diversity, as they can be simulated based on the writer's subjective interpretations. In contrast, Oracle Character Recognition (OrCR) models trained on scanned datasets demonstrate broader applicability in archeological scenarios due to their authenticity, providing archeologists with objective reference results. However, existing scanned OBI datasets are plagued by severe class imbalance, redundancy, and corruption caused by historical degradation. These issues significantly impair the performance of OrCR models, including those based on CNN frameworks[12–14] and modern Transformer architectures[15–17].

Data augmentation is a powerful technique for improving the quality and diversity of OBI datasets. Unlike artificial hand-printed data, which is inherently subjective, data augmentation enhances dataset quality and size while preserving the original data characteristics, making it more representative of real-world distributions. Additionally, data augmentation can be seamlessly integrated into general deep learning frameworks as a preprocessing step, such as VGG[18], Overfeat[19], ResNet[20], ViT[21], and Swin-Transformer[22]. Most importantly, data augmentation mitigates noise and corrosion effects in OBI samples while capturing their inherent styles, resulting in more diverse and realistic data.

This paper proposes a novel two-stage decomposition Generative Adversarial Network (GAN) framework for OBI data augmentation, which learns a unidirectional mapping from low-quality to high-quality samples. Our method employs a two-stage learning and generation framework to map low-quality samples from the reality domain to high-quality samples in the standard domain. This process enhances dataset quality without requiring external supervision or pre-defined standard samples. Experimental results demonstrate that our algorithm effectively reduces corrosion traces, improves dataset diversity, and generalizes well to other OBI datasets, thereby indirectly enhancing the performance of downstream classification models.

[1]Artificial Intelligence School, Jilin University, ChangChun, China. [2]Department of Information Engineering and Computer Science, University of Trento, Trento, Italy. [3]School of Engineering, University of Minho, Guimarães, Portugal. [4]College of Computer Science and Technology, Jilin University, ChangChun, China. [5]Exhibition Department, Jilin Provincial Museum, ChangChun, China. [6]Key Laboratory of Ancient Chinese Script, Culture relics and Artificial Intelligence, School of Archaeology, Jilin University, ChangChun, China. [7]Jilin University Library, Jilin University, ChangChun, China. ✉e-mail: lct33@jlu.edu.cn; xuhao@jlu.edu.cn

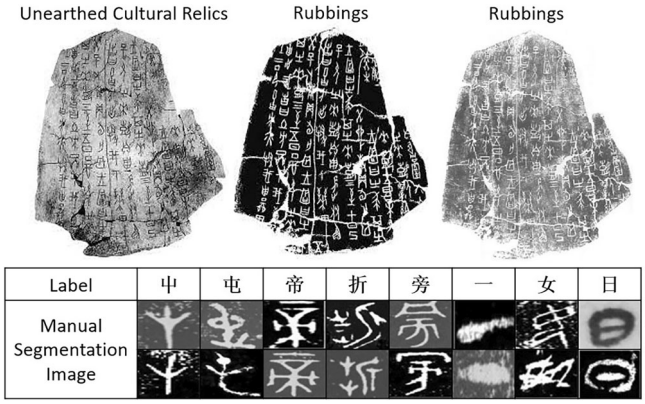


Fig. 1 | Oracle bone inscriptions on the artifacts.

Methods

Existing OBI augmentation models

Since 2022, OBI data augmentation has primarily focused on enhancing GAN frameworks to address various sub-tasks, including denoising[9,23], style transfer[24–26], restoration[27–29], and sample expansion[30,31]. The ultimate goal of these efforts is to improve the accuracy of recognition models[32,33]. While all these methods establish a mapping relationship from the source domain to the target domain, they differ in the mapping process and the number of generators employed.

On the one hand, models based on a single GAN typically establish a one-way mapping relationship between the source and target domains. These models often rely on external knowledge, such as unified fonts, evolutionary information, attention mechanisms, and graphic transformations, to compensate for deviations. For example, TH-GAN[24] generates characters by transferring font styles from one type to another. However, due to the unstandardized font styles and the presence of corrosion marks, this method lacks generalization ability for OBI. Sundial-GAN[33] generates different forms of oracle bone inscriptions based on their evolution through five historical stages until the modern form. This method relies on era sequence labels and corresponding samples. However, different OBI categories exhibit distinct evolutionary patterns, making it challenging to generalize beyond the given dataset. EA-GAN[27] restores ancient characters using example attention mechanisms[34], but its ability to expand sample diversity is limited. AGTGAN[30] employs unsupervised image transformation for data augmentation. By incorporating rotation, translation, and other graphic transformations into the generator's learning process, it outputs diverse samples while preserving the original corrosion style of OBI. RCRN[9] utilizes skeleton extraction and U-Net generative networks to repair severely corroded OBI images. CharFormer[23] introduces an attention mechanism[21] and a skeleton corrector for denoising OBI samples. Li et al.[31] augment data by mixing majority and minority samples at the patch level, which enhances dataset creativity and even combines some non-existent singular samples.

While single GAN-based models provide a direct mapping from source to target, they exhibit several limitations. Firstly, many models, such as TH-GAN and Sundial-GAN, rely on external annotations (e.g., standard font styles, evolutionary stages), making them less flexible for real-world applications where such information is unavailable. Secondly, models like EA-GAN and RCRN are primarily designed for restoration tasks but lack the ability to generate diverse new samples, limiting their utility for dataset expansion. Finally, reliance on pre-defined transformations (e.g., AGTGAN's rotation and translation) restricts adaptability to complex degradation patterns found in oracle bone inscriptions. These constraints highlight the need for a more generalizable approach that can handle diverse and naturally degraded OBI images without external supervision.

On the other hand, models based on multiple GANs typically establish a transitional mapping relationship between the source and target domains, enabling the generative model to learn in stages. For example, DDGAN[28] employs a generator with two discriminators to restore images, allowing the discriminative model to learn in stages. DualGAN[29] extends this idea by adding another generator, forming a dual complete GAN (Generative Adversarial Network) to restore damaged calligraphy characters. While DualGAN's approach is similar to our proposed Two-Stage GAN, it requires pre-damaged samples as specific references, which is feasible for calligraphy but impractical for OBI samples, as their pre-excavation appearance is unknown.

While multi-GAN architectures introduce intermediate transformations for improved restoration and augmentation, they also present several challenges. For instance, DualGAN requires pre-damaged reference samples, which are unavailable for real-world OBI inscriptions, limiting its practical applicability. Additionally, architectures like DDGAN introduce training instability due to multiple discriminators and training stages, making them less efficient for large-scale dataset augmentation. Therefore, an effective OBI enhancement model for multi-GANs should be able to be stably trained without explicit references or external labels and generate high-quality samples.

