



Deep Reconstruction of Image Style for Intangible Cultural Heritage and Adaptive Enhancement Model for Visual Features

Lili Zhang¹ and Yu Fang^{1,*}

¹ Jingchu University of Technology, Jingmen City, Hubei Province, China 448000

SUMMARY: *Digital reproduction of ICH images should consider both the style fidelity and cultural semantic expression. Given that the existing methods are prone to losing the unique visual grammar of ICH in reconstruction, this paper proposes a joint model for deep reconstruction of ICH artistic image style and adaptive enhancement of visual features to address this problem. Gram matrix is used to extract the embedding of ICH style from the model, and then combined with the improved ICH-AdaIN mechanism to achieve decoupled reconstruction of content and style. An adaptive vision-enhancement module that is guided by cultural attention will be added to strengthen the foundation of culture. Experiments have been conducted on 1,280 image datasets of the four types of ICH, such as New Year pictures and paper-cuts; compared with the U-Net baseline, the proposed method reduced LPIPS-style by 27% and FID by 15.3, increased local SSIM to 0.91 in key areas, and reduced the standard deviation of background noise by 12%. The expert's score reached 4.3/5 and had risen from the original 2.8/5. According to the above results, the model can reconstruct high-quality images that meet the visual requirements of ICH, and is thus suitable for use in digital protection and activation applications.*

KEYWORDS: *Intangible heritage; Image style reconstruction; Adaptive enhancement; ICH-AdaIN; Cultural attention*

1 Introduction

Intangible Cultural Heritage (ICH) is a necessary link in the continuous flow of life for the entire country; it is a storehouse of cultural memory and also contains beautiful aesthetic ideas, technological wisdom and spiritual symbols (Li et al., 2025). Digital technology is expected to be implemented soon, and thus serves as an effective way to promote and operate ICH. Most of the digital works at present are still in a state of "static preservation", such as high-precision scanning and database archiving. Although it can show us what life is like, it does not show us its internal dynamism and beauty in this way (Wang et al., 2023). Generative Adversarial Networks (GAN) and its variations are now capable of generating very realistic images. Style Transfer Algorithms can also use brushstrokes

Van Gogh or Monet's works, etc., are all photographs. The above results appear to be ready-made tools for regenerating ICH images. However, in practice, the general model often substantially alters the original structure of the image during the transfer process (Xiao et al., 2024). At the same time, most current studies of ICH art simplify "style" to colour division or pattern display and overlook the rich, formalised and symbolic visual systems in such art (Cao et al., 2021). The fundamental characteristics of this type of works are repeated patterns, fixed

*200608024@jcut.edu.cn

<https://doi.org/10.65102/is2026943>

composition rules, particular colour schemes and cultural connotations (Song, 2020).

The two basic deficiencies of the prior art are as follows: First, the model design does not align with the characteristics of the artistic noumenon in ICH. Most of the mainstream style transfer frameworks are focused on optimizing the style of Western paintings or photographs, and their loss functions and network structures do not consider the co-expression requirements of high-frequency details and low-frequency structures in ICH images (Zhao, 2025). Second, strengthen the separation of strategy and cultural context. Traditional image enhancement methods are generally applied uniformly around the world, and it is not feasible to identify and enhance the visual components with significant cultural connotations selectively (Liang et al., 2020). Therefore, a new computing model that can both accurately reconstruct the original artistic style of intangible cultural heritage (ICH) and enhance its essential visual features adaptively is urgently needed. Therefore, the relevant research should move away from simple pixel-level fitting and begin to study the "visual language" of ICH images in terms of understanding and reconstruction. The so-called "deep reconstruction" refers to high-resolution output and, at the same time, to the faithful reproduction of the internal structure of the style during generation (Su et al., 2022). Including but not limited to: cultural conventions of the colour system, technological characteristics of lines, and symbolic logic of patterns (Zuo et al., 2022). Adaptive enhancement refers to adjusting the degree of enhancement in regions with a high semantic value in the image, and thus preventing the loss or damage of cultural symbols during digital reproduction.

