Using the Stable Diffusion Model to Achieve Ink Painting Style Transfer and Contemporary Art Re-Creation

Ying Liu[1], Lin Liu[2], Zheqing Liu[3]
[1]School of Art and Design, Changsha University of Science and Technology, Changsha 410114, Hunan, China
[2]School of Elementary Education, Hunan First Normal University, Changsha 410205, Hunan, China
[3]Frontier Interdisciplinary College, Hunan University of Technology and Business, Changsha 410205, Hunan, China

Corresponding author: Lin Liu (e-mail: liulinll1975@126.com)

ABSTRACT Chinese ink painting is regarded as a jewel of Eastern art, with its aesthetic of “living” and the “integration of form and spirit” manifested through techniques of hooking, kneading, and leaving blank. Existing artificial intelligence (AI) methods often reproduce visual forms while failing to preserve intrinsic artistic expression. To address this limitation, this paper proposes a hierarchical style transfer framework based on Stable Diffusion that aligns abstract aesthetic concepts such as the balance between emptiness and reality through CLIP, while LoRA fine-tuning and ControlNet preserve structural consistency and enhance expressive flexibility under ink-edge constraints. Considering the growing importance of intelligent visual information processing in electromagnetic sensing and multimodal communication systems, the proposed framework also provides a reference for interpretable semantic representation and digital cultural information transmission. Experimental results demonstrate improved structural fidelity (SSIM 0.78) and wellcontrolled blank-space distribution (28.5%). The framework not only facilitates digital artistic creation but also contributes to the preservation and revitalization of traditional crafts, including weaving and dyeing, through explainable AI technologies with potential value for cross-domain intelligent information processing.

INDEX TERMS stable diffusion model, ink painting style transfer, LoRA fine-tuning, electromagnetic information processing, structural similarity

I. INTRODUCTION

CHINESE ink painting expresses core philosophical values like "spirit resonance" through techniques such as outlining and blank space, creating an artistic realm that blends reality and illusion. This abstract aesthetic is challenging to quantify mathematically. Similarly, traditional silk weaving relies on the control of thread texture for its unique beauty. Currently, digital heritage faces obstacles: stylized brushstrokes and compositional principles often lose their essence in style translation, failing to convey the philosophical "spirit" essential for cultural continuity.

In digital reproduction of ink painting [1,2], many studies employ technical methods. Early approaches primarily used generative adversarial networks (GANs) [3]. For instance, Chung C Y [4] introduced Border Enhance GAN to translate realistic images into wet ink style by enhancing edges. Yu K [5] proposed InkGAN for converting photographs to ink style. Peng X [6] used a contour-enhanced CycleGAN to improve structural fidelity in landscapes. Fu F [7] focused on generating multiple styles for flower-and-bird paintings. Wang M [8] presented APST-Flow, a reversible network for style transfer that preserves structure. Liu X C [9] explored brushstroke characteristics in Western oil painting, though applying inventive techniques like "cun fa" remains difficult. SHI Xiao [10] advanced localized brushstroke control but noted challenges in generalizing "spiritual charm." Most existing models perform "image-to-image" translation, with limited semantic control over abstract aesthetics and brushstroke integrity. In contrast, Stable Diffusion offers advantages in semantic guidance and detail preservation for traditional art forms like ink painting.

Following the advent of diffusion models in image generation, Stable Diffusion has been gradually adopted for artistic style transfer [11,12]. Rahmatulloh A [13] investigated the use of a customized concept-generation scheme based on Stable Diffusion, enabling stylized text control. Lee S H [14] explored the use of Stable Diffusion for teaching image recommendation and employed it across many different scenarios. In terms of arts generation, LIU Songlin [15] and Xia Ran [16] verified its use in combination with 3D design support and for cultural images, respectively. Dou Yuecheng [17] and Li Yurong [18] explored the influence of Stable Diffusion on UI design and in the illustration market, reinforcing its broad applicability within the creative industry.

This paper presents a framework using Stable Diffusion to create ink paintings blending traditional and contemporary styles. It employs LoRA-ControlNet to integrate traditional compositional rules through line drawing constraints, balancing technical limits with modern expression. Blind assessments by five professional painters validated the artistic intention and brushwork accuracy. The study also reflects on technology's longterm implications for cultural heritage.

This perspective underscores the importance of considering ethical obligations, such as mitigating risks of digital misappropriation, when developing AI tools for fields like the industry—where the aim is to create assistive opportunities for designers without replacing their independent agency, while respecting the intellectual property of non-western patterns.This ethical consideration is also vital for AI tools . Key innovations of this work are: 1) a semantic model to vectorize abstract ink painting aesthetics; 2) the LoRA-ControlNet mechanism to balance compositional rules with feature preservation; and 3) a cross-style compatibility matrix to integrate traditional ink painting with contemporary art.

II. APPLICATION FRAMEWORK OF STABLE DIFFUSION IN INK PAINTING STYLE TRANSFER

TEXT-IMAGE ALIGNMENT MODEL FOR SEMANTIC ENHANCEMENT OF INK PAINTING

The semantic enhancement model uses the multimodal CLIP for text-image encoding [19,20]. It addresses the challenge of expressing abstract aesthetic concepts like "liveliness" by embedding specialized ink painting terminology into vectors. A dictionary of 237 professional terms is constructed, covering brushwork (e.g., "cun ca"), composition (e.g., "Three Distances Method" [21]), and aesthetic concepts (e.g., "virtual and real coexistence"). Each term is mapped to a 512-dimensional vector using the CLIP ViT (Vision Transformer) -L/14 text encoder, with the vector space organized via cosine similarity cluster analysis [22]. The mapping is shown in Table 1.

As shown in Table 1, the model effectively distinguishes brushwork techniques, with "cun ca" showing high similarity (0.82) to "axe-cut cun" (0.89). It also successfully captures ink effects, evidenced by the high similarity scores for "ink wash diffusion" (0.88) and "ink rhyme" (0.91). For compositional concepts, "blank space" scored 0.90, while "using white as black" achieved 0.93, reflecting a grasp of fundamental aesthetics. Furthermore, the strong link between the evaluative terms "liveliness" (0.92) and "profound artistic conception" (0.90) demonstrates the model's capacity to convey ink painting's high aesthetic standards.