Architecture of the proposed model

To address training instability and dependence on extra annotations, we propose a two-stage decomposition GAN for OBI augmentation, guided by the standard domain embedded in the dataset. Unlike previous methods that rely on reference samples or pre-defined transformations, our approach introduces a novel reference-free strategy. It leverages skeleton-based transitions and a progressive learning framework to enhance real-domain samples without external labels or paired data. The cumulative effect of two sequential generators significantly boosts the diversity of the augmented outputs. Additionally, we design a specialized loss function, a tailored generator-discriminator architecture, and a balanced training scheme to ensure stability and effectiveness. These innovations enable scalable, high-quality, and annotation-free augmentation for real-world OBI decipherment tasks.

As illustrated in Fig. 2, our framework employs a two-stage GAN-based approach to progressively enhance low-quality OBI samples. The first stage, Reality-to-Skeleton GAN, generates structural features by converting input images into skeleton representations via a generator-discriminator pair (G_{RS}, D_{RS}). The second stage, Skeleton-to-Standard GAN, refines texture details to restore high-quality images using another pair (G_{SS}, D_{SS}). Both generators (G_{RS}, G_{SS}) and discriminators (D_{RS}, D_{SS}) share identical network architectures but are trained independently. This decomposition enables the model to sequentially learn structural and stylistic features, thereby improving reconstruction fidelity.

Preprocessing

Preprocessing involves dataset separation, normalization, and skeletonization. Given an OBI dataset D, we manually separate the samples into a standard domain $s \in S$ (high-quality images) and a reality domain $r \in R$ (low-quality images) based on image clarity and structural integrity. To enhance the model's generalization ability, we perturb reality domain samples with Gaussian noise as defined in Equation (1):

$$\tilde{r} = r + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

where $\tilde{r}$ denotes the perturbed sample, ϵ is the additive noise, and σ^2 controls the noise intensity.

To extract the structural skeleton of standard domain samples s, we apply morphological operations formulated in Equation (2):

$$I_{\text{sk}} = \bigcup_{k=0}^{N} [(s \ominus kB) \setminus ((s \ominus kB) \circ B)] \tag{2}$$

where $\ominus$ denotes morphological erosion, k iterates until $s \ominus kB = \emptyset$ (extracting the centerline structure), $\circ$ represents the morphological opening operation, and B is a 3×3 square structuring element.

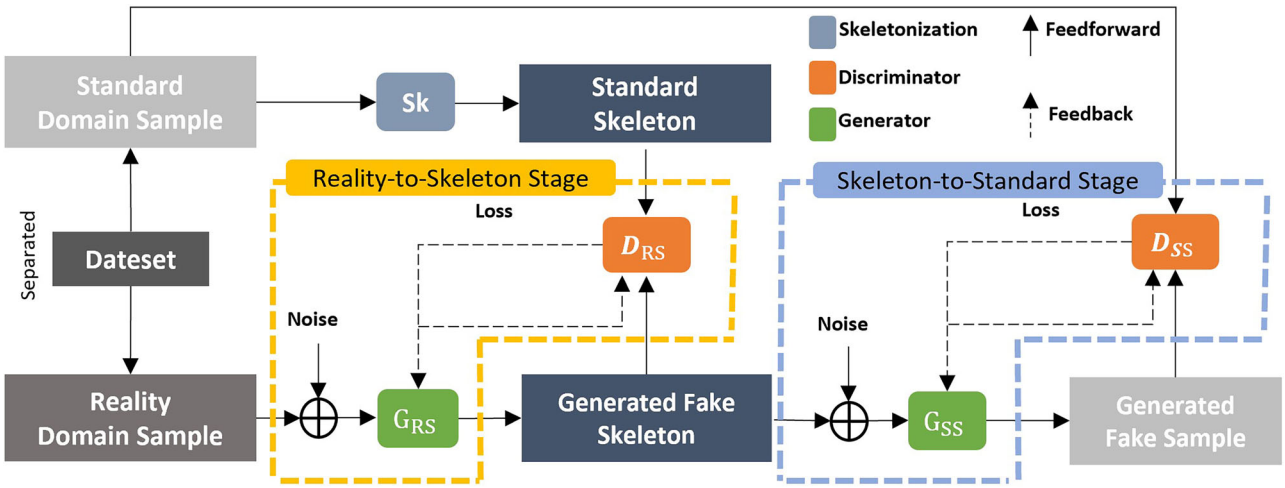


Fig. 2 | Flowchart of two-stage decomposition GANs.

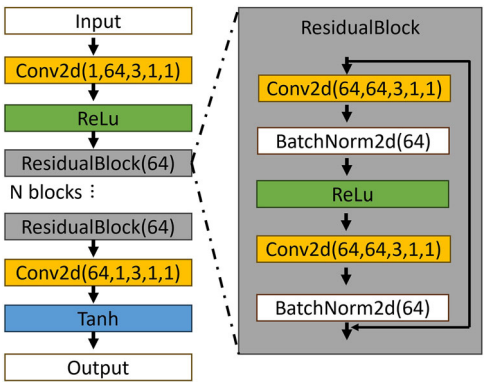


Fig. 3 | Generate structure of G_{RS} and G_{SS}.

Generator

The generator's primary function is to synthesize fake images that closely resemble real images. Our framework employs two generators: a Reality-to-Skeleton generator (G_{RS}) and a Skeleton-to-Standard generator (G_{SS}). The input to G_{RS} is a noisy reality sample $\tilde{r}$, and its target is to generate a fake skeleton image F_{sk} that approximates the ground truth skeleton I_{sk}. The input to G_{SS} is the output of G_{RS} with added noise, denoted as $\tilde{F}_{sk}$, and its target is to generate a fake standard image F_{st} that approximates the high-quality sample $s \in S$. The generators are formally defined as:

$$F_{sk} = G_{RS}(\tilde{r}; \theta_{RS}), \quad F_{st} = G_{SS}(\tilde{F}_{sk}; \theta_{SS}) \tag{3}$$

where $\tilde{r}$ and $\tilde{F}_{sk}$ are obtained by adding Gaussian noise as defined in Equation (1). The generator parameters θ_{RS} and θ_{SS} are updated via gradient descent, as described in Equation (4). The optimization objective is to minimize the error between the generated features (F^f_{sk}, F^f_{st}) and the target features (I^f_{sk}, I^f_{st}).

$$\theta \leftarrow \theta - \eta \nabla_\theta L(I^f, F^f) \tag{4}$$

where θ denotes the parameters of G_{RS} or G_{SS}, η is the learning rate, and ∇ represents the gradient. L represents the loss function, and different loss functions can be designed according to the constraint objectives. The learning rate is adjusted dynamically based on the training epoch t:

$$\eta_t = \eta_0 \cdot \gamma^t \tag{5}$$

where η_0 is the initial learning rate, γ is the decay factor, and t is the current epoch. Specific values for η_0 and γ are provided in section "Training Setup".