With the deepening of research on image processing of cultural heritage, many scholars have put forward various approaches from the perspectives of image retrieval, image quality enhancement, classification and recognition, and information hiding. Chigurupati Navyasree (2020) has created a good algorithm based on several features to solve the retrieval problem for minority ICH images. Li and others (2025) focused on the quality of image reconstruction and proposed a deblocking algorithm that combines gradient consistency (GC) and scale-space reconstruction (SSR) to improve visual coherence significantly; it also provides a reference for high-fidelity ICH image restoration. Cao et al. (2025) have also addressed the problem of image understanding by taking an open-set scene and introducing an angular interval loss for hardness perception and a label noise filtering mechanism to improve the classification robustness of the model under the interference of unknown categories. El-Hoseny et al. (2024) have integrated image enhancement and information security to put forward the Stego-OptDehaz algorithm, which can reduce smog and embed metadata simultaneously while satisfying the requirements of good visual clarity and information steganography. Generally speaking, the existing work focuses on a single task and lacks a systematic model for the compound goal of "style reconstruction-feature enhancement-semantic preservation" of ICH images.

A joint model for depth reconstruction and adaptive enhancement of visual characteristics in the style of ICH art will be proposed in this paper. At the core of this model is: on the one hand, by incorporating the prior knowledge of ICH, a style-decoupling representation mechanism is constructed; thus, in the reconstruction process, it can separately address the content structure and style attributes without mutual interference. On the other hand, a type of attention module for cultural perception has been designed that is guided by typical visual elements of ICH to achieve selective enhancement of important areas. This paper does not seek to substitute the inheritor's skills with algorithms, but wishes for digital technology to genuinely understand and respect the visual logic of intangible cultural heritage (ICH) and thus provide more culturally authentic visual support in scenes of preservation, education and innovation.

2 Related theory and technical basis

ICH art images are not the same as other visual materials, and they have their own paradigms and symbols. The northern woodblock New Year pictures are a typical case; they are not realistic, but rather serve folk functions such as blessings and exorcism through highly stylized visual language (Zuo & Liu, 2023). Character Modelling focuses on the proportions of the large head and small body, and composition generally adopts a symmetrical and full-bodied form. Among them, some colours are also traditionally linked to specific symbols (Alkishriwo, 2020). Paper-cutting is also such; it achieves the effect of positive and negative shapes through the structure of connecting shapes with lines and supporting lines with surfaces (Paul, 2023). Only the edge detection algorithm is employed for contour extraction; thus, the beauty of Yin and Yang is lost. For this reason, before technical modelling, it is necessary to be clear that the style of an ICH image is not a statistical texture or hue distribution, but a set of visual grammar systems that contain cultural semantics. Therefore, in the following feature extraction and reconstruction stages, we need to be more conscious of these structural and symbolic features and aim to optimise them rather than just optimising pixel similarity (Zhang et al., 2025).

GAN has significantly improved the visual realism of synthetic images through the game between discriminator and generator, but the training process is unstable and it is easy to ignore the rationality of low-frequency structure (Ding et al., 2024). The StyleGAN series introduced later supports hierarchical style control to modify coarse-grained and fine-grained features at different stages during the generation process, which is suitable for layered reconstruction of ICH images. StyleGAN's style space is generated by unsupervised learning, and its hidden variables do not directly correspond to cultural features that can be easily observed by people, such as palindrome patterns or indigo main colours (Zhang, 2024). Although AdaIN (Adaptive Instance Normalisation) and other style transfer methods can achieve quick style fusion, they may rely too heavily on global statistics for style definition and fail to address local areas with high semantic information in ICH images; as a result, they often lead to an "average style" and lose key details (Xia et al., 2025). Therefore, it is not feasible to meet the two demands of style uniformity and detail preservation for ICH images by directly employing the existing framework.

New ideas for solving the above problems have been proposed in the development of visual feature enhancement technology. Traditional image enhancement usually uses manually designed filters or global adjustments of parameters, and is not content-aware. In recent years, the attention mechanism has been introduced and thus this problem has been resolved (Zhou et al., 2024). CBAM (Convolutional Block Attention Module) is composed of channel attention and spatial attention in series, and thus can dynamically focus on important feature channels and locations (Iseringhausen et al., 2020). If the above mechanism is combined with ICH visual prior, it can help the model realise an improvement in cultural orientation, and thus automatically identify and emphasise visual areas with cultural significance during the process of image reconstruction (Wang et al., 2020). However, the general-purpose attention mechanism is still based on data-driven learning. If the ICH samples in the training data are too few or the labels are absent, the attention weights may not align with the cultural logic (Xiang et al., 2020). Therefore, ICH knowledge guidance should be added to the attention module explicitly.