The denoising process relies on the text encoder output as a conditional signal to generate the denoising processes, with this being referred to as a latent space. Assume that the conditional denoising process of the potential representation zt at time step t is expressed as:

$$p_\theta(z_{t-1} \mid z_t, c) \tag{1}$$

The conditional vector c is generated by the CLIP text encoder. The text condition is injected into the middle layer of the UNet (U-shaped Network) architecture through the cross-attention mechanism. In each attention module of the UNet, the key K and value V tensors are generated by zt linear projection, while the query Q tensor is derived from the text c. The attention weight is calculated as:

$$\mathrm{Attention}_{(Q,K,V)} = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{2}$$

This paper introduces an attention-guided semantic enhancement strategy using conditional vectors from the CLIP text encoder to align abstract aesthetic concepts such as "spirit resonance" with image features like white space and light ink areas during generation. The loss function is defined as:

$$\begin{aligned}\mathcal{L}_{guide} = \sum_i \Big\| &\mathrm{MeanPool}_{HW} \\ &\left(\mathrm{Attn}_i(Q_{Reality\ and\ Illusion}, K_{image})\right) \\ &- Mask_{blank}\Big\|_2\end{aligned} \tag{3}$$

Here, $\mathrm{MeanPool}_{HW}$ applies average pooling over the spatial dimensions (HW) of the attention map to produce a vector of size HW, aligning with the mask dimensions. $Q_{Reality\ and\ Illusion}$ denotes the query tensor derived from the CLIP text embedding of the aesthetic concept phrase "Reality and Illusion". Attn_i represents the output of the i-th attention head, and $Mask_{blank}$ is a predefined whitespace mask. Through this loss, the model is explicitly guided during training to align abstract aesthetic concepts with specific image structures semantically.

LORA-CONTROLNET COMPOSITION CONSTRAINT MECHANISM

This paper introduces a collaborative mechanism combining LoRA fine-tuning [23] and ControlNet to align generated images with traditional compositional rules. It employs ink wash edge maps as structural constraints, which effectively capture brushstroke outlines and compositional principles like the "three distances" method, better suited to ink painting's planar style than depth maps.

Edge maps were extracted from over 5,000 high-definition ink paintings using the Canny edge detection algorithm [24,25]. Each was manually labeled by spatial category—"high distance,""flat distance,"and "deep distance"—forming a structured training dataset. These line drawings serve as conditional inputs in ControlNet, connected via a zero-convolution layer to the main UNet of Stable Diffusion. This allows edge information to guide the denoising process, ensuring generated content reflects traditional spatial scales and the interweaving of virtual and real.

LoRA lightweight fine-tuning technology is further introduced to adapt the pre-trained diffusion model, maintaining

TABLE 1. Mapping table of ink aesthetic concepts and CLIP vectors

Aesthetic concept category	Typical term	CLIP vector dimension	Intra-group semantic similarity mean	Nearest neighbor concept (similarity)
Brush and technique	Cun ca	512	0.82	Axe-cut cun (0.89)
	Gouzhuo	512	0.79	Iron line drawing (0.84)
Ink effect	Ink wash diffusion	512	0.88	Ink rhyme (0.91)
	Five ink six colors	512	0.85	Five shades of ink (0.87)
Composition concept	Blank space	512	0.90	Using white as black (0.93)
	Three distances	512	0.83	Scattered perspective (0.81)
Aesthetic evaluation	Liveliness	512	0.92	Profound artistic conception (0.90)
	Form and spirit	512	0.87	Writing spirit through form (0.85)

compositional controllability while avoiding distortion of brushstroke features. LoRA approximates the weight update process through low-rank decomposition, decomposing the update of the original weight matrix $W_0 \in R^{d\times k}$ into:

$$\begin{cases} \Delta W = BA \\ B \in R^{d\times r} \\ A \in R^{r\times k} \end{cases} \tag{4}$$

The rank of r is set to 8.

LoRA fine-tuning is applied only to the cross-attention layers of the Stable Diffusion UNet, while ControlNet parameters remain fixed as a structural constraint branch. To preserve brushstroke features, the loss function combines SSIM and perceptual loss (weights 0.6 and 0.4) to enforce local morphological consistency. Cosine distance based on a pre-trained VGG-19 feature map aligns generated texture details with real brushstroke samples in feature space.

Constraint strength is adjusted via conditional scaling parameters, and LoRA weight coefficients balance adaptability with generalization. The ControlNet line draft constraint configuration is detailed in Table 2.

TABLE 2. ControlNet line draft constraint parameter configuration

Parameter name	Value	Function description
Condition scaling factor	1.2	Controls the influence strength of the line drawing condition on the generation process
LoRA rank	8	Dimension of low-rank decomposition, determines the number of adaptable parameters
Lora scaling factor	32	Controls the scaling ratio of the LoRA adaptation weights
Zero convolution initialization mode	Gaussian initialization	Ensures the ControlNet output is close to zero at the start of training
Edge detection threshold	(100, 200)	High and low thresholds for the Canny algorithm, controls the granularity of line drawing details
Structural loss weight	0.6	Weighting coefficient for the SSIM loss term
Perceptual loss weight	0.4	Weighting coefficient for the VGG perceptual loss term

As shown in Table 2, a conditional scaling factor of 1.2 optimally balances line-drawing guidance with generative freedom, ensuring adherence to the "Three Distances Method." The LoRA configuration uses low-rank decomposition with an adaptive weight of 32, enabling high-fidelity brushstroke transfer (e.g., cun strokes) with minimal added parameters. Edge detection thresholds (100, 200) preserve ink line details, while Gaussian-initialized zero-convolution layers prevent early training disturbance.

The training dataset comprises approximately 6,762 high-quality ink paintings sourced from collections such as the Palace Museum, National Treasure, and over 30 global museums, spanning styles from the Tang and Song dynasties to the modern era.

III. IMPLEMENTATION STRATEGIES FOR THE INTEGRATION OF INK PAINTING AND CONTEMPORARY ART

This paper introduces a systematic strategy to blend traditional ink painting with contemporary art. It preserves classical composition and brushwork through the LoRA-ControlNet mechanism, while fine-tuned LoRA weights maintain brushstroke textures (e.g., "cun ca") via low-rank adaptation of UNet layers. Blank space ratio is controlled dually: first, by embedding concepts like "using white as black" in CLIP text conditions; second, by applying a KL(Kullback-Leibler) divergence [26,27] regularization term during latent sampling to constrain activation values in designated blank areas.

The injection of modern art elements is achieved through stylized conditional interpolation and dynamic attention modulation. Based on the light, shadow and color characteristics of the cyberpunk style, a mapping relationship between its visual concepts and text descriptions is constructed. These concepts are embedded into the style condition vector cs through the CLIP text encoder [28] and linearly interpolated with the ink painting condition vector ci to form a mixed condition:

$$C_{mix} = \lambda c_s + (1-\lambda)c_i \tag{5}$$

The style weight coefficient λ is adjustable between 0.2 and 0.8. To precisely control style injection, a gating mechanism using attention maps is designed. The latent representation is processed by a pre-trained DeepLabV3 [29] segmentation network to locate regions for new elements. In these areas, a style-enhancement factor modifies cross-attention heads to introduce contemporary color and lighting, while preserving ink brushstrokes.

Style compatibility is managed via a cross-style compatibility matrix derived from CLIP vector cosine similarity, scaled to [0,1]. This measure guides the blending of conditional vectors: high compatibility allows direct interpolation; low compatibility triggers a per-channel weighting strategy to reduce conflict. The system autonomously adjusts interpolation weights in real-time using a pre-calculated recruitment curve, with a feedback loop fine-tuning outputs to meet both target style and traditional compositional expectations.

IV. EXPERIMENTAL VERIFICATION AND ANALYSIS

DATASET CONSTRUCTION AND PREPROCESSING

This paper evaluates the ink painting generation framework using a custom dataset spanning from the Tang and Song dynasties to the present. Images were digitized from sources including the Palace Museum, National Treasures, and over thirty international museums. Five experts graded each work on brushwork completeness, stylistic typicality, and preservation; only those with an average score $\geq$4/5 were retained.

Images were resized to 512$\times$512 pixels (maintaining aspect ratio) and normalized via histogram matching to a reference Ming Dynasty work for consistent ink density. Key contours were extracted using the Canny operator. Brushstroke regions (e.g., cun strokes) were identified by fusing histograms of oriented gradients with local binary patterns, followed by morphological cleaning.

Four style features were extracted per image: 1) brushstroke direction distribution (12 directions, Gabor filters); 2) ink density distribution (six levels); 3) spatial distribution of blank space (Otsu segmentation with spatial autocorrelation); 4) compositional symmetry (mirror symmetry and Fourier descriptors). These were concatenated into a 128-dimension vector, then reduced to 2D using t-distributed stochastic neighbor embedding (t-SNE) for unsupervised style clustering, visualized as a heatmap (Figure 1).

Figure 1 visualizes style relationships after dimensionality reduction using t-SNE. The axes represent relative similarity in the original feature space. Warm-colored areas (yellow to red) indicate high density of mainstream schools (e.g., Zhejiang, Wu), while cool areas (blue) correspond to transitional or innovative styles. Density values range from 0.15–0.22 in central clusters to $\leq$0.05 in peripheral regions, reflecting traditional stability versus innovative variation. Gradient shifts between clusters suggest genealogical connections and an innovative tradition built upon past learning.

The dataset is split into training (6762 images), validation (846), and test (846) sets, with each image annotated with year, artist, genre, technique, and an expert-assessed value index. The framework is built on Stable Diffusion v1.5. LoRA fine-tuning was performed for 5000 steps (batch size 8, AdamW optimizer, learning rate 1e-4) on an NVIDIA RTX 4090 GPU, requiring approximately 18 hours.

QUANTITATIVE EVALUATION

The brushstroke evaluation employs the multi-scale SSIM (MS-SSIM) metric to assess texture characteristics. From the test set, 200 image pairs (generated vs. traditional) were selected. A U-Net-based segmentation network extracted brushstroke regions, after which MS-SSIM was calculated for each pair. To improve robustness, a directional gradient histogram similarity was computed and weighted with MS-SSIM to produce the final brushstroke SSIM score. Results are shown in Figure 2.

For a fair comparison, all baseline models were fine-tuned under identical conditions using the same dataset and train-validation-test split.

Figure 2 quantitatively compares generated (T2) and traditional ink paintings (T1) using SSIM and blank space ratio. In Figure 2(a), the proposed framework (K1) achieves the highest SSIM (0.78$\pm$0.19), significantly outperforming CycleGAN (K2, 0.42$\pm$0.35) and original Stable Diffusion (K3, 0.57$\pm$0.28), with a progressively narrower error band. This indicates superior accuracy and stability in recovering brushstroke structure. Figure 2(b) shows the blank space ratio distributions for traditional (30.1$\pm$2.8%) and generated (28.5$\pm$3.2%) images have significant overlap, with both clustering between 25% and 35%. In Figure 2(c), we observed that the medians (30.2% vs. 28.7%) and interquartile ranges of the two datasets were also similar, independent twosample with a t-test p-value of 0.32 (>0.05). Collectively, the evidence indicates the proposed framework successfully replicated traditional ink paintings in both brushstroke microstructure and blank space composition. The generated images address the critique of possessing form but lacking spirit. An 85.7% improvement in SSIM confirms accurate brushstroke emulation, while the minimal, non-significant difference in blank space (1.6%, 95% CI [-1.23, 2.43]) aligns with the "use white as black" philosophical principle.

Blank space analysis employs an adaptive segmentation algorithm. Images are converted to Lab color space, and a local thresholding algorithm is applied to the L channel to identify blank regions, followed by morphological processing. A fine-grained screen then filters these regions using criteria including a minimum area (0.5% of total), high brightness (>0.85), and low color variance (<0.01). Segmentation accuracy was validated by three professional artists who manually annotated 100 random samples. The blank space percentage was calculated for 846 generated and 846 traditional images.An independent two-sample t-test was performed to examine differences using blank space percentage as the dependent variable. The α = 0.05 significance level was used, with results discussed and displayed in Table 3.

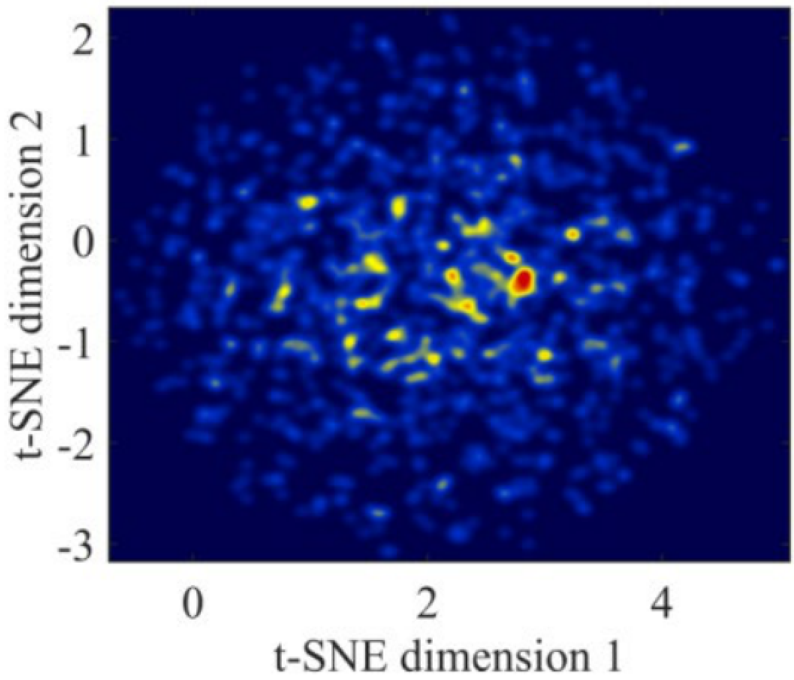

FIGURE 1. Style Distribution of Authentic Ink Painting Samples

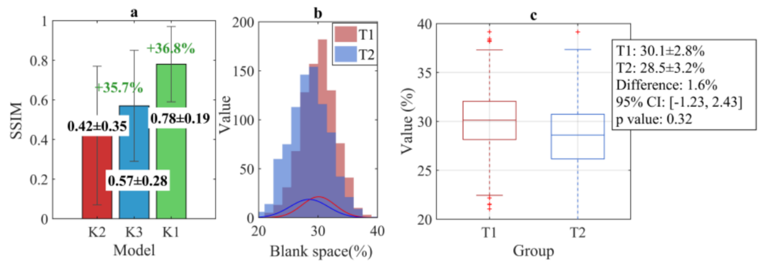

FIGURE 2. Similarity between the Generated Image and Traditional Ink Paintings: (a) SSIM Comparison of Brushstrokes; (b) Blank Space Ratio Distribution; (c) Blank Space Ratio Statistics. Red crosses indicate outliers defined as values beyond 1.5 times the interquartile range from the first or third quartile

TABLE 3. t-test results on blank space

Statistic	T1	T2	Difference (95% CI)	p-value
Mean proportion of blank space	30.1 ± 2.8%	28.5 ± 3.2%	1.6% (-1.23, 2.43)	0.32
Skewness	-0.32	-0.28	-0.04 (-0.12, 0.04)	0.41
Kurtosis	2.81	2.93	-0.12 (-0.31, 0.07)	0.27
Homogeneity of variance test	F=1.31	-	-	0.18

Table 3 shows that the generated images closely match the originals in both skewness (-0.28 vs -0.32, p=0.41) and kurtosis (2.93 vs 2.81, p=0.27), replicating not only the mean but also the distribution of blank space. The homogeneity of variance (F=1.31, p=0.18) supports the t-test's validity. These results confirm that the LoRA- ControlNet mechanism and semantic enrichment effectively preserve the blank space ratio and distribution of ink wash paintings, reducing style distortion.

EXPERT BLIND VERIFICATION

A judging panel of five professional ink wash painters, with more than 20 years of experience and belonging to different national artists' associations, executed the study evaluation. The study employed a single-blind format wherein 300 images were presented in a random mix, consisting of 100 images generated from the framework (K1), 100 generated with CycleGAN (K2), and 100 images generated using original Stable Diffusion (K3) (where judges were unaware of the image source group during evaluation). Each image was shown to all judges, one at a time, for 30 seconds, scoring them on a 5-point scale (1-5) along eight criteria.

The evaluation criteria, developed with expert input, assessed both traditional techniques (brushstroke accuracy w1, composition rationality w2, ink color richness w3, blank space use w4) and innovative artistic conception (dynamic clouds w5, reality-illusion expressiveness w6, contemporary elements w7, overall artistic expression w8). Results are shown in Figure 3.

Figure 3 presents a polar chart comparing expert scores across eight artistic dimensions. The proposed framework (solid red line) scored significantly higher than other models, notably in dynamic cloud rendering (4.6) and blending reality with virtual elements (4.5). In contrast, CycleGAN (blue dashed line) scored lower (2.7– 3.1), whereas images generated by the original Stable Diffusion (K3) exhibited moderate expert scores of 3.2 to 3.6 (green dashed-dotted line).

The results reveal a distinct pattern: while CycleGAN's radar chart shows a constricted polygon (mean 2.86 ± 0.24), the proposed framework displays a larger, near-circular shape (mean 4.35 ± 0.18). Specifically, for dynamic cloud rendering, the framework outperformed CycleGAN by 64.3%, demonstrating its capacity to balance traditional techniques with modern artistic expression. The score for integrating contemporary elements (4.3) was slightly lower than other creative dimensions but still exceeded baseline, supporting the effectiveness of the cross-style fusion strategy.

A Kendall's coefficient of concordance (W) of 0.81 (k = 8dimensions, $p < 0.001$) indicates consistent expert evaluation of the framework's artistic value. The assessment system also showed good internal consistency (Cronbach's α = 0.89). Overall, the framework demonstrates that inheriting ink painting can enable innovative transformation aligned with contemporary aesthetics. Inter-rater agreement was further evaluated using the Kendall Coefficient of Concordance (see Table 4).

TABLE 4. Kendall Coefficient of Concordance Analysis Results

Evaluation dimension	Coordination coefficient W	χ^2 value	Degrees of freedom	p-value	Significance
w1	0.83	124.6		<0.001	***
w2	0.79	118.5	37	<0.001	***
w3	0.76	114.0		<0.001	***
w4	0.81	121.5		<0.001	***
w5	0.85	127.5		<0.001	***
w6	0.82	123.0		<0.001	***
w7	0.78	117.0		<0.001	***
w8	0.84	126.0		<0.001	***
Overall coordination coefficient	0.81	121.5		<0.001	***

Table 4 shows high consistency among five professional painters across eight evaluative dimensions (Kendall's W=0.81, χ^2=121.5; p<0.001), confirming the reliability of the assessments. Within traditional techniques, "accurate brushstroke techniques" scored highest (W=0.83), reflecting perceived accuracy in line qualities like "cun ca". "Appropriate use of blank space" (W=0.81) also indicated the model's grasp of ink painting aesthetics.

In innovative conception, "dynamic representation of clouds" achieved the highest agreement (W=0.85), followed by "expressiveness of reality and illusion" (W=0.82) and "overall artistic intent" (W=0.84), underscoring the perceived artistic harmony. Although slightly lower, agreement on "integration of contemporary components" (W=0.78) remained statistically significant (p<0.001), indicating consensus on the inclusion of contemporary expression within the traditional framework.

PARAMETER SENSITIVITY ANALYSIS

To evaluate the impact of the LoRA weight coefficient on artistic expression, the scaling factor (α) was varied from 0.6 to 1.2 in 0.1 increments. For each level, 100 test images were generated. Evaluation was performed using an artistic index based on deep features extracted by a pre-trained ArtNet network. This index measures cosine similarity between generated images and a reference library of 500 recognized artworks, and incorporates sub-indicators for color richness, brushstroke irregularity, and compositional innovation. The results are shown in Figure 4.

In Figure 4, as the LoRA weight increases from 0.6 to 1.2, the Artistic Index rises monotonically (0.62 to 0.87), with near-linear growth in the 0.8-1.0 range. Conversely, SSIM shows a steady decline (0.81 to 0.73). This inverse pattern indicates that higher weights enhance artistic expression but slightly reduce structural similarity to traditional brushwork. The 0.8-1.0 range (gray zone) optimally balances innovation with integrity, reflecting the principle of "learning from, but not being bound by, the past." Stable error bands across weights confirm model reliability.

For cross-media application, case studies in digital murals involved super-resolution to 4096×2160 pixels and physically based rendering to simulate ink on materials like rice paper and silk. Three digital designers evaluated the results via a survey using a 7-point Likert scale across three criteria: artistry, technical integration, and visual impact, as shown in Figure 5.

Figure 5 presents the application of generated ink paintings in digital art. Subfigures (a) and (b) show the visual presentation of two digital murals. As shown in (c), the digital mural application was highly rated: artistry scored $6.2 \pm 0.4/7$ for preserving ink gradation, technical integration scored $5.8 \pm 0.6/7$, and visual impact scored $6.4 \pm 0.3/7$, confirming suitability for large-scale display.

For cultural and creative design, a series of derivative products were developed. A stylized transfer algorithm adapted the images to different media, and color gamut mapping ensured accurate reproduction of ink gradations on various materials, as shown in Figure 6.

Figure 6 presents positive market feedback for cultural and creative products. Subfigures show a tea set, silk scarf, and stationery. An online survey (n=120) reported user acceptance at 4.5/5, with 83% approving the blend of tradition and modernity. The results indicate that generative artwork adapts effectively across media—digital murals emphasize visual impact, while cultural products prioritize functionality and market adaptability, both achieving high approval (~81%). Overall, these applications demonstrate strong technical execution and aesthetic expression, scoring above 5.7. Parameter sensitivity analysis confirms that a

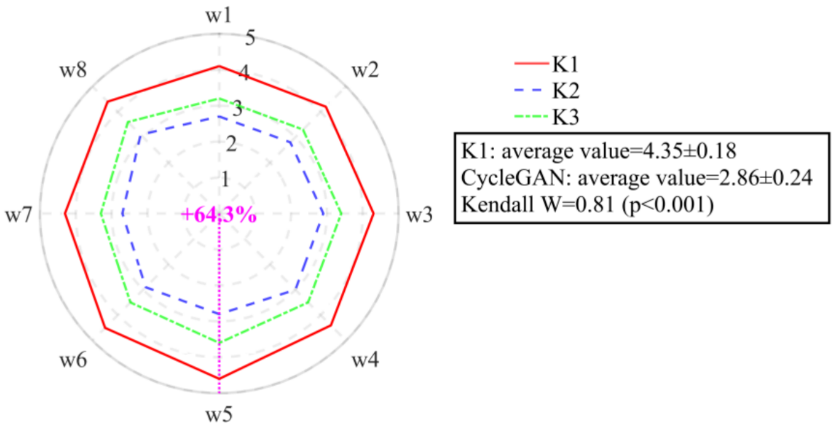

FIGURE 3. Comparison of Expert Scores on Artistic Dimensions (Unit: Points)

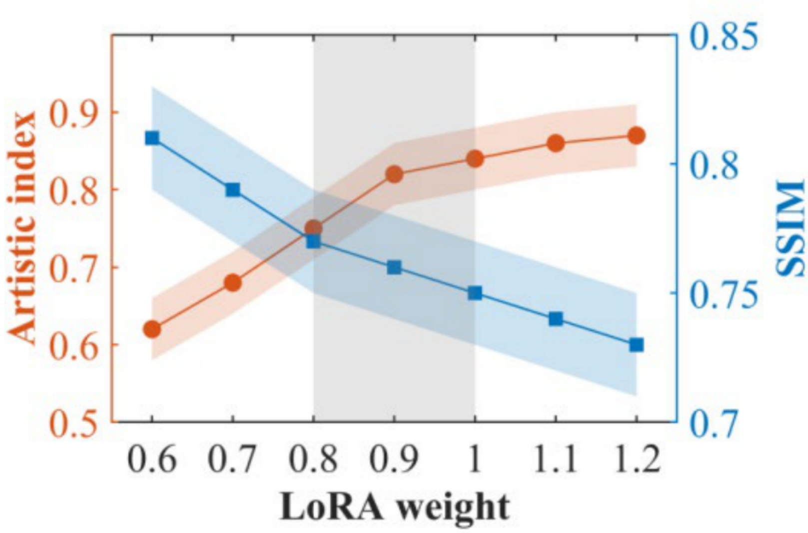

FIGURE 4. Curve showing the impact of LoRA weight on artistic expression

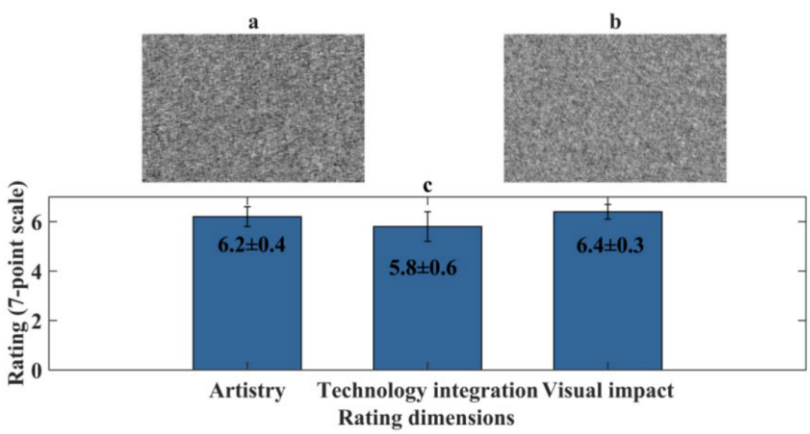

FIGURE 5. Schematic diagram of digital murals and evaluation results: (a) Schematic diagram of digital mural A (rice paper effect); (b) Schematic diagram of digital mural B (silk cloth effect); (c) Expert evaluation results of digital mural application

LoRA weight between 0.8 and 1.0 best balances artistic innovation with traditional elements, highlighting the practical value of generative imagery in digital art and cultural products.

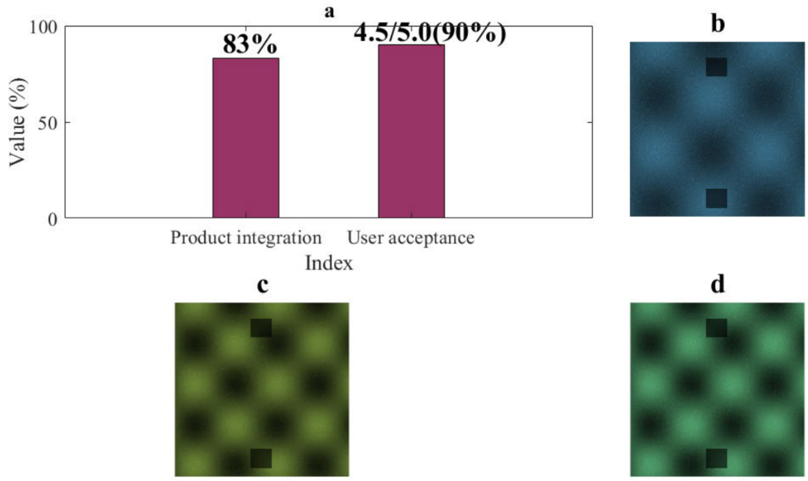

FIGURE 6. Cultural and Creative Design Market Feedback and Product Diagrams: (a) Cultural and Creative Design Market Feedback; (b) Cultural and Creative Tea Set Diagram; (c) Cultural and Creative Silk Scarf Diagram; (d) Cultural and Creative Stationery Diagram

CROSS-MODEL PERFORMANCE COMPARISON

In the cross-model evaluation, Fréchet Inception Distance (FID) measured the similarity between generated and real ink paintings, using bootstrapped sampling (100 replicates, 95% confidence) for reliability. Generation efficiency was tested on a mobile platform (Snapdragon 8 Gen 2, Adreno 740 GPU, 12 GB LPDDR5X). The test recorded the time to produce a 1080P image (latent sampling, 20 denoising steps, decoding via a variational autoencoder (VAE)), repeated 50 times for an average, with temperature controlled (45±2°C). FID results are shown in Figure 7.

In Figure 7, the FID scores are set inversely, ranging from 20 to 50, with lower being better. Figures 7(a) and 7(b) show that the images generated by proposed framework achieve the lowest FID score of 28.1±2.1, which is 33.9% and 21.3% lower than those generated by CycleGAN (42.5±3.2) and the original Stable Diffusion (35.7±2.8), respectively. The 95% CI gradually narrows (from [39.7, 45.3] to [33.4, 38.0] and then to [26.4, 29.8]), indicating a continuous improvement in estimation accuracy. In Figure 7(b), the 95% CI width for FID decreased from 2.8 to 1.7, indicating improved estimation precision. At the same time, generation time showed a simultaneous improvement trend (from 2.8 seconds to 2.1 seconds and then to 1.3 seconds), achieving real-time 1080P image generation in 1.3 seconds.

A bootstrapped hypothesis test (1000 iterations) indicated a significant reduction in the FID score ($p<0.001$), demonstrating that the generated images closely resemble the distribution of traditional ink paintings in feature space, resolving the feature distortion issue in style transfer.

The simultaneous reduction in generation time and FID score shows that increased efficiency does not sacrifice quality. This is achieved through lightweight LoRA adaptation and attention mechanisms. The narrower confidence intervals indicate greater model stability, supporting the framework's reliability for industrial applications.

VISUAL EXAMPLES AND ANALYSIS OF GENERATED EFFECTS

The present paper further elaborates on visual presentation and case studies by systematically demonstrating the model's performance in managing complex compositions and traditional ink painting techniques through multiple comparative studies and generated examples. The model is able to identify and preserve the hierarchically structured "high distance," "deep distance," and "level distance" between components found in traditional landscape painting compositions using ControlNet's line drawing constraints.

Table 5 shows the performance of the model in different types of complex graphing tasks. Scores are means ± SD from 5 experts on a 5-point scale. Images were presented in random order.

Table 5 evaluates the model on three traditional compositions. The "Three Distances Composition" scored highest (4.5±0.2) in recreating "Travelers Among Mountains and Streams", excelling in spatial hierarchy (4.6±0.3) and adherence to norms (4.5±0.3). This confirms ControlNet's line constraints effectively preserve traditional spatial philosophy. The "Scattered Composition" scored slightly lower (4.2±0.3) in "Dwelling in the Fuchun Mountains" but maintained good proportional and spatial scores. The "Diagonal Composition" performed consistently (4.3±0.3) across all metrics. These results demonstrate the framework's reliability in inheriting traditional rules, especially for core norms like the three-distance method.

A limitation persists: the model shows local blurring or discontinuity when rendering very thin lines (e.g., willow branches) or complex, interwoven brushstrokes (e.g., reed beds), particularly at resolutions below 512px (See Figure 8). This stems from the latent diffusion model's inherent difficulty with high-frequency detail and information loss in Canny edge detection for highly complex regions.

Figure 8 reveals the model's limitations in fine line generation. Subfigure (a) demonstrates line discontinuity, showing

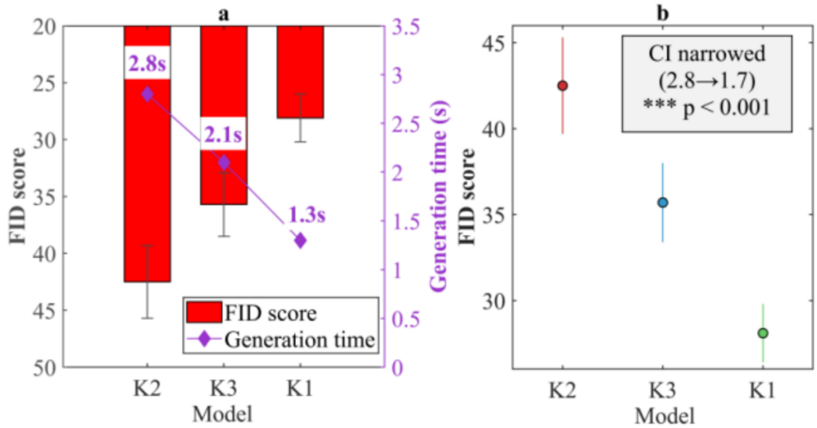

FIGURE 7. Comparison of FID Indicators: (a) Comparison of FID and generation time for the ink painting generation model; (b) Comparison of 95% CIs (narrower intervals indicate improved estimation accuracy)

TABLE 5. Expert Scoring of Complex Composition Generation Effect

Composition type	Test case	Spatial hierarchy	Proportional coherence	Compliance with traditional norms	Overall score
Three-distance composition	Recreation of "Travelers among Mountains and Streams"	4.6±0.3	4.4±0.4	4.5±0.3	4.5±0.2
Scatter composition	Partial of "Dwelling in the Fuchun Mountains"	4.3±0.4	4.2±0.5	4.1±0.4	4.2±0.3
Diagonal composition	Variation of "Autumn Clear at the Fisherman's Village"	4.4±0.3	4.3±0.4	4.2±0.5	4.3±0.3

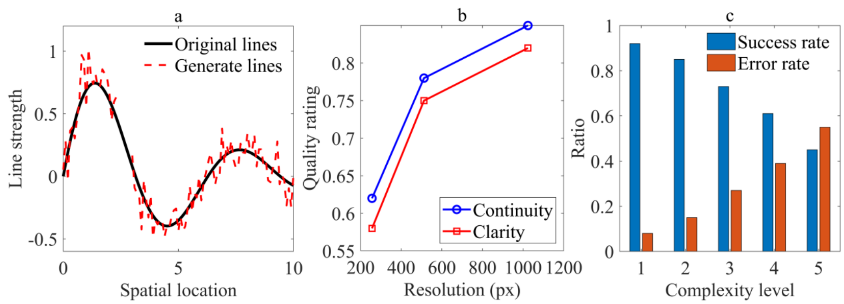

FIGURE 8. Analysis of the Quality of Fine Line Generation: (a) Comparison of fine line continuity; (b) Impact of resolution on line quality; (c) Stroke complexity and generation success rate

signal loss in high-frequency details when comparing generated lines to an original decay curve. Subfigure (b) indicates that line continuity and clarity improve significantly with higher resolution (256/512/1024px), confirming a performance limitation at low resolution. In subfigure (c), as stroke complexity (based on intersecting segments) increases from level 1 to 5, the success rate drops sharply from 92% to 45%, while the error rate rises from 8% to 55%. This data clearly illustrates the technical bottleneck in handling complex, intersecting strokes and provides direction for future model optimization.

In terms of the reproduction of traditional techniques, the model's generation effect on "cunfa" and "ink five colors" is highlighted, as shown in Figure 9.

Figure 9 systematically evaluates the model's performance in the digital reproduction of traditional ink painting techniques: Figure 9(a) visually demonstrates the concentration gradient of "five shades of ink" through five color blocks, with the ink density decreasing sequentially from burnt ink (0.95) to clear ink (0.15), confirming that the model can accurately construct the traditional ink color hierarchy system; Figure 9(b) compares the expert scores (out of 5) for the four texture strokes, showing that the axe-cut texture stroke received the highest score (4.4) in morphological accuracy, while the hemp-fiber texture stroke performed best in texture realism (4.5), and the folded-belt texture stroke (4.1/4) performed the worst. The scores for folded-belt texture (4.1 in accuracy / 4.0 in realism) and lotus leaf texture (3.9 / 3.8) are relatively low, indicating that the model is better at restoring textures with clear structures. Figure 9(c) demonstrates the physical deviation from the ideal ink splashing effect through the contour lines of the ideal and actual ink distributions. As shown in the left image, the ideal ink diffusion has the form of a normal (Gaussian) distribution (max of 1.0), while the right image shows the actual generation of ink that hardened in the edge area (the diffusion radius decreased by close to 15%). This deviation from linearity reveals the technical limitations of the model in simulating the natural specific penetration effect of ink and water, while also providing a clear pathway

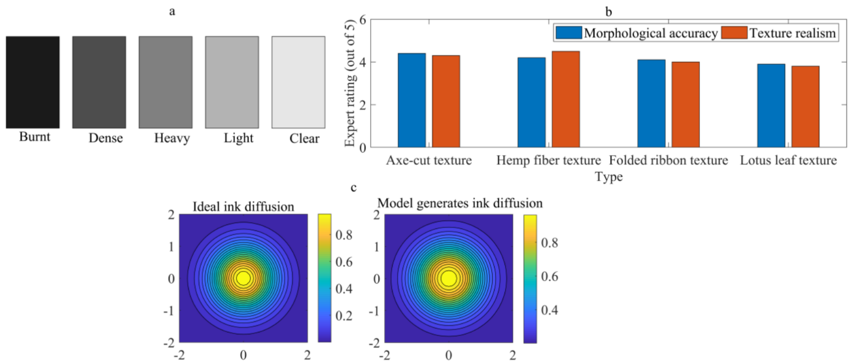

FIGURE 9. Analysis of the generation effect of traditional techniques: (a) Distribution of five shades of ink; (b) Comparison of the quality of different brushstroke techniques; (c) Deviation in the physical simulation of ink splash effect

for future improvements in the modelling of Xuan paper materials.

ABLATION EXPERIMENT

To systematically assess the contributions of each component of the multi-layer style transfer framework used in this paper, the CLIP semantic enhancement module, LoRA fine-tuning module, and ControlNet constraint module were removed or replaced, and their performance on the test set was evaluated relative to one another. Evaluation metrics were stroke structure similarity (SSIM), blank ratio accuracy, FID score, and mean expert score.

The experimental results are shown in Table 6 below:

TABLE 6. Comparison of Ablation Test Results

Model configuration	SSIM (Stroke)	Blank ratio (%)	FID (down)	Mean expert rating (5-point scale)
Complete framework (K1)	0.78	28.5	28.1	4.35
Without CLIP semantic enhancement (A)	0.65	24.8	38.2	3.42
Without LoRA fine-tuning (B)	0.71	26.3	33.5	3.78
Without controlNet constraints (C)	0.59	22.1	41.6	3.15
Original stable diffusion	0.57	21.9	35.7	3.32

The complete framework in Table 6 performs well for stroke SSIM (0.78), white space ratio (28.5% of the image), image realism (FID = 28.1), and art score (4.35/5.0), thereby demonstrating that all three layers have an effective collaborative impact. The removal of any single layer leads to considerable drops in mean performance, such that for the absence of the CLIP layer (configuration A), SSIM decreases to 0.65 and the white space ratio deteriorates 3.7% ; for the absence of LoRA adaptation (configuration B), SSIM drops to 0.71 and FID increases to 33.5; and configuration C, the absence of controlNet leads to the worst results, SSIM is 0.59, white space ratio = 22.1%, FID = 41.6, and expert art score = 3.15, which establishes the importance of CLIP and LoRA for conveying abstract aesthetics and stroke details and visual style, and finally, the contributions to traditional composition rules and layout from controlNet.

V. CONCLUSION

This paper addresses three overarching difficulties typically associated with AI-generated Chinese ink paintings regarding (1) Abstract Aesthetic Modeling, (2) Traditional Painting Technique Preservation, and (3) Maintaining the ‘Form-Spirit’ Resemblance, which are issues in the marketplace. This paper’s solutions include developing a multi-level style transfer framework based on Stable Diffusion, constructing CLIP text encoder (a specialized text encoder with enhanced terminology), lightweight LoRA fine-tuning, and ControlNet latent-space conditional structural control. Results from our experiments show stroke SSIM reached 0.78, and the white space ratio (28.5±3.2%) showed no statistically significant difference from traditional works (p>0.05, see Table 3). Cross-media validation showed 83% of works successfully blended traditional techniques with innovation. The optimal LoRA weight range was 0.8-1.0, balancing innovation with technical integrity. The framework demonstrated efficient real-time generation, indicating Stable Diffusion’s potential for both preserving and innovating traditional culture. This approach offers a scalable model for AI-generated art and could inform similar applications, such as innovating traditional motifs in the industry.

AUTHOR CONTRIBUTIONS

Conceptualization –Ying Liu, Lin Liu and Zheqing Liu; methodology – Ying Liu, Lin Liu and Zheqing Liu; investigation – Ying Liu and Lin Liu; writing-original draft preparation – Ying Liu, Lin Liu and Zheqing Liu. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research was supported by,

1. General Project of Degree and Graduate Reform of Higher Education in Hunan Province in 2022: Exploration and Practice of Aesthetic Education for Chinese Painting in Art Design Major from the Perspective of Ideology and Politics (No. 2022JGSZ067).

2. General Project of Teaching Reform of Higher Education in Hunan Province in 2020: Research on the Teaching Reform of the Course "Modern Educational Technology" for Public funded Normal University Students under the Background of Education Informatization 2.0 (No. HNJG-2020-1101).

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] J. Zhang, J. Yu, Y. Miao, and R. Peng, "Self-Adaptive Computational Aesthetic Evaluation of Chinese Ink Paintings Based on Deep Learning," Journal of Computer-Aided Design & Computer Graphics, vol. 33, no. 9, pp. 1349-1360, 2021, doi: 10.3724/SP.J.1089.2021.18815.

[2] X. Y. Zhang, "The Analysis of the Graphic Creation Teams of Digital Ink in Chinese Universities and Colleges in the Past Twenty Years," Journal of Education and Teaching, vol. 6, no. 2, pp. 15-35, 2024, doi: 10.6994/JET.202412_6(2).0002.

[3] B. Liu, "GAN-Based Scene Text Artistic Style Transfer," Computer & Digital Engineering, vol. 52, no. 5, pp. 1523-1528, 2024.

[4] C. Y. Chung and S. H. Huang, "Interactively transforming Chinese ink paintings into realistic images using a border enhance generative adversarial network," Multimedia Tools and Applications, vol. 82, no. 8, pp. 11663-11696, 2023, doi: 10.1007/s11042-022-13684-4.

[5] K. Yu, Y. Wang, S. Zeng, C. Liang, X. Bai, D. Chen, et al., "InkGAN: Generative Adversarial Networks for Ink-And-Wash Style Transfer of Photographs," Advances in Artificial Intelligence and Machine Learning, vol. 3, no. 2, pp. 1220-1233, 2023, doi: 10.54364/aaiml.2023.1171.

[6] X. Peng, S. Peng, Q. Hu, J. Peng, J. Wang, X. Liu, et al., "Contour-enhanced CycleGAN framework for style transfer from scenery photos to Chinese landscape paintings," Neural Computing and Applications, vol. 34, no. 20, pp. 18075-18096, 2022, doi: 10.1007/s00521-022-07432-w.

[7] F. Fu, J. Lv, C. Tang, and M. Li, "Multi-style Chinese art painting generation of flowers," IET Image Processing, vol. 15, no. 3, pp. 746-762, 2021, doi: 10.1049/ipr2.12059.

[8] M. Wang, Y. Shao, and H. Liu, "APST-Flow: A Reversible Network-Based Artistic Painting Style Transfer Method," Computers, Materials and Continua, vol. 75, no. 3, pp. 5229-5254, 2023, doi: 10.32604/cmc.2023.036631.

[9] X. C. Liu, Y. C. Wu, and P. Hall, "Painterly style transfer with learned brush strokes," IEEE Transactions on Visualization and Computer Graphics, vol. 30, no. 9, pp. 6309-6320, 2023, doi: 10.1109/TVCG.2023.3332950.

[10] X. Shi, X. Y. Hu, Y. H. Huang, R. Yang, and F. Z. Zhu, "Art Painting Image Style Transfer Based on Stroke and Contour Constraints," Journal of Chaohu University, vol. 25, no. 6, pp. 101-110+128, 2024.

[11] J. Y. Zeng and X. P. Shen, "Application and Research on Virtual Human Image Pre-design Based on Stable Diffusion," Modern Information Technology, vol. 8, no. 3, pp. 169-175, 2024, doi: 10.19850/j.cnki.2096-4706.2024.03.036.

[12] S. S. Li, "Application of AI Painting in Art Creation–Taking Stable Diffusion as an Example," Modern Information Technology, vol. 8, no. 8, pp. 133-137, 2024, doi: 10.19850/j.cnki.2096-4706.2024.08.029.

[13] A. Rahmatulloh, "Custom concept text-to-image using stable diffusion Model in generative artificial intelligence," JICO: International Journal of Informatics and Computing, vol. 1, no. 1, pp. 1-11, 2025, [Online]. Available: https://iaico.org/index.php/JICO/article/view/2.

[14] S. H. Lee and K. S. Song, "Exploring the possibility of using ChatGPT and Stable Diffusion as a tool to recommend picture materials for teaching and learning," Journal of the Korea Society of Computer and Information, vol. 28, no. 4, pp. 209-216, 2023, doi: 10.9708/jksci.2023.28.04.209.

[15] S. L. Liu and L. Y. Yang, "AI Drawing Helps the Development of 3D Design: Taking Stable Diffusion as an Example," Journal of Kunming Metallurgy College, vol. 41, no. 1, pp. 61-69, 2025, doi: 10.3969/j.issn.1009-0479.2025.01.009.

[16] R. Xia, Y. Y. Chang, P. F. Li, and X. Z. Wei, "Research on cultural graph technology based on Stable Diffusion and its application in multiple scenarios," Journal of Zhoukou Normal University, vol. 42, no. 2, pp. 50-54, 2025, doi: 10.13450/j.cnki.jzknu.2025.02.009.

[17] Y. C. Dou, "Research on UI design optimization based on Stable Diffusion," Science and Technology Vision, vol. 14, no. 35, pp. 61-64, 2024, doi: 10.3969/j.issn.2095-2457.2024.35.017.

[18] Y. R. Li, "On the Impact of Stable Diffusion Algorithm on the Illustration Market and Aesthetic Environment," Art Science and Technology, vol. 36, no. 15, pp. 244-246, 2023, doi: 10.3969/j.issn.1004-9436.2023.15.080.

[19] P. Gao, S. Geng, R. Zhang, T. Ma, R. Fang, Y. Zhang, et al., "Clip-adapter: Better vision-language models with feature adapters," International Journal of Computer Vision, vol. 132, no. 2, pp. 581-595, 2024, doi: 10.1007/s11263-023-01891-x.

[20] P. Shieh, M. R. Hill, W. Zhang, S. L. Kristufek, and J. A. Johnson, "Clip chemistry: Diverse (bio)(macro) molecular and material function through breaking covalent bonds," Chemical Reviews, vol. 121, no. 12, pp. 7059-7121, 2021, doi: 10.1021/acs.chemrev.0c01282.

[21] S. A. Abdullah and A. K. Jumaat, "Selective image segmentation models using three distance functions," Journal of Information and Communication Technology, vol. 21, no. 1, pp. 95-116, 2022, doi: 10.32890/jict2022.21.1.5.

[22] L. Khrissi, N. El Akkad, H. Satori, and K. Satori, "Clustering method and sine cosine algorithm for image segmentation," Evolutionary Intelligence, vol. 15, no. 1, pp. 669-682, 2022, doi: 10.1007/s12065-020-00544-z.

[23] X. Y. Liu, R. Zhu, D. Zha, J. Gao, S. Zhong, M. White, et al., "Differentially private low-rank adaptation of large language model using federated learning," ACM Transactions on Management Information Systems, vol. 16, no. 2, pp. 1-24, 2025, doi: 10.1145/3682068.

[24] F. Wu, C. Zhu, J. Xu, M. W. Bhatt, and A. Sharma, "Research on image text recognition based on canny edge detection algorithm and k-means algorithm," International Journal of System Assurance Engineering and Management, vol. 13, no. Suppl 1, pp. 72-80, 2022, doi: 10.1007/s13198-021-01262-0.

[25] L. I. Mochurad, "Canny edge detection analysis based on parallel algorithm, constructed complexity scale and CUDA," Computing and Informatics, vol. 41, no. 4, pp. 957–980, 2022, doi: 10.31577/cai_2022_4_957.

[26] Z. Xie and S. Song, "FedKL: Tackling data heterogeneity in federated reinforcement learning by penalizing KL divergence," IEEE Journal on Selected Areas in Communications, vol. 41, no. 4, pp. 1227-1242, 2023, doi: 10.1109/JSAC.2023.3242734.

[27] M. Al-Jarrah, E. Alsusa, and C. Masouros, "A unified performance framework for integrated sensing-communications based on KL-divergence," IEEE Transactions on Wireless Communications, vol. 22, no. 12, pp. 9390-9411, 2023, doi: 10.1109/TWC.2023.3270390.

[28] X. Yang, "Image-to-Image Search Method Based on CLIP Model," Computer Science and Application, vol. 15, no. 1, pp. 177-186, 2025, doi: 10.12677/csa.2025.151018.

[29] L. Cheng, R. Xiong, J. Wu, X. Yan, C. Yang, Y. Zhang, et al., "Fast segmentation algorithm of USV accessible area based on attention fast DeepLabV3," IEEE Sensors Journal, vol. 24, no. 15, pp. 24168-24177, 2024, doi: 10.1109/JSEN.2024.3410403.