The architecture of generators G_{RS} and G_{SS} is illustrated in Fig. 3. Each generator comprises three main components: (1) an initial convolutional layer with a 3×3 kernel, stride 1, and padding 1, which extracts low-level features while preserving spatial resolution; (2) Residual block N is a hyperparameter, stacked for efficient feature conversion and reuse. The specific number of stacking N depends on the local computing performance; and (3) a final convolutional layer with a Tanh activation function, constraining the output values to the range $[-1, 1]$. The residual block is defined as:

$$y = \mathcal{F}(\tilde{r}, W_i) + \tilde{r} \tag{6}$$

where y is the output of the residual block, and $\mathcal{F}(\tilde{r}, W_i)$ represents a non-linear mapping composed of two convolutional layers and batch normalization, as shown in Equation (7):

$$\mathcal{F}(\tilde{r}, W_i) = \mathrm{BN}(\sigma(\mathrm{W}_2 * \mathrm{BN}(\sigma(\mathrm{W}_1 * \tilde{\mathrm{r}})))) \tag{7}$$

Here, * denotes the convolutional operation, BN(·) represents batch normalization, and σ(·) is the ReLU activation function. Batch normalization stabilizes training, accelerates convergence, and mitigates internal covariate shift by normalizing feature distributions after each convolutional layer. Skip connections are introduced to enhance feature reuse, as described in Equation (8).

$$y_{final} = X_{initial} + \widetilde{X}_{residual} \tag{8}$$

Where $\widetilde{X}_{residual}$ accumulates transformations from all residual blocks with Equation (9).

$$\widetilde{X}_{residual} = \mathcal{F}_N(\mathcal{F}_{N-1}(\ldots \mathcal{F}_1(\widetilde{X}_{initial}))) \tag{9}$$

$\mathcal{F}_i(\cdot)$ represents *ith* residual block mapping. In this paper, five residual blocks ($N = 5$) are used to capture non-linear feature transformations while preserving key structural information from the input.

The loss function serves as a critical reference for updating the generator parameters. To ensure the generated images closely resemble the target images in terms of details, we design a composite loss function comprising three components: adversarial loss ($\mathcal{L}_{GAN}$), content loss ($\mathcal{L}_C$) [35], and perceptual loss ($\mathcal{L}_P$) [36,37]. These losses are weighted by coefficients λ_{GAN}, λ_C, and λ_P to form the total generator loss:

$$\mathcal{L}_G = \lambda_{GAN}\mathcal{L}_{GAN} + \lambda_C\mathcal{L}_C + \lambda_P\mathcal{L}_P \tag{10}$$

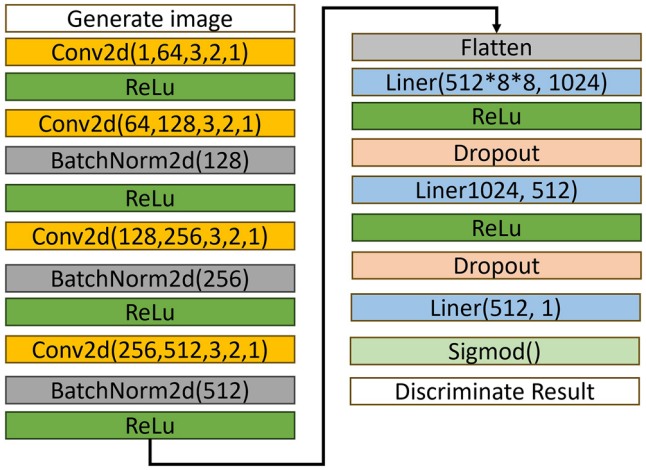


Fig. 4 | Discriminator structure of D_{RS} and D_{SS}.

The adversarial loss $\mathcal{L}_{GAN}$ ensures the generated images are realistic by minimizing the following objective:

$$\mathcal{L}_{GAN} = \mathbb{E}_{z\sim p_z}[\log D(G(z))] \tag{11}$$

where $\mathbb{E}_{z\sim p_z}$ denotes the expectation over the latent variable distribution p_z, ensuring the generator's output is indistinguishable from real data under the discriminator's judgment. The content loss $\mathcal{L}_C$ measures the L1 distance between the generated image and the target image s:

$$\mathcal{L}_C = \| G(z) - s\|_1 \tag{12}$$

Finally, the perceptual loss $\mathcal{L}_P$ extracts feature maps from a pre-trained VGG network [18] to measure the similarity between the generated and target images in a high-dimensional feature space:

$$\mathcal{L}_P = \| \phi(G(z)) - \phi(s)\|_2 \tag{13}$$

where ϕ denotes the feature extraction function of the VGG network.

Discriminator

The discriminator's primary role is to distinguish real images from generated ones, providing adversarial feedback to refine the generator's parameters. It guides the training process by quantifying the difference between real and generated data through a loss function. Our framework employs two CNN-based discriminators, D_{RS} and D_{SS}, whose inputs are F_{sk} and F_{st}, respectively, as defined in Equation (3).

The architecture of the discriminators is illustrated in Fig. 4. Both D_{RS} and D_{SS} are based on a CNN framework, with inputs F_{sk} and F_{st}, respectively, as defined in Equation (3). Each convolutional layer in the discriminator is defined as:

$$h_i = \mathrm{LU}(W_i * h_{i-1} + b_i) \tag{14}$$

where h_0 is the input image, W_i and b_i are the weights and bias of the i-th layer, and * denotes the convolution operation. LU(·) represents the LeakyReLU activation function [38]. Batch normalization is then applied to the convolutional feature map:

$$h_i' = \frac{h_i - \mu_i}{\sqrt{\sigma_i^2 + \epsilon}} \tag{15}$$

The final convolutional output is flattened into a vector v, which is passed through two fully connected layers:

$$v_1 = f(W_1 v + b_1), \quad v_2 = f(W_2 v_1 + b_2) \tag{16}$$

The output layer is activated by the sigmoid function $\sigma(\cdot)$:

$$D(x) = \sigma(W_3 v_2 + b_3) \tag{17}$$

The discriminator's objective is to distinguish real images from generated ones. Its loss function measures the accuracy of this classification: for real images, the discriminator output should be close to 1, while for generated images, it should be close to 0. Given a discriminator D, a generator G, a real image $\mathbf{x}$, and a latent input $\mathbf{z}$, the discriminator's loss function is defined as:

$$\mathcal{L}_D = -\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}}[\log D(\mathbf{x})] - \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}}[\log(1 - D(G(\mathbf{z})))] \tag{18}$$

where $\mathbb{E}[\cdot]$ denotes the expectation over the real data distribution p_{data} and the latent variable distribution $p_{\mathbf{z}}$. $D(\mathbf{x})$ represents the probability that the discriminator assigns to a real image, ideally close to 1, while $D(G(\mathbf{z}))$ represents the probability assigned to a generated image, ideally close to 0.

Training setup

The model was implemented using the PyTorch framework. The training was conducted on an Intel i9 CPU and an NVIDIA RTX 3070 GPU with 8 GB of memory. The number of residual blocks N is 5. The loss weights are set as $\lambda_{GAN} = 1$, $\lambda_C = 0.75$, and $\lambda_P = 1$. These values were chosen based on follow-up experiments at section "Ablation and visualization experiment", where we tested different loss weight combinations and observed that setting $\lambda_C = 0.75$ helped balance the content loss and adversarial loss, preventing the generator from overfitting to the adversarial component while still preserving structural details in the generated images. The initial learning rates are $\eta^0_{G_{RS}} = 0.00025$ for the Reality-to-Skeleton generator, $\eta^0_{G_{SS}} = 0.00025$ for the Skeleton-to-Standard generator, $\eta^0_{D_{RS}} = 0.0003$ for the skeleton discriminator, and $\eta^0_{D_{SS}} = 0.0003$ for the target discriminator. These values were selected based on a grid search over 0.0001, 0.0002, 0.00025, 0.0003, 0.0005, where we found that a slightly higher learning rate for the discriminators ($\eta_D > \eta_G$) stabilized training and reduced mode collapse, aligning with findings from prior studies on GAN training stability[39]. The learning rate decay factor is $\gamma = 0.995$, and the noise intensity added to the input image is $\sigma = 0.05$, which introduces minor perturbations to improve model generalization. The generator is trained three times per epoch, while the discriminator is trained once. This ratio was chosen because training the generator more frequently prevents the discriminator from becoming too dominant, which can lead to vanishing gradients and poor generator updates[40]. Label smoothing is applied with ranges of (0.9, 1.0) for true labels and (0, 0.1) for fake labels, following best practices in GAN training to mitigate overconfidence in discriminator predictions[41]. Additionally, data augmentation techniques such as rotation, contrast adjustment, and horizontal flipping are employed to enhance the model's robustness. These transformations enable the generator to learn structural and stylistic patterns under varying conditions, improving the quality of generated images.

Results

Data preparation

We evaluate our method on two public datasets: OBC306[8] and RCRN[9]. OBC306 contains 306 categories with 309,551 samples and is primarily used for classification and restoration tasks. RCRN consists of 362 categories with 1606 samples, manually collected from rubbing materials. There is no overlap between the two datasets. The datasets are randomly split into training, validation, and test sets with a ratio of 8:1:1. Figure 5 displays samples

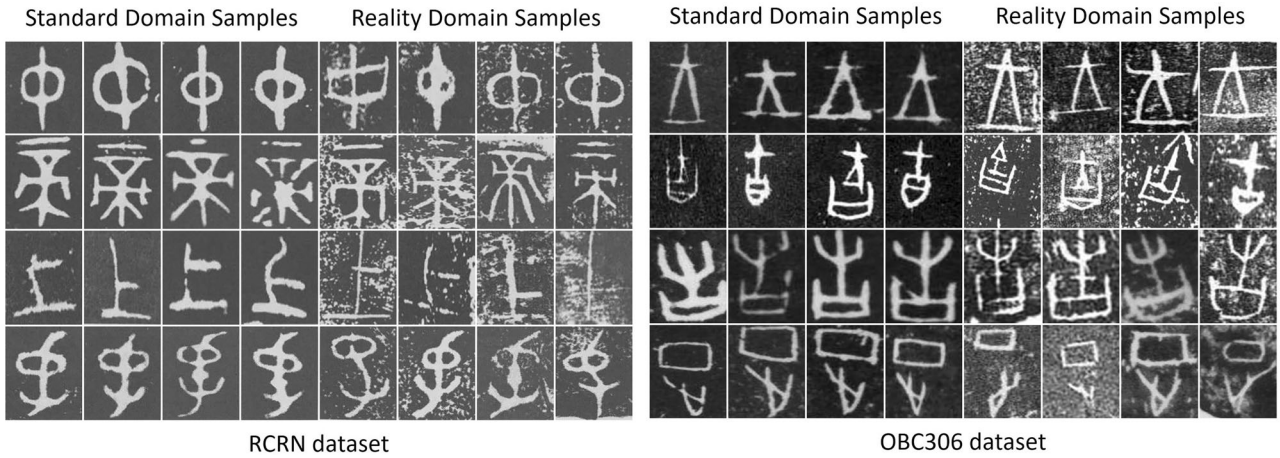


Fig. 5 | Dataset sample display. Each dataset was separated into the Standard and Reality Domains.

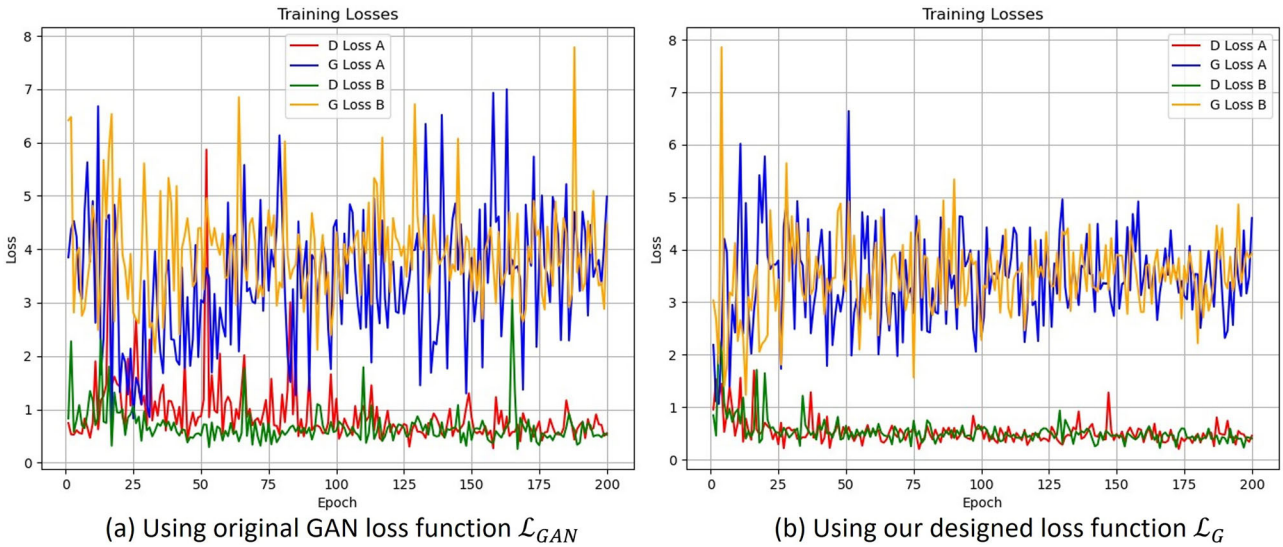


Fig. 6 | Loss curves during the training of the generator and discriminator under different loss functions. a Training losses under the original GAN loss function. **b** Training losses under the designed loss function.

from the RCRN and OBC306, divided into two domains: **Standard** and **Reality**. The standard domain comprises well-defined, clean character images, while the reality domain includes noisy, distorted, or incomplete characters.

Training stability analysis

After initializing the hyperparameters of the generators (G_{RS}, G_{SS}) and discriminators (D_{RS}, D_{SS}), training proceeds for 200 epochs. To balance the learning dynamics, the discriminator updates its weights once per epoch, while the generator updates its weights three times. Figure 6 shows the loss curves for G_{RS}, G_{SS}, D_{RS}, and D_{SS}. The red line (D loss A) represents D_{RS}, the blue line (G loss A) represents G_{RS}, the green line (D loss B) represents D_{SS}, and the yellow line (G loss B) represents G_{SS}. In the first 75 epochs, all loss curves exhibit significant oscillations without clear convergence, a common phenomenon in GAN training due to the adversarial nature of the generator-discriminator interaction. Compared to Fig. 6a, the variance of G_{SS} and D_{SS} losses is significantly reduced in Fig. 6(b), demonstrating the improved training stability achieved by our proposed loss function $\mathcal{L}_G$ (Equation (10)).

As shown in Fig. 6a, the loss curves of both generators and discriminators exhibit significant volatility, particularly for G_{SS} and D_{SS} (yellow and green lines). In contrast, D_{RS} (red line) and G_{RS} (blue line) gradually stabilize, indicating a balanced adversarial relationship between G_{RS} and D_{RS}. The losses of D_{SS} (green line) and G_{SS} (yellow line) fluctuate more prominently, especially for G_{SS}. This is likely due to the dependency of G_{SS} on the output of G_{RS}, which complicates the training process. Since the input to G_{SS} is the output of G_{RS}, changes in G_{RS} directly affect the input distribution of G_{SS}, leading to increased loss fluctuations.

As shown in Fig. 6b, our designed loss function $\mathcal{L}_G$ (Equation (10)), which incorporates perceptual loss and content loss, significantly stabilizes the training of G_{SS} and D_{SS}. The loss values calculated using $\mathcal{L}_G$ converge gradually after epoch 50, with significantly reduced oscillation amplitudes compared to $\mathcal{L}_{GAN}$ in Fig. 6a. These results demonstrate that $\mathcal{L}_G$ it enhances training stability, particularly for generators, which are more sensitive to distribution shifts. This improvement is critical for generating high-quality oracle bone script images.

Evaluation indicators

Since our low-quality images lack direct ground truth references, we evaluate the generated images using three complementary metrics that assess distributional similarity, diversity, and perceptual quality:

- Fréchet Inception Distance (FID) [42]: Measures the similarity between generated and real image distributions using features from a pre-trained Inception-v3 network. FID evaluates the distribution of

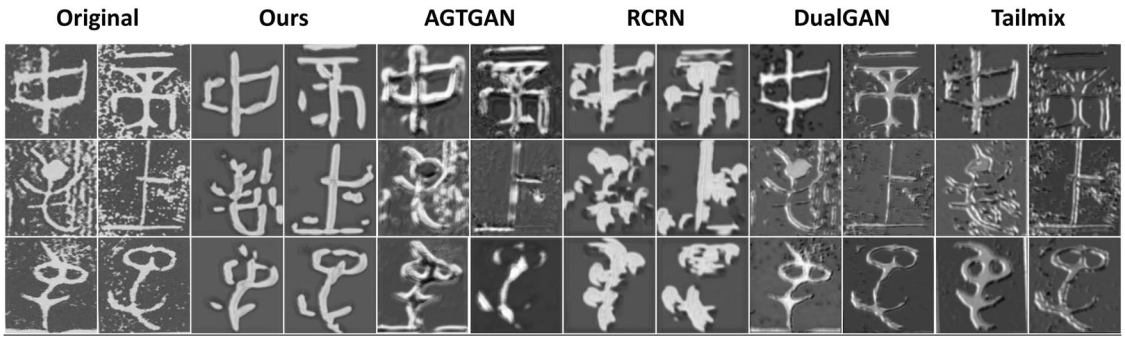


Fig. 7 | Comparison of generation performance of state-of-the-art models.

Table 1 | Performance comparison of generated samples.

Method	Year	RCRN[9]			OBC306[8]		
		FID↓	IS↑	BS↓	FID↓	IS↑	BS↓
Baseline(RD)	–	270	1.65	72	330	1.47	83
AGTGAN[30]	2022	188↓30.4%	1.94↑17.6%	61↓15.3%	240↓27.3%	1.66↑12.9%	74 ↓10.8%
RCRN[9]	2022	193↓28.5%	1.74↑05.5%	34↓52.8%	252↓23.6%	1.54↑04.7%	44↓47.0%
DualGAN[28]	2022	190↓29.6%	1.78↑07.9%	29↓59.7%	230↓30.3%	1.58↑07.4%	41↓50.6%
TailMix[31]	2023	182↓32.6%	2.10↑27.2%	59↓18.1%	226↓31.5%	1.82↑23.8%	73↓12.1%
Ours	–	174↓35.6%	2.20↑33.3%	24↓66.7%	205↓37.9%	1.85↑25.9%	37↓55.4%

Green values indicate improvements in noise reduction and image quality, while red values indicate improvements in diversity expansion.

features such as shape, texture, and structure in generated images compared to real oracle bone inscription images. A lower FID score indicates that the generated images are closer to real ones in terms of feature distribution.

- Inception Score (IS)[41]: Evaluates both the diversity and quality of generated images by computing the KL divergence between conditional and marginal class distributions. Since oracle bone inscriptions are not naturally classified into distinct categories, IS mainly helps assess the diversity and clarity of the generated characters rather than their classification consistency. A high IS suggests that the GAN generates distinct and well-defined oracle bone inscription symbols.
- Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) Score (BS)[43]: Measures perceptual quality by analyzing natural scene statistics, quantifying noise and artifacts in images. While BRISQUE does not explicitly assess structural fidelity, a lower BS score generally correlates with higher perceptual quality, meaning the generated images appear sharper, clearer, and more visually appealing.

By combining these three metrics, we obtain a comprehensive evaluation of the generated images, covering feature distribution similarity (FID), diversity and clarity (IS), and perceptual quality (BS). This ensures a well-rounded assessment of the model's performance.

Comparison of generation performance with state-of-the-art models

Our framework transforms reality-domain samples into the standard domain through two generators (G_{RS} and G_{SS}), enhancing dataset quality. We compare our model against four state-of-the-art methods: AGTGAN[30], RCRN[9], DualGAN[44], and TailMix[31]. Six low-quality test samples are used for qualitative analysis. Figure 7 shows that our model achieves effective denoising compared to DualGAN and RCRN, while AGTGAN and TailMix retain noise as a stylistic feature. Quantitative results in Table 1 demonstrate that our method significantly outperforms others.

In Table 1, we evaluate RD(reality-domain) samples in the dataset as a baseline. Then we input the real-domain samples into different state-of-the-art generation models to generate new samples and finally evaluate the indicators of the generated samples for comparison. As shown in Table 1, DualGAN and RCRN achieve lower BS scores (e.g., 29 and 34 vs. baseline RD's 72), indicating substantial noise reduction. However, their FID (190 and 193) and IS (1.78 and 1.74) scores remain close to the baseline (FID: 270, IS: 1.65), suggesting limited diversity improvement. In contrast, AGTGAN and TailMix enhance diversity (FID: 188 and 182, IS: 1.94 and 2.10) but retain higher noise (BS: 61 and 59). Our method strikes a balance, achieving the best FID (174), IS (2.20), and BS (24). Overall, our model reduces the BS by 66.7% and FID by 35.6%, compared to the reality-domain baseline, while improving the IS by 33.3%. This indicates a superior balance between noise reduction and diversity preservation on OBI datasets.

Ablation and visualization experiment

In this section, we perform an ablation study on our two-stage model to analyze the impact of different loss weight configurations on the generated results at each stage.

Figure 8 shows the generated images under different loss combinations: *-A*: Original input image. *-B*: Skeletonized image using Equation (2). *-C/D*: G_{RS}/G_{SS} with only GAN loss ($\lambda_{GAN} = 1, \lambda_C = 0, \lambda_P = 0$). *-E/F*: G_{RS}/G_{SS} with combined losses ($\lambda_{GAN} = 1, \lambda_C = 1, \lambda_P = 1$). *-G/H*: G_{RS}/G_{SS} with GAN and content losses ($\lambda_{GAN} = 1, \lambda_C = 1, \lambda_P = 0$).

Comparing *C/D* (only GAN loss) with *E/F* (combined losses), we observe that adding content and perceptual losses improves structural clarity (e.g., sharper strokes) and style consistency (e.g., better alignment with standard domain features). However, *G/H* (GAN + content losses) shows that content loss, while smoothing edges, may lead to detail loss (e.g., blurred textures). The experimental results demonstrate that our designed loss function $\mathcal{L}_G$ effectively preserves both structural and stylistic characteristics. Specifically, the combination of $\mathcal{L}_{GAN}$, $\mathcal{L}_C$, and $\mathcal{L}_P$ (as in columns E/F) achieves the best balance between realism and consistency,

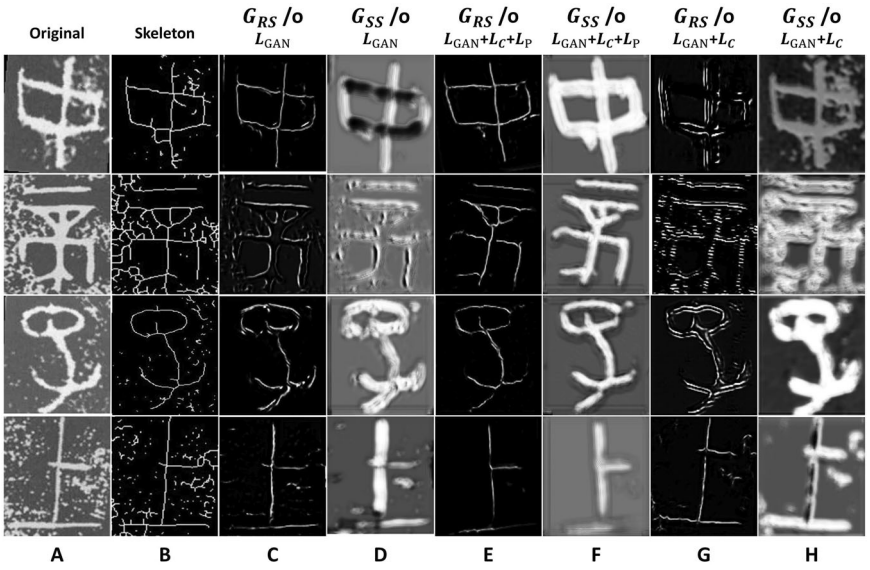


Fig. 8 | The results of the generator selecting different combinations of loss functions, where /o denotes **ONLY USE**.

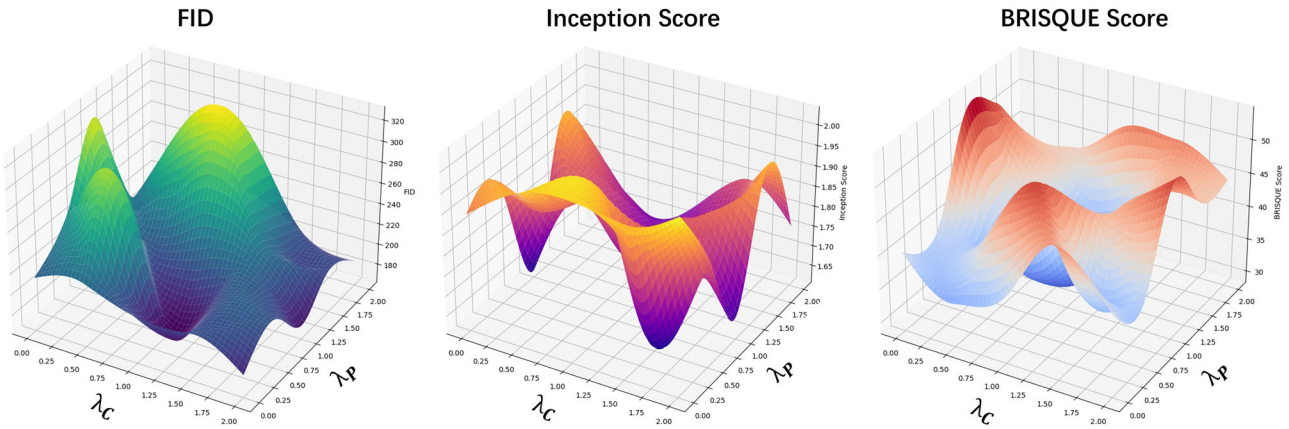


Fig. 9 | Loss weight ratio experiment. λ_C is the weight of content loss, λ_P is the weight of perceptual loss.

Table 2 | Classification performance (Top-1 and Top-3 accuracy) on different models and datasets, with(denoted as *w/Ours*) and without (denoted as Original) our proposed augmentation method

Form	Method	Years	OBC306[8]				RCRN[9]			
			Original		w/ Ours		Original		w/ Ours	
			Top1	Top3	Top1	Top3	Top1	Top3	Top1	Top3
Supervised Learning	ResNet101[20]	2016	69.50	74.92	**74.21**	**80.19**	64.45	71.72	**71.45**	**77.77**
	TailMix[31]	2023	82.24	86.68	**85.25**	**88.74**	76.24	82.27	**83.36**	**86.50**
	A-DeConv[45]	2024	86.45	88.14	**88.26**	**89.92**	83.36	85.58	**86.27**	**89.98**
Unsupervised Learning	UFCNet[46]	2024	86.88	88.51	**88.00**	**90.27**	83.75	86.04	**86.74**	**90.20**
	UDCN[44]	2024	87.25	90.02	**89.27**	**92.24**	85.27	88.26	**87.15**	**91.00**

The bold values are the results after using our augmentation method.

while using $\mathcal{L}_{GAN}$ $+\mathcal{L}_C$ (G/H) leads to smoother but potentially less detailed results.

The generator's total loss consists of weighted contributions from the GAN loss $\mathcal{L}_{GAN}$, content loss $\mathcal{L}_C$, and perceptual loss $\mathcal{L}_P$. To find the optimal weight value combination, we fix $\lambda_{GAN} = 1$ and vary λ_C and λ_P in the range [0, 2] with a step size of 0.25. We evaluate all combinations to identify the best performance to find the best one. Figure 9 visualizes FID, IS, and BS under different weight combinations. The global optimal value combination in the range [0,2] is found at $\lambda_C = 0.75$ and $\lambda_P = 1$.

Classification experiment

Table 2 shows the impact of our augmentation method on classification performance. We evaluate our augmentation method on the OBC306[8] and RCRN[9] datasets, using an 8:1:1 split for training, validation, and testing. Our approach is integrated into both supervised (ResNet101, TailMix, A-DeConv) and unsupervised (UFCNet, UDCN) OBI classification models. We use Top-n accuracy to evaluate performance, measuring whether the correct label is among the top n predictions. Our augmentation method consistently improves classification performance. For example, ResNet101's

Top-1 accuracy increases from 69.50% to 74.21% on OBC306 and from 64.45% to 71.45% on RCRN. Across all models, we achieve an average Top-1 accuracy improvement of 2.56%.

Discussion

We propose a novel oracle bone inscription (OBI) augmentation method based on two-stage decomposition GANs. Our method enhances dataset quality and diversity by mapping low-quality (reality domain) samples to high-quality (standard domain) samples through an intermediate skeleton representation. Unlike prior approaches, our method does not require external references and relies solely on the intrinsic style of the dataset.

Our main contributions include:

- **Novel Two-Stage Decomposition GANs Framework**: We propose a two-stage framework that maps real-domain samples to the standard domain via skeletonization, enhancing data quality and preserving style details.
- **Stable Training Strategy with Weighted Loss Control**: We design a loss function combining adversarial, content, and perceptual losses, stabilizing multi-GAN training.
- **Superior Generation Performance**: Our method outperforms state-of-the-art generative models on two public datasets, improving sample quality and reducing noise.
- **Improved Classification Performance**: Integrating our augmentation model into classification networks improves accuracy, with an average Top-1 gain of 2.56%.

Overall, our method enhances the quality and diversity of OBI datasets while improving downstream recognition tasks, demonstrating its broad applicability. Future work will explore extending this approach to other ancient script datasets and further optimizing the training process.

Data availability

The RCRN dataset is available at https://doi.org/10.1145/3503161.3548344. The OBC306 dataset is available at https://jgw.aynu.edu.cn/home/down/index.html.

Code availability

The underlying code for this study is available in the repository TSDGAN and can be accessed via this link: https://github.com/gaojiaying1218/TSDGAN.git.

Received: 18 February 2025; Accepted: 5 May 2025;
Published online: 03 June 2025

References

1. Fu, X., Zhou, R., Yang, X. & Li, C. Detecting oracle bone inscriptions via pseudo-category labels. *Herit. Sci.* **12**, 107 (2024).
2. Anzhu, L. et al. Oracle-bone inscriptions and cultural memory. *Front. Art. Res.* **2**, 63–73 (2020).
3. Li, J. et al. A comprehensive survey of oracle character recognition: challenges, benchmarks, and beyond. arXiv preprint arXiv:2411.11354 (2024).
4. Guo, J., Wang, C., Roman-Rangel, E., Chao, H. & Rui, Y. Building hierarchical representations for oracle character and sketch recognition. *IEEE Trans. Image Process.* **25**, 104–118 (2015).
5. Han, W. et al. Self-supervised Learning of Orc-Bert Augmentor for Recognizing Few-Shot Oracle Characters. *Asian Conference on Computer Vision,* 652–668 (Cham, Springer International Publishing, 2020).
6. Guan, H. et al. An open dataset for the evolution of oracle bone characters: Evobc. arXiv preprint arXiv:2401.12467 (2024).
7. Wang, P. et al. An open dataset for oracle bone character recognition and decipherment. *Sci. Data* **11**, 976 (2024).
8. Huang, S., Wang, H., Liu, Y., Shi, X. & Jin, L. Obc306: A large-scale oracle bone character recognition dataset. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*. 681–688 (IEEE, 2019).
9. Shi, D. et al. RCRN: Real-world Character Image Restoration Network via Skeleton Extraction. *Proceedings of the 30th ACM International Conference on Multimedia,* 1177–1185 (Association for Computing Machinery, 2022).
10. Diao, X. et al. Toward zero-shot character recognition: a gold standard dataset with radical-level annotations. *Proceedings of the 31st ACM International Conference on Multimedia,* 6869–6877 (2023).
11. Wang, M. & Deng, W. A dataset of oracle characters for benchmarking machine learning algorithms. *Sci. Data* **11**, 87 (2024).
12. Wang, K., Yi, Y., Liu, J., Lu, L. & Song, Y. Multi-scene ancient chinese text recognition. *Neurocomputing* **377**, 64–72 (2020).
13. Wang, K., Yi, Y., Tang, Z. & Peng, J. Multi-scene ancient chinese text recognition with deep coupled alignments. *Appl. Soft Comput.* **108**, 107475 (2021).
14. Guo, Z. et al. An improved neural network model based on inception-v3 for oracle bone inscription character recognition. *Sci. Program.* **2022**, 7490363 (2022).
15. Li, B. et al. Vit-based classification of oracle bones for periodization of the Shang dynasty. In *International Conference on Algorithms, High Performance Computing, and Artificial Intelligence (AHPCAI 2023)*, vol. 12941, 575–579 (SPIE, 2023).
16. Ma, Y. et al. Stef: a swin transformer-based enhanced feature pyramid fusion model for dongba character detection. *Herit. Sci.* **12**, 206 (2024).
17. Zheng, Y., Chen, Y., Wang, X., Qi, D. & Yan, Y. Ancient chinese character recognition with improved swin-transformer and flexible data enhancement strategies. *Sensors* **24**, 2182 (2024).
18. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).
19. Sermanet, P. et al. Overfeat: Integrated recognition, localization and detection using convolutional networks. *2nd International Conference on Learning Representations* (ICLR 2014).
20. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (IEEE, 2016).
21. Dosovitskiy, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020).
22. Liu, Z. et al. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10012–10022 (IEEE, 2021).
23. Shi, D. et al. CharFormer: A Glyph Fusion based Attentive Framework for High-precision Character Image Denoising. *Proceedings of the 30th ACM International Conference on Multimedia,* 1147–1155 (Association for Computing Machinery, 2022).
24. Cai, J. et al. TH-GAN: Generative Adversarial Network Based Transfer Learning for Historical Chinese Character Recognition. *2019 International Conference on Document Analysis and Recognition (ICDAR),* 178–183 (IEEE Computer Society, 2019).
25. Li, J. et al. Diff-oracle: Diffusion model for oracle character generation with controllable styles and contents. arXiv preprint arXiv:2312.13631 (2023).
26. Wen, Q. et al. ZiGAN: Fine-grained Chinese Calligraphy Font Generation via a Few-shot Style Transfer Approach. *Proceedings of the 29th ACM International Conference on Multimedia,* 621–629 (ACM, 2021).
27. Wenjun, Z., Benpeng, S., Ruiqi, F., Xihua, P. & Shanxiong, C. Ea-gan: restoration of text in ancient chinese books based on an example attention generative adversarial network. *Herit. Sci.* **11**, 1–13 (2023).
28. Chen, S., Yang, Y., Liu, X. & Zhu, S. Dual discriminator gan: Restoring ancient yi characters. *Trans. Asian Low. Resour. Lang. Inf. Process.* **21**, 1–23 (2022).

29. Su, B., Liu, X., Gao, W., Yang, Y. & Chen, S. A restoration method using dual generate adversarial networks for chinese ancient characters. *Vis. Inform.* **6**, 26–34 (2022).
30. Huang, H. et al. AGTGAN: Unpaired Image Translation for Photographic Ancient Character Generation. *Proceedings of the 30th ACM International Conference on Multimedia,* 5456–5467 (Association for Computing Machinery, 2022).
31. Li, J. et al. Towards better long-tailed oracle character recognition with adversarial data augmentation. *Pattern Recognit.* **140**, 109534 (2023).
32. Wang, W. et al. Improving oracle bone characters recognition via a cyclegan-based data augmentation method. In *International Conference on Neural Information Processing*. 88–100 (Springer, 2022).
33. Chang, X. et al. Sundial-GAN: A Cascade Generative Adversarial Networks Framework for Deciphering Oracle Bone Inscriptions. *30th ACM International Conference on Multimedia, MM 2022,* 1195–1203 (Association for Computing Machinery, Inc, 2022).
34. Vaswani, A. et al. Attention is all you need. *Adv. Neural Inf. Process. Syst*. 30 https://arxiv.org/abs/1706.03762 (2017).
35. Gatys, L. A., Ecker, A. S. & Bethge, M. A. neural algorithm of artistic style. arXiv preprint arXiv:1508.06576 (2015).
36. Johnson, J., Alahi, A. & Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14*, 694–711 (Springer, 2016).
37. Wu, Q., Fan, C., Li, Y., Li, Y. & Hu, J. A novel perceptual loss function for single image super-resolution. *Multimed. Tools Appl.* **79**, 21265–21278 (2020).
38. Maas, A. L., Hannun, A. Y. & Ng, A. Y. Rectifier nonlinearities improve neural network acoustic models, In *Proc. ICML*, Vol. 30, 3 (Citeseer, 2013).
39. Goodfellow, I. J. et al. Generative adversarial nets. *Adv. Neural Information Process*. Syst 27. https://arxiv.org/abs/1406.2661 (2014).
40. Mescheder, L., Geiger, A. & Nowozin, S. Which training methods for gans do actually converge? In *International Conference on Machine Learning.* 3481–3490 (PMLR, 2018).
41. Salimans, T. et al. Improved techniques for training GANs. *Adv. Neural Inf. Process. Syst.* 29. https://arxiv.org/abs/1606.03498 (2016).
42. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. & Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Adv. Neural Information Process. Syst*. 30 https://arxiv.org/abs/1706.08500 (2017).
43. Mittal, A., Moorthy, A. K. & Bovik, A. C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **21**, 4695–4708 (2012).
44. Wang, M., Deng, W. & Su, S. Oracle character recognition using unsupervised discriminative consistency network. *Pattern Recognit.* **148**, 110180 (2024).
45. Zhuo, S. & Zhang, J. Attention-based deformable convolutional network for chinese various dynasties character recognition. *Expert Syst. Appl.* **238**, 121881 (2024).
46. Zhou, Y. et al. UFCNet: Unsupervised network based on Fourier transform and convolutional attention for oracle character recognition. In *Proceedings of the 1st Workshop on Machine Learning for Ancient Languages (ML4AL 2024)*, 98–106 (2024).

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No.62476111), the National Natural Science Foundation of China (No.62077027), the Department of Science and Technology of Jilin Province, China (20230201086GX), the "Paleography and Chinese Civilization Inheritance and Development Program" Collaborative Innovation Platform (No.G3829), and the National Social Science Foundation project "Research on Chinese Pre Qin Language and Culture Based on Jinwen Data" (No. 23VRC033), and has also received funding from the interdisciplinary cultivation project for young teachers and students at Jilin University, "Research on Jinwen Data Based on Artificial Intelligence" (No. 2024-JCXK-04).

Author contributions

J.G. conceived and conducted the experiments and contributed to writing, T.Z. conducted the experiment(s), X.Z. separated the dataset into standard domain and reality domain, C.L. verified and collected the dataset, F.G. analyzed the results, and H.X. analyzed the results.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Chuntao Li or Hao Xu.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025