The absence of an evaluation mechanism is also one of the main reasons for the landing of technology. At present, the main methods used to measure the general similarity or perceptual quality of image generation tasks are FID (Fréchet Inception Distance), LPIPS (Learned Perceptual Image Patch Similarity) and others, but these indicators cannot determine whether the generated New Year pictures meet the requirements of tradition. Therefore, only the

automation indicators are insufficient to demonstrate the actual results of the model in terms of culture. Therefore, it should be clear in the theoretical preparation stage that the technical goal is to produce pictures with good visual effects and also to create pictures that align with cultural cognition and acceptance.

3 ICH art image style deep reconstruction model

The style of images for intangible cultural heritage art is not entirely based on simple colour or texture variations. Yangliuqing New Year paintings are a typical case; the four stages of their creation are outlining, engraving, printing and painting, and they have a visual feature of strong lines and soft colours (Dong et al., 2025). Weixian County's paper-cutting art is based on a hollow-out method of "yin engraving as the primary approach and yang engraving as an auxiliary method" to create image structures. If only general-purpose generative models are used for reconstruction, it is easy to reduce the visual ideas in these crafts to imitations of surface textures. Therefore, a deep reconstruction network based on style disentanglement was designed to accurately integrate the distinct stylistic semantics of ICH into the reconstruction process and preserve the structural information of the original image.

At the beginning of reconstructing digital images, the content and style of the input ICH image need to be separated effectively. During reconstruction, the proposed depth network will analyze the semantic structure of the input image in real time, dynamically modulate the feature representation of each layer in combination with a reference style embedding, and thus generate high-quality output with the native ICH style in pixel space. Finally, a complete-structured and consistently-styled ICH artistic image can be reconstructed on the image plane. In order to enhance the realism of the reconstruction results and restore the visual grammar of ICH works more accurately, a multi-scale feature fusion and style perception mechanism need to be introduced in the network. However, this will result in an increase in the number of model parameters and computational complexity that cannot be met by the resources of a typical deployment environment. Therefore, the lightweight attention guidance strategy introduced in this paper is employed to address the above problem (as shown in Figure 1).

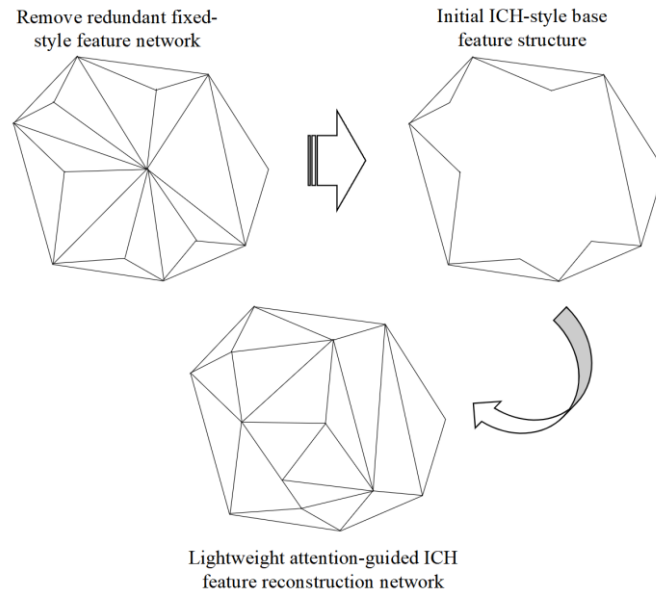


Figure 1: Adaptive Feature Enhancement Mechanism

The model is based on U-Net, because its encoder-decoder structure is naturally suitable for image-to-image mapping task, and jump connection can effectively transmit high-frequency details. However, the traditional U-Net lacks the explicit modeling ability of "style". For this reason, ICH style guidance mechanism is introduced into its hop connection path. Given the input image X , the encoder $E(\cdot)$ extracts the content features layer by layer:

$$F_i^c = E_i(F_{i-1}^c), \quad i = 1, 2, \dots, N \quad (1)$$

Among them, $F_0^c = X$. These features mainly carry the "content" information such as the geometric composition and object layout of the image, while the style needs to be independently extracted from the high-quality ICH reference image S .

In this study, Gram matrix is used as style embedding, because it can effectively capture the co-occurrence relationship of repeated patterns in ICH. If $\phi(S)$ is the feature response of the relu4₁ layer in VGG-19, then the style representation is:

$$G(S) = \frac{1}{CHW} \sum_{h,w} \phi(S)_{:,h,w} \cdot \phi(S)_{:,h,w}^T \quad (2)$$

This matrix is independent of space and can thus be translated to fit the translational invariance of ICH patterns perfectly.

The key innovation lies in the ICH style adaptive normalization (ICH AdaIN) during the decoding stage. Different from the standard AdaIN, which only uses the global mean and variance for style transfer, this model makes style embedding G generate affine parameters $\gamma_i(G)$ and $\beta_i(G)$ through a lightweight MLP, and modulates the jumping features of each layer at the channel level:

$$\text{ICH - AdaIN}(F_i^c, G) = \gamma_i(G) \cdot \frac{F_i^c - \mu(F_i^c)}{\sigma(F_i^c)} + \beta_i(G) \quad (3)$$

This way, different channels in the reconstruction process can be strengthened at different times. For example, it enhances the contrast between the "red" and "black" channels in nianhua (Chinese New Year painting) reconstruction and strengthens the separation between "positive shape" and "negative space" channels in paper-cutting restoration.

The final output image \hat{Y} is generated by the decoder $D(\cdot)$:

$$\hat{Y} = D(\text{ICH - AdaIN}(F_1^c, G), \dots, \text{ICH - AdaIN}(F_N^c, G)) \quad (4)$$

In order to consider structural fidelity, style consistency and visual realism, the loss function is a multi-objective weighting as shown below:

$$\mathcal{L}_{\text{rec}} = \lambda_1 \|\hat{Y} - Y\|_1 + \lambda_2 \mathcal{L}_{\text{style}} + \lambda_3 \mathcal{L}_{\text{adv}} + \lambda_4 \mathcal{L}_{\text{perc}} \quad (5)$$

Among them, $\mathcal{L}_{\text{style}} = \|G(\hat{Y}) - G(Y)\|_F$ constrains the consistency of style statistics, \mathcal{L}_{adv} is the PatchGAN discriminative loss to improve local texture authenticity, and $\mathcal{L}_{\text{perc}}$ uses VGG features to ensure that high-level semantic structures are not destroyed.

4 Visual characteristics AVEM

A reconstruction model cannot be used to show the cultural core of an ICH image. For instance, in a figure of shadow play, the viewer's eyes will automatically go to the face, the pattern of the

hat, and the held weapons. These places have roles, identities and stories. If the entire image is uniformly increased, it will be less prominent as a cultural relic (Dong & Zhang, 2025). Therefore, AVEM, a visual feature, is proposed and added to the end of the decoder to achieve "culture-sensitive" detail enhancement.

Firstly, in practice, it is necessary to standardise the training data set and ensure consistency of style labels; then, according to the characteristics of ICH art types, set the corresponding style embedding extraction method and loss weight; next, use the constructed depth reconstruction network to perform content-style collaborative generation of the input image; finally, carry out visual quality assessment and cultural semantic verification of the output results, such as a style fidelity check and a key pattern integrity analysis. The general flow of the proposed ICH image reconstruction and enhancement is shown in Figure 2.

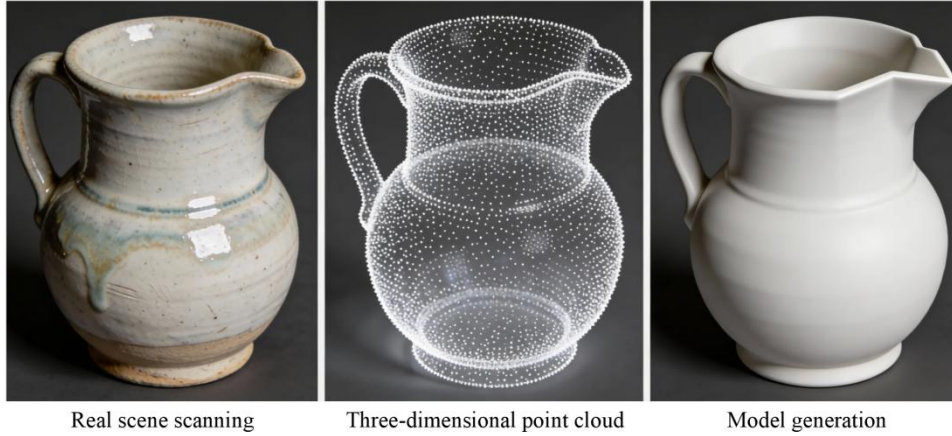


Figure 2: Depth Reconstruction and Enhancement of ICH Images

AVEM contains two collaborative components. The first is the Cultural Attention Generator (CAG), whose goal is to learn the "visual center of gravity" in ICH images. Given the reconstructed intermediate feature F_{mid} , CAG predicts the spatial attention map through two convolution layers:

$$A = \sigma(\text{Conv}_{1 \times 1}(\text{Conv}_{3 \times 3}(F_{mid}))) \quad (6)$$

In the training stage, weak supervision signal is introduced: Through the pre-labeled key areas of ICH images as attention guidance, A automatically focuses on the parts with high semantic value. This design avoids the high cost of manually labeling the whole picture, and ensures the alignment of attention and cultural semantics.

Subsequently, the Multi Scale Feature Enhancer (MFR) utilizes A to perform differential processing on high and low frequency features. Let F_{high} be the high-resolution feature and F_{low} be the low-resolution feature. Firstly, the attention map is sampled to the corresponding scale, and the adaptive weight is generated:

$$W_{high} = 1 + \alpha \cdot \text{Up}(A), \quad W_{low} = 1 - \beta \cdot \text{Up}(A) \quad (7)$$

where α, β are learnable scalars, which ensure that the key areas get stronger detailed response, while the non-core areas are moderately smooth, and prevent the background noise from being amplified.

The augmented attributes are combined element by element with weights:

$$F_{\text{enh}} = W_{\text{high}} \odot F_{\text{high}} + W_{\text{low}} \odot \text{Up}(F_{\text{low}}) \quad (8)$$

Therefore, it will be more feasible in practice: when processing an embroidery image, AVEM can automatically enhance the texture of the long line coiled with gold thread and only lightly sharpen the plain background fabric; thus, the traditional aesthetic of "distinguishing primary from secondary" will be visually achieved.

To add the improvement effect to the end-to-end optimisation, an enhancement consistency term is added to the total loss function:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{rec}} + \lambda_5 \mathcal{L}_{\text{enh}} \quad (9)$$

Wherein the \mathcal{L}_{enh} structure is:

$$\mathcal{L}_{\text{enh}} = \|\text{SSIM}(\hat{Y}_{\text{enh}}, Y) - 1\|_1 + \gamma \|\nabla \hat{Y}_{\text{enh}} - \nabla Y\|_1 \quad (10)$$

The SSIM term ensures local structural similarity, while the gradient term ∇ forces edge alignment to avoid introducing false edges or blurring the true contour during the enhancement process.

By integrating deep reconstruction and adaptive enhancement seamlessly, the model can restore ICH images and at the same time highlight their cultural connotations to achieve both technical reliability and aesthetic value in the digital reproduction results.

5 Experimental design and result analysis

A series of experiments have been carried out on a self-constructed dataset of ICH art images to systematically assess the impact of the proposed model for digital regeneration. The four types of ICH in this dataset are: Yangliuqing New Year paintings, Weixian paper-cutting, Xiang embroidery, and Shaanxi shadow puppetry; there are a total of 1,280 image pairs. Among them, 80% were used for training and 20% were used for testing. All of the above experiments were performed on an NVIDIