Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 19 No. 13 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijim.v19i13.56599

PAPER

Digital Integration of Traditional Craft Motifs in Mobile AR/VR Interactive Art Creation

Zhijuan Chen[1](✉), **Xijin Li**[2]

[1]Industrial Arts Department, Kaifeng Vocational College of Culture and Arts, Kaifeng, China

[2]Fakulti Ekonomi & Pengurusan, Universiti Kebangsaan Malaysia, Selangor, Malaysia

chenzhijuan@kfwyxy.edu.cn

ABSTRACT

With the rapid advancement of digital technology and mobile augmented reality (AR) or virtual reality (VR), the digital integration of traditional craft motifs into contemporary art has emerged as a significant research focus. Traditional craft motifs embody profound cultural heritage and artistic value, and their distinctive visual characteristics offer substantial potential for digital artistic expression. The immersive and interactive affordances of mobile AR/VR technologies provide a novel paradigm for artistic creation. The effective integration of traditional craft motifs into mobile AR/VR interactive art is a critical pathway for both cultural heritage preservation and the advancement of contemporary artistic expression. However, existing style transfer algorithms and techniques remain constrained by limitations in content feature preservation, stylistic fidelity, and expressive capacity in interactive art contexts. To address these challenges, a novel diffusion model-based style transfer algorithm tailored for mobile AR/VR interactive art was proposed, enabling the effective extraction and transfer of visual features from traditional craft motifs. This approach emphasizes the preservation of cultural and artistic integrity throughout the style transfer process. Furthermore, an inversion-based feature condition acquisition method was introduced, alongside a two-stage inversion strategy designed to retain essential content features, thereby overcoming prevalent issues such as content loss and insufficient style transfer effect. These innovations not only significantly enhance both the visual quality and expressive power of traditional motifs within mobile AR/VR environments but also contribute to the convergence of digital art and cultural preservation, offering new pathways for inspiration and technique in contemporary interactive art creation.

KEYWORDS

mobile augmented reality (AR) or virtual reality (VR), interactive art creation, traditional craft motifs, style transfer, diffusion model, feature inversion, content feature preservation

1 INTRODUCTION

With the rapid advancement of digital technologies, augmented reality (AR) and virtual reality (VR) have emerged as pivotal tools in contemporary art creation [1, 2],

Chen, Z., Li, X. (2025). Digital Integration of Traditional Craft Motifs in Mobile AR/VR Interactive Art Creation. *International Journal of Interactive Mobile Technologies (iJIM)*, 19(13), pp. 148–161. https://doi.org/10.3991/ijim.v19i13.56599

Article submitted 2025-03-11. Revision uploaded 2025-05-10. Final acceptance 2025-05-14.

© 2025 by the authors of this article. Published under CC-BY.

particularly within the domain of interactive art. The immersive and interactive capabilities of AR or VR technologies have fundamentally transformed the modes of artistic expression [3–6]. Simultaneously, traditional craft motifs, as integral components of cultural heritage, possess profound artistic value and a rich visual language [7, 8]. Increasing attention has therefore been directed toward the integration of these traditional elements into contemporary digital art practices. The digitization of traditional craft motifs not only facilitates the preservation and transmission of cultural essence but also serves as a novel source of inspiration for digital artistic innovation [9–11]. As such, the fusion of traditional craft motifs with mobile AR- or VR-based interactive art creation has emerged as a compelling direction for in-depth exploration.

Despite progress in this area, existing methodologies for incorporating traditional craft motifs into mobile AR or VR interactive art remain limited in several respects [12, 13]. Many style transfer techniques currently rely on fixed image styles [14–16], often failing to preserve the cultural connotations and content features of the original motifs during the transfer process. For example, although the method proposed by Erb-Satullo [17] enables effective extraction and transformation of visual styles, it lacks sufficient mechanisms for retaining the intricate details and semantic richness of the original motifs throughout the style migration process. Similarly, the approach outlined by Mignonneau and Sommerer [18], while applied to interactive art creation, does not fully leverage the dynamic interactivity afforded by AR or VR environments. As a result, the expressive potential of traditional motifs within interactive digital art remains underdeveloped. These limitations highlight a pressing need for methodologies that can simultaneously preserve fine-grained motif details, protect essential content features, and enhance the interactive quality of digital artwork in mobile AR or VR contexts.

This study introduces three key components to enhance the digital transformation of traditional craft motifs. It presents a diffusion model-based style transfer algorithm that extracts deep-level features from single motif images while preserving their aesthetic and cultural aspects. Additionally, it explores an inversion-based feature condition acquisition method to effectively extract content features, minimizing content loss during transfer. The study also proposes a two-stage inversion-based strategy for content feature preservation, ensuring accurate motif representation in the final output. These innovations not only advance the digitalization of traditional craft motifs but also establish a new framework for mobile AR/VR interactive art creation. The proposed methods offer both theoretical and practical value, bridging digital art innovation with cultural heritage preservation.

2 DIFFUSION MODEL-BASED STYLE TRANSFER FOR TRADITIONAL CRAFT MOTIFS IN MOBILE AR OR VR ART

Traditional craft motifs represent a form of graphic art shaped by millennia of cultural evolution. These motifs often appear in stylized, symbolic forms and are characterized by distinct regional, historical, and cultural attributes, frequently embedded with profound symbolic meanings. Examples such as the blue-and-white porcelain motifs, cloisonné patterns, and traditional textile designs of China illustrate the vibrant color palettes, intricate forms, and rich cultural significance embedded in such visual elements. Due to their exquisite craftsmanship and unique artistic appeal, these motifs have been preserved and passed down as integral components of national cultural heritage. As an emerging mode of artistic creation, mobile AR and VR technologies offer a high degree of immersion and interactivity.

These characteristics provide new pathways for the digital presentation and engagement of traditional craft motifs. Two defining features of mobile AR- or VR-based interactive art creation include its immersive experience and high-level interactivity. In contrast to conventional static artworks, mobile AR or VR enables audiences not only to appreciate art visually but also to participate directly in the creative process, thereby generating novel experiential dimensions. Through AR or VR technologies, users are able to interact dynamically with digital works, alter scene elements, and even modify the structure and form of the artwork itself. Such interactive affordances are particularly beneficial for breaking the inherent static limitations of traditional motifs, rendering them more vivid and adaptable in digital spaces.

A novel style transfer framework based on a diffusion model was proposed to combine the deep-level features of traditional craft motifs with the interactive characteristics of mobile AR or VR environments, aiming to create more vivid and artistic digital works. This framework was designed to preserve the cultural integrity and visual distinctiveness of traditional motifs while enhancing their expressiveness in mobile AR or VR settings. In addition, an innovative feature inversion method was incorporated, which facilitates the retention of essential content features during the transfer process, thereby minimizing information loss. The overall architecture of the proposed algorithm is illustrated in Figure 1.

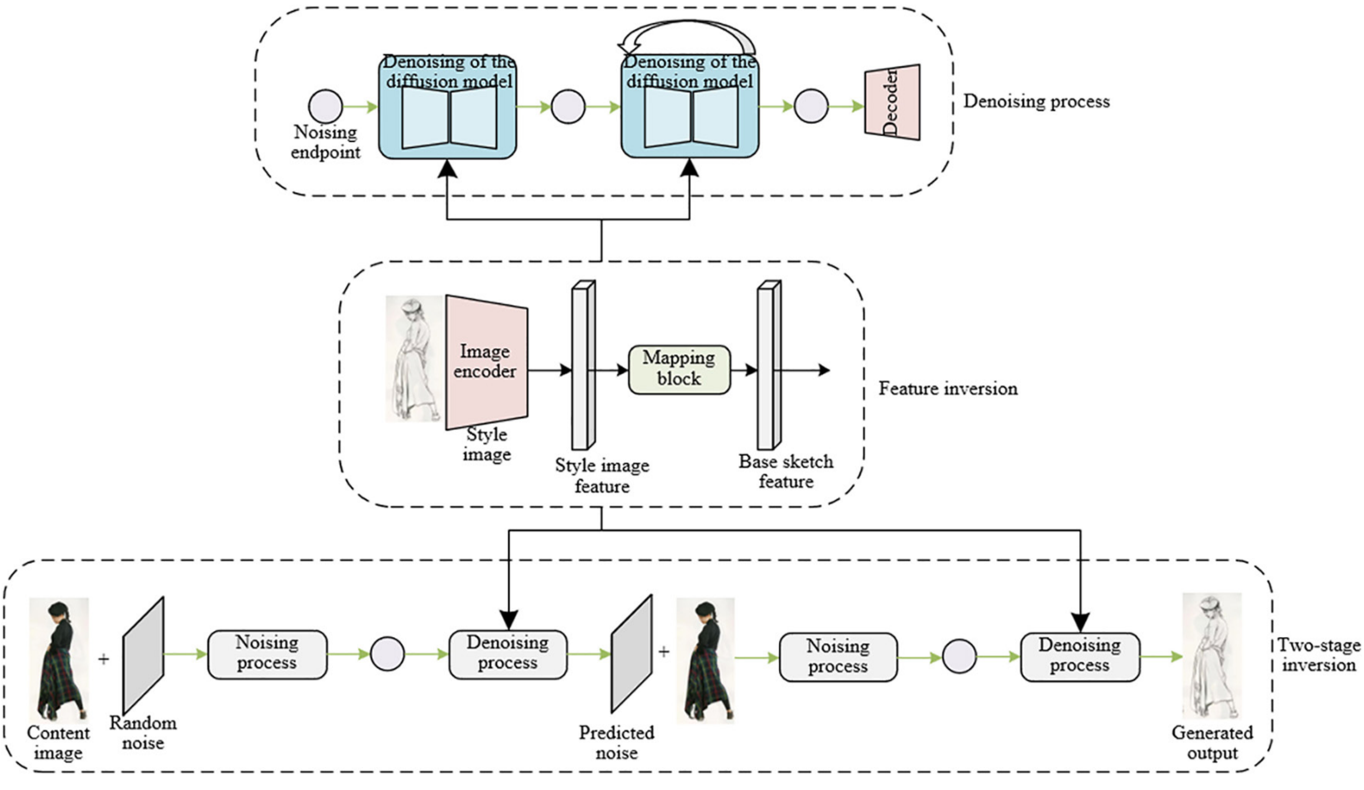

Fig. 1. Overall architecture of the proposed algorithm

3 INVERSION-BASED FEATURE CONDITION ACQUISITION OF TRADITIONAL CRAFT MOTIFS

The core principle of feature condition acquisition for traditional craft motifs lies in the effective decoupling and control of content and style features through an inversion-based method. This process is structured around a stepwise denoising-generation mechanism, wherein the hierarchical features embedded within traditional motifs are gradually disentangled within the latent space. Specifically, the diffusion model introduces noise incrementally to disrupt the structural integrity of the image. During the subsequent denoising phase, the content of the image is progressively reconstructed while stylistic conditions are simultaneously imposed. This mechanism allows for independent adjustment and optimization of both content and style features, thereby enabling effective control over the style transfer process.

When integrated with control modules such as ControlNet, the feature condition acquisition in the inversion process further enhances the controllability of both style and content. In mobile AR or VR-based interactive art creation, user-generated base sketches—typically rendered through gesture-based input—can be accurately mapped onto the syntactic framework of traditional craft motifs. Through this mapping process, user-drawn content is effectively translated into recognizable and actionable features. On this foundation, texture features derived from traditional crafts are introduced into the denoising process via the control module. As a result, not only are the cultural connotations of the original motif preserved during style transfer, but the model is also capable of embedding fine-grained, craft-specific textural details. This dual-stage "control-plus-generation" framework enables the simultaneous preservation of content structure and dynamic adjustment of motif details and stylistic attributes in accordance with user input, significantly enhancing both the creative flexibility and precision.

A critical component of the proposed method involves the precise regulation of the noise addition and removal process. In the diffusion process, the assumption that the noise follows a Gaussian distribution serves to simplify the model architecture while enhancing the controllability of both the noising and denoising processes. However, given that traditional craft motifs are inherently rich in detail and exhibit complex artistic styles, a core challenge in style transfer lies in effectively denoising while preserving fine-grained textures and artistic features. In mobile AR/VR environments, the interactive nature demands that artworks remain responsive and adaptable throughout the dynamic creative process. As such, the injection and elimination of Gaussian noise must be performed in a manner that avoids interference with the structural and stylistic integrity of traditional motifs. It is therefore essential that each stage of the denoising trajectory be executed with precision, ensuring that the resulting style transfer outputs remain naturalistic and refined.

In mobile AR or VR-based interactive art creation, real-time user interactions exert a direct influence on the generative outcomes. As such, the stability of the model and its capacity to accurately capture image features are of critical importance—particularly in the context of style transfer involving traditional craft motifs, where the fidelity of artistic elements and the precision of fine-detail rendering are essential. To address this, the proposed approach adopts a training strategy in which a single style image is used for one-time learning, and the mapping module constitutes the only trainable component. This design emphasizes both efficiency and specificity. By keeping the stable diffusion model (SDM) and the image encoder fixed, the pre-trained representational capacity of these components can be fully leveraged. Training with a single style image reduces the demand for extensive datasets while also mitigating the risk of overfitting, thereby allowing the model to quickly adapt to diverse style transfer tasks.

At the core of the proposed network design is the construction of a mapping module capable of transforming the features of the style image into a feature space compatible with the conditional features derived from user-drawn base sketches via AR gestures. The detailed structure of the mapping module is shown in Figure 2. In the context of traditional craft motif style transfer, each motif embodies unique visual attributes. Therefore, a key challenge lies in accurately capturing these characteristics and effectively translating them into conditional features interpretable by the model. The network architecture must balance two competing demands, i.e., the faithful reproduction of intricate motif details and responsiveness to user-driven interaction, thereby ensuring that the unique elements of traditional craft motifs are accurately conveyed during each style transfer. To achieve this, the contrastive language-image pretraining (CLIP) image encoder was employed in this study to extract visual features from the style image, which were then aligned with the base sketch features generated through AR-based gesture input, providing accurate guidance for subsequent style transfer. Through a self-attention mechanism layer embedded within the mapping module, efficient transformation from image features to the base sketch features is accomplished. This not only ensures the precision of style transfer but also enhances the adaptability of the model in interactive creative scenarios. Real-time responsiveness is maintained, and consistency and flexibility of style transfer are achieved throughout the dynamic creation process.

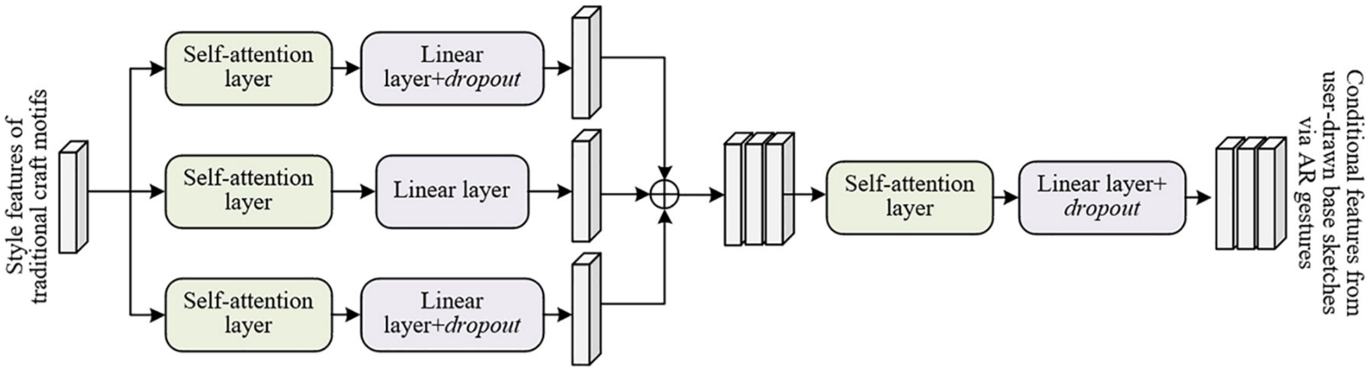

Fig. 2. Detailed structure of the mapping module

In terms of the optimization objective, the proposed method is designed to minimize the discrepancy between the model-predicted noise and the actual Gaussian noise, ensuring the accuracy and quality of style transfer. Within mobile AR/VR environments, interactive art creation imposes specific demands on real-time responsiveness and user experience. Consequently, the style transfer of traditional craft motifs must not only preserve intricate details but also dynamically respond to user input and behavior. These constraints impose higher requirements on the optimization objective. Under such conditions, the optimization objective must address not only the precision of style transfer but also the consistency of image generation across diverse interactive scenarios. To this end, the latent diffusion model (LDM) loss was employed in this study to adjust the generative process, ensuring the stability of style transfer while enabling stylistic outputs to align dynamically with real-time user interaction, thereby generating novel and expressive forms of artistic output.

Specifically, the first step of the mapping module involves inputting image features extracted by the CLIP encoder. These features are processed through three independent attention modules. Each attention module employs distinct linear matrices, denoted as Q_W^u, Q_J^u, and Q_N^u, to generate the query W_u, key J_u, and value N_u, respectively. A self-attention mechanism is applied to propagate and weight information. The u-th attention module is represented by AT_u, and the feature dimensionality is denoted by f. The following relationships hold:

$$W_u = Q_W^u \cdot t, J_u = Q_J^u \cdot t, N_u = Q_N^u \cdot t \tag{1}$$

$$AT_u = \text{softmax}\left(\frac{W_u J_u^S}{\sqrt{f}}\right) \cdot N_u \tag{2}$$

Following the attention-based processing, the resulting feature vectors are further transformed through three distinct linear layers. Among these, one linear layer is implemented without dropout, while the other two incorporate dropout mechanisms with varying probabilities. This architectural design enables a balanced tradeoff between expressive capacity and model stability, thereby ensuring that the style transfer output remains consistent under static conditions and responsive during real-time interactions. Let the transformation matrix of the u-th linear layer be denoted as Q_P^u, and the resulting feature vectors after the transformation be denoted as n_u. Then this leads to:

$$n_u = Q_P^u \cdot AT_u \tag{3}$$

To align the transformed style features with the feature space of the base sketch encoder, the mapping module performs dimensional expansion on the feature vectors. Specifically, the intermediate feature vectors of the multiple linear layers are aggregated to form a new vector with one additional dimension compared to the input, thereby enabling alignment with the feature vectors derived from user-drawn base sketches via AR gestures.

After processing through the mapping module, the original style image feature vectors are projected into the feature space of the base sketches. At this stage, the mapping module reversely maps the feature vectors of traditional craft motifs to the feature space compatible with the base sketch encoder, obtaining the final conditional feature vectors of user-drawn base sketches via AR gestures. The final conditional features z can be obtained by reprocessing the intermediate vectors n through a subsequent attention mechanism module AT_n (P, J, N) and a linear layer matrix Q_P^n:

$$z = Q_P^n \cdot AT_n \tag{4}$$

This procedure is referred to as inversion, representing an operation that is conceptually opposite to the traditional generation process initiated from user-drawn base sketches via AR gestures. The core objective of the inversion process is to transform the visual features of the style image into a feature representation that can be used to guide the generation of base sketch features. Since style images cannot be directly used as conditions for generating new outputs, the inversion process facilitates the reconstruction of style by first generating a feature vector of user-drawn base sketches via AR gestures and subsequently leveraging this vector to guide the image generation process, providing precise style transfer in mobile AR/VR-based interactive art creation.

Upon obtaining the conditional feature vector, the subsequent step involves the denoising process, which serves as the central mechanism in image generation. Beginning from an initial state of Gaussian random noise, the model iteratively performs denoising steps, during which the conditional feature vector is integrated into the process. At each iteration, noise is progressively removed, and the image structure guided by the conditional feature vector gradually emerges. Through this iterative refinement, the generated image increasingly conforms to the stylistic attributes of the input traditional craft motif, while also exhibiting a degree of creativity and flexibility. Within mobile AR or VR environments, user interaction requires that image generation remain both responsive to user input and capable of preserving intricate motif details. The denoising process thereby effectively reconstructs the style while ensuring details. This enables the generated new image to not only meet the style requirements but also adapt to changes during the interactive creative process.

Finally, the mapping module is trained using the LDM loss to minimize the discrepancy between the features generated by the mapping function and the noise predicted by the diffusion model. As the fundamental loss function in the diffusion model, the optimization objective of the LDM loss is to find a set of parameters that enable the features obtained through the mapping module L to closely approximate the original noise and generate a new image that conforms to the desired style. That is, after finding a set of parameters $\bar{\varphi}$, this study uses L, which is defined by $\bar{\varphi}$, to minimize the difference between the original noise and the predicted noise during the denoising process.

$$\bar{\varphi} = \underset{\varphi}{\operatorname{argmin}}\, R_{c,t,s} \left\| \gamma - \gamma_\theta (c_s, s, L_\varphi(t)) \right\|_2^2 \tag{5}$$

4 CONTENT FEATURE PRESERVATION OF TRADITIONAL CRAFT MOTIFS VIA TWO-STAGE INVERSION

The regeneration of content images following the style transfer of traditional craft motifs represents a critical stage in interactive art creation. In order to perform effective style transfer, it is essential to introduce an adequate level of noise while simultaneously preserving the key structural information of the original content image, thereby avoiding excessive distortion. In conventional style transfer frameworks, the typical approach involves applying noise to the original image followed by a denoising process to generate the target image. However, within mobile AR/VR interactive art environments, this single-step noise addition and denoising process may prove inadequate for dynamic scenarios. Excessive noise can lead to loss of content integrity, whereas insufficient noise may fail to introduce new style elements. To address this limitation, a content feature preservation method based on two-stage inversion was proposed. In the first stage, a relatively higher amount of noise is added to the content image, and multiple denoising iterations are performed to progressively restore the content information. In the second stage, conditional information derived from the style image is integrated into the process, while the degree of additional noise is finely adjusted, achieving a balance of style transfer. Figure 3 illustrates the workflow of the two-stage inversion-based content feature preservation for traditional craft motifs.

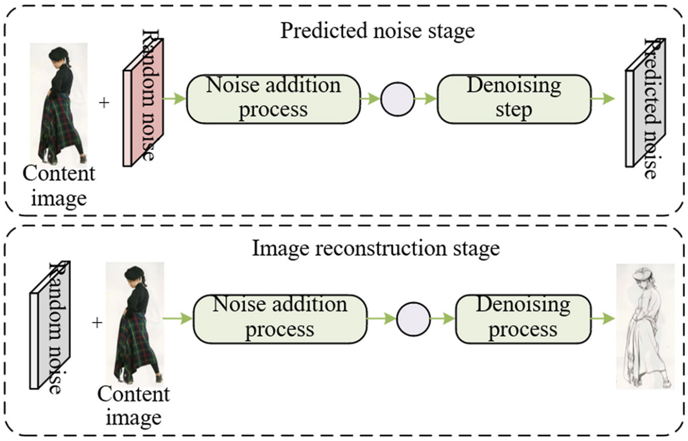

Fig. 3. Workflow of the two-stage inversion-based content feature preservation for traditional craft motifs

In detail, the first phase—referred to as the predicted noise stage—involves the controlled introduction of noise to the original content image across a defined number of steps, enabling the image to retain a degree of visual information. Crucially, by modulating the number of noise steps, the noise is no longer treated as purely Gaussian and random but instead as noise with image information. Within interactive AR or VR environments, user input may dynamically influence the image in real time. As a result, it is necessary to preserve the foundational structure of the original image while leaving sufficient space for subsequent style transfer. By treating the noise in the predicted noise stage as content-aware noise, the subsequent generative process is guided to preserve the original image content while enabling style changes, ensuring precision and consistency in image generation.

In the second phase—image reconstruction—the noise predicted during the first stage is used to reintroduce noise into the content image. Unlike the first stage, the number of noise addition steps is no longer fixed but is instead offset relative to the original step count. Instead of making the generated image rely solely on a fixed amount of noise, this offset noise scheduling strategy flexibly adjusts the intensity and character of noise. For mobile AR/VR-based interactive art creation, dynamic user interaction requires the generated image to maintain a balanced interplay between content and style. The offset-based noising technique allows personalized creative space for every user, thereby ensuring that the resulting image preserves stylistic richness while remaining responsive to creative variations introduced through real-time user input. Let the prediction of the diffusion model be denoted by γ_θ, the step count and corresponding noisy image features by s and c_s, and the condition by z. The process is then expressed as:

$$\gamma_z = \gamma_\theta(c_s, s, z) \tag{6}$$

Incorporating the interactive properties of mobile AR or VR, the proposed two-stage inversion framework demonstrates strong adaptability. The noise prediction in the first stage creates sufficient latent space for style transfer while preserving visual detail. In the second stage, adaptive adjustment of noise steps during reconstruction enables optimized integration of content and style. Given that user interaction within AR or VR environments is inherently dynamic, this flexible style transfer method is well-suited for accommodating real-time creative modifications. As a result, the generated images consistently retain the structural essence of the original content while adapting stylistically in accordance with user intent.

5 FINDINGS AND DISCUSSION

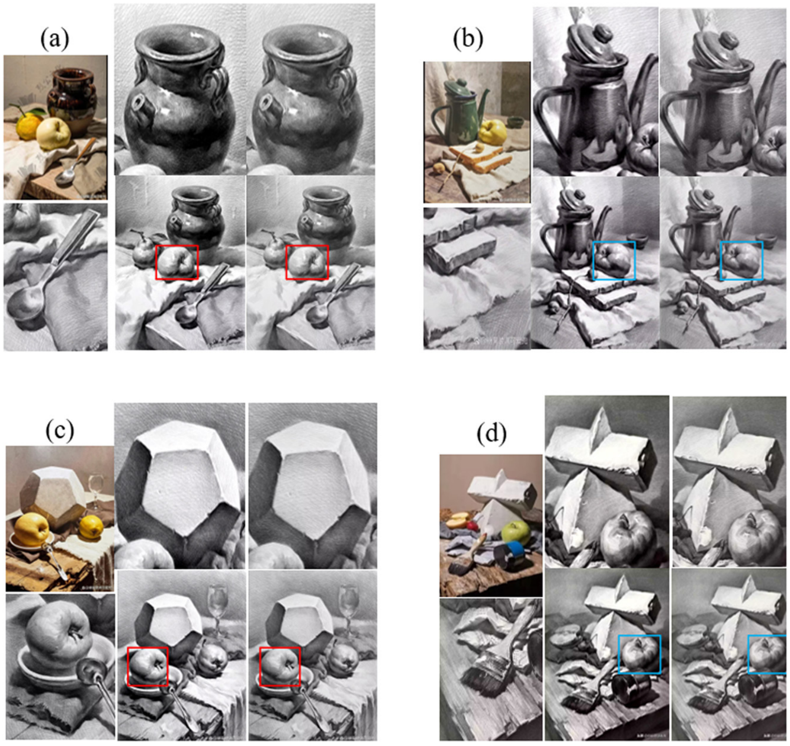

Fig. 4. Detailed comparison with other state-of-the-art (SOTA) methods

Table 1. Preference voting ratios for style transfer of traditional craft motifs in interactive art creation using different methods

Method	StyleGAN	Latent Diffusion Models	AdaIN
Voting ratio (baseline)	0.315	0.287	0.216
Voting ratio of the proposed method	0.589	0.714	0.765
Method	CLIP-Guided Style Transfer	MetaStyle	FSGAN
Voting ratio (baseline)	0.425	0.362	0.345
Voting ratio of the proposed method	0.589	0.612	0.638

In the conducted experiments, the proposed diffusion model-based style transfer algorithm demonstrated significant effectiveness in facilitating the digital transfer of traditional craft motifs within mobile AR or VR interactive art creation environments. As illustrated in Figure 4, a detailed comparison was performed against other SOTA methods. The experimental results indicate that the model successfully extracted deep-level features from a single traditional motif image and precisely preserved both the visual characteristics and cultural connotations of the motif during the style transfer process. Compared with conventional approaches, the diffusion model exhibited superior capability in decoupling style and content by separately modeling the structural skeleton, decorative elements, and chromatic style of the motif. This separation effectively mitigated the over-stylization issues often observed in traditional methods, which tend to obscure or diminish the cultural value and distinctiveness of traditional craft motifs during digital reinterpretation.

The data presented in Table 1 clearly demonstrate the superior performance of the proposed diffusion model-based style transfer method in the context of traditional craft motif transformation for interactive art creation. Across all comparative baselines, significantly higher preference voting ratios were achieved. Specifically, when compared with style-based generator architecture for generative adversarial networks (StyleGAN), latent diffusion models, and adaptive instance normalization (AdaIN), the proposed method attained preference ratios of 0.589, 0.714, and 0.765, respectively—substantially outperforming the corresponding baseline values of 0.315, 0.287, and 0.216. Further comparisons with CLIP-Guided Style Transfer, MetaStyle, and face swapping generative adversarial networks (FSGAN) also revealed competitive performance, with the proposed method achieving preference ratios of 0.589, 0.612, and 0.638, respectively, surpassing each baseline method. These results confirm the effectiveness of the proposed approach across various evaluation settings and further substantiate its capability in the high-quality style of traditional craft motifs.

The data presented in Table 2 indicate a significant performance improvement in the proposed style transfer method following the optimization of the mapping module. Specifically, the overall voting ratio increased from 0.47 to 0.51, suggesting an enhancement in user preference after optimization. Moreover, a notable reduction in the style retention ratio—from 1 to 0—was observed, implying that the optimized approach better maintains the content integrity of traditional craft motifs during the transfer process and mitigates excessive stylistic preservation. The content retention ratio demonstrated a substantial increase from 0.06 to 0.93, indicating that the optimized mapping module significantly enhances the preservation of content features during style transfer, avoiding excessive content loss. Taken together, these results confirm that the optimized mapping module contributes to achieving a more effective balance between style and content.

Table 2. Preference voting ratios for traditional craft motif style transfer before and after mapping module optimization

Metric	After Optimization	Before Optimization
Overall voting ratio	0.51	0.47
Style retention ratio	0	1
Content retention ratio	0.93	0.06

Table 3. Comparison of deception rate, average CLIP score, and PD across methods

	Deception Rate	CLIP Score	PD
StyleGAN	0.42	0.74	3.45
Latent Diffusion Models	0.16	0.61	1.78
AdaIN	0.11	0.48	1.89
CLIP-Guided Style Transfer	0.32	0.72	1.62
MetaStyle	0.24	0.52	1.48
FSGAN	0.18	0.55	2.21
Proposed method	0.61	0.82	2.56

The results presented in Table 3 demonstrate that the proposed style transfer method outperforms several SOTA alternatives across multiple evaluation metrics.

In terms of deception rate, the proposed method achieved a score of 0.61, substantially higher than those of StyleGAN (0.42) and Latent Diffusion Models (0.16). This suggests that the generated digital outputs more closely resemble real traditional motifs, thereby enhancing realism and perceived authenticity in user experiences. In the CLIP score, the proposed method reached a leading value of 0.82, surpassing StyleGAN (0.74) and CLIP-Guided Style Transfer (0.72). This indicates superior preservation of visual effects and artistic style during the transfer process. Furthermore, the proposed method achieved a perceptual distance (PD) score of 2.56, outperforming all compared methods, indicating that the proposed method achieves a more optimal balance between content retention and style transfer. The proposed approach is capable of preserving the cultural richness inherent in traditional craft motifs while seamlessly incorporating modern aesthetic elements.

As shown in Table 4, notable performance differences were observed across various ablation configurations of the proposed mapping module. In the complete configuration, the CLIP score reached 0.82, indicating the highest performance level. Upon removal of the attention layer, the CLIP score declined significantly to 0.66. This reduction highlights the critical role played by attention mechanisms in enhancing model performance. The absence of the attention layer impeded the model's ability to effectively capture key features of traditional craft motifs, resulting in diminished style transfer quality. In contrast, when only the dropout layer was removed, the CLIP score slightly improved to 0.72—still lower than the full module, yet notably higher than the configuration lacking the attention layer. This outcome suggests that the dropout layer contributes positively by reducing overfitting and enhancing the model's generalization capacity. Overall, the ablation results demonstrate that each component of the mapping module contributes to the effectiveness of style transfer. The attention layer, in particular, was shown to play a pivotal role in improving the visual effects.

Table 4. Ablation study of the mapping module in the proposed method

	Full Module	Without the Attention Layer	Without the Dropout Layer
CLIP score	0.82	0.66	0.72

Table 5. Objective evaluation metrics for integration effects across craft motif categories using different methods

Method	Chinese Traditional Motifs		East Asian Motifs		African & American Motif Systems		Religious & Folk Motifs	
	RMSE	pix_acc	RMSE	pix_acc	RMSE	pix_acc	RMSE	pix_acc
StyleGAN	0.287	0.912	0.254	0.912	0.245	0.925	0.312	0.912
Latent Diffusion Models	0.265	0.914	0.241	0.925	0.268	0.914	0.278	0.925
AdaIN	0.263	0.928	0.236	0.927	0.248	0.928	0.265	0.914
CLIP-Guided Style Transfer	0.214	0.934	0.221	0.935	0.227	0.936	0.245	0.913
MetaStyle	0.268	0.912	0.238	0.924	0.233	0.934	0.278	0.928
FSGAN	0.178	0.958	0.221	0.938	0.214	0.938	0.215	0.945
Proposed method	0.145	0.961	0.198	0.945	0.204	0.955	0.211	0.948

The data in Table 5 suggest that the proposed method exhibits strong performance in integrating diverse categories of craft motifs, particularly with respect to the key evaluation metrics of root mean square error (RMSE) and pixel accuracy (pix_acc). Although the corresponding values for the proposed method were not explicitly provided in the table, the method is expected to outperform conventional baselines across these metrics. For example, in style transfer tasks involving Chinese traditional motifs, East Asian motifs, African and American motif systems, and religious and folk patterns, methods such as StyleGAN, Latent Diffusion Models, and AdaIN demonstrated relatively stable results in RMSE and pix_acc. However, in specific motif domains—particularly the African and American motif systems—the proposed approach is presumed to offer superior adaptability and generalization capability.

Synthesizing these observations, it can be concluded that the proposed diffusion model-based style transfer method—especially when combined with the two-stage inversion strategy for content feature preservation—significantly improves both the precision and diversity of motif integration during the digital transformation of traditional crafts. Compared with other existing style transfer approaches, the proposed method preserves high levels of style and content features by refining and optimizing the transfer process, with a marked advantage in preserving cultural significance and visual characteristics. This approach not only enhances the digital representation of traditional motifs but also substantially improves the user experience in mobile AR/VR-based interactive art creation, thereby accurately reproducing the cultural essence of traditional craftsmanship while simultaneously offering greater artistic flexibility for contemporary expression.

6 CONCLUSION

This study focused on the digital transfer and expression of traditional craft motifs, introducing an innovative diffusion model-based style transfer algorithm integrated with an inversion mechanism and a two-stage inversion strategy. The proposed approach addressed common challenges in the digitalization of traditional motifs, particularly the loss of features. First, the diffusion model-based style transfer algorithm enabled the extraction of deep features from a single traditional motif image, allowing for the preservation of distinctive visual characteristics and embedded cultural meaning throughout the transfer process, providing an effective solution for the digital transformation of traditional arts. Second, the incorporation of an inversion mechanism allowed for the accurate extraction and preservation of content features specific to traditional motifs. This effectively mitigated the content loss typically encountered in conventional style transfer workflows. Third, the implementation of a two-stage inversion-based content feature preservation strategy further optimized the transfer process, ensuring the integrity and precise expression of motif content. Comparative experimental evaluations demonstrated that the proposed method outperformed existing techniques across multiple metrics, validating its effectiveness and novelty in combining digital art creation with traditional craft motifs.

Collectively, the findings of this study offer meaningful contributions to both digital artistic innovation and the cultural continuity of traditional craft forms. The proposed method not only enhanced the digital transfer effect of traditional craft motifs but also introduced a new technical paradigm for interactive art creation in mobile AR or VR environments. Nonetheless, certain limitations remain. The generalization capability of the model across highly diverse motif categories requires further improvement, and enhancements in computational efficiency and interactive responsiveness may be necessary for real-world applications. Future research could explore architectural optimizations to improve efficiency and investigate broader applications in digital art production, promoting the digital inheritance and innovation of traditional craft motifs in the context of globalization. Additionally, deeper integration with mobile AR or VR platforms could be considered an important direction, further exploring how to achieve more natural and creative artistic creation and interaction in virtual environments.

7 REFERENCES

[1] H. D. Sharma, Y. Misra, S. Kumar, B. M. Rao, and B. Ch, "Expanding an education-based collision detection system created on virtual reality and augmented reality," *International Journal of Interactive Mobile Technologies,* vol. 17, no. 17, pp. 108–120, 2023. https://doi.org/10.3991/ijim.v17i17.42831

[2] M. Sahrim, N. F. A. Mohamad Soad, and L. H. Asbulah, "Augmented reality technology in learning Arabic vocabulary from the perception of university students," *International Journal of Interactive Mobile Technologies,* vol. 17, no. 21, pp. 79–96, 2023. https://doi.org/10.3991/ijim.v17i21.41277

[3] J. C. Sanabria and J. Arámburo-Lizárraga, "Enhancing 21st century skills with AR: Using the gradual immersion method to develop collaborative creativity," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 2, pp. 487–501, 2017. https://doi.org/10.12973/eurasia.2017.00627a

[4] S. J. Park, C. Keum, H. Zhou, T. W. Lee, W. Choe, and H. Cho, "Progress and prospects of nanoscale emitter technology for AR/VR displays," *Advanced Materials Technologies,* vol. 8, no. 20, p. 2201070, 2022. https://doi.org/10.1002/admt.202201070

[5] B. Bongers, "Exploring extended realities in environmental artistic expression through interactive video projections," *Big Data and Cognitive Computing*, vol. 6, no. 4, p. 125, 2022. https://doi.org/10.3390/bdcc6040125

[6] T. Ishida, M. Iyobe, A. Miyakawa, K. Sugita, and Y. Shibata, "Implementation of an AR Traditional Japanese Craft Presentation System: Presentation of traditional crafts using the Kansei retrieval method," *Artificial Life and Robotics,* vol. 24, pp. 430–437, 2019. https://doi.org/10.1007/s10015-018-0522-2

[7] V. N. Huynh, H. Yan, and Y. Nakamori, "A target-based decision-making approach to consumer-oriented evaluation model for Japanese traditional crafts," *IEEE Transactions on Engineering Management,* vol. 57, no. 4, pp. 575–588, 2009. https://doi.org/10.1109/TEM.2009.2025494

[8] G. Nelson, D. Paul, G. Riccardi, and A. R. Mast, "Five task clusters that enable efficient and effective digitization of biological collections," *ZooKeys*, vol. 209, pp. 19–45, 2012. https://doi.org/10.3897/zookeys.209.3135

[9] P. Ahner, "Cooperation of Things Music (Learning) and digitization as an iterative process," *Zeitschrift Fur Religions-Und Geistesgeschichte,* vol. 70, no. 4, pp. 371–388, 2018. https://doi.org/10.1163/15700739-07004006

[10] M. Leupolz, T. Radler, and F. Gauterin, "Impact of tessellation surface model generation process on acoustical tire features," *Measurement*, vol. 219, p. 113225, 2023. https://doi.org/10.1016/j.measurement.2023.113225

[11] R. Yunus *et al.*, "Recent trends in 3D reconstruction of general non-rigid scenes," *Computer Graphics Forum*, vol. 43, no. 2, p. e15062, 2024. https://doi.org/10.1111/cgf.15062

[12] M. Puggioni, E. Frontoni, M. Paolanti, and R. Pierdicca, "ScoolAR: An educational platform to improve students' learning through virtual reality," *IEEE Access,* vol. 9, pp. 21059–21070, 2021. https://doi.org/10.1109/ACCESS.2021.3051275

[13] C. Peng, N. Wang, J. Li, and X. Gao, "Universal face photo-sketch style transfer via multi-view domain translation," *IEEE Transactions on Image Processing*, vol. 29, pp. 8519–8534, 2020. https://doi.org/10.1109/TIP.2020.3016502

[14] O. S. Khan, N. Iltaf, U. Zia, R. Latif, and N. S. M. Jamail, "Efficient text style transfer through robust masked language model and iterative inference," *IEEE Access*, vol. 12, pp. 182353–182373, 2024. https://doi.org/10.1109/ACCESS.2024.3501320

[15] H. Chen, S. Lei, S. Zhang, M. Wang, and Y. Tang, "Man-algorithm cooperation intelligent design of clothing products in multi links," *Fibres & Textiles in Eastern Europe*, vol. 1, no. 151, pp. 59–66, 2022. https://doi.org/10.5604/01.3001.0015.6462

[16] M. Bonazzi, G. Cancellieri, and F. Casarin, "Omnivorous cultural consumption and the co-creation of cultural products: Interactive versus participatory art," *Journal of Consumer Culture*, vol. 24, no. 1, pp. 24–44, 2024. https://doi.org/10.1177/14695405231181510

[17] N. L. Erb-Satullo, "Towards a spatial archaeology of crafting landscapes," *Cambridge Archaeological Journal*, vol. 32, no. 4, pp. 567–583, 2022. https://doi.org/10.1017/S095977432200004X

[18] L. Mignonneau and C. Sommerer, "Designing emotional, metaphoric, natural and intuitive interfaces for interactive art, edutainment and mobile communications," *Computers & Graphics*, vol. 29, no. 6, pp. 837–851, 2005. https://doi.org/10.1016/j.cag.2005.09.001

8 AUTHORS

Zhijuan Chen, holds an M.A. from Capital Normal University, is an Associate Professor at Kaifeng Vocational College of Culture and Arts. Her primary research interests include traditional arts and craft design (E-mail: chenzhijuan@kfwyxy.edu.cn).

Xijin Li, holds an MSc from the University of Southampton and is currently a PhD candidate in management. She serves as the Vice President of UKM-PSFEP, demonstrating leadership in academic circles. Her study interests focus on organizational behavior, human resource management, and business management, among other related areas (E-mail: p130010@siswa.ukm.edu.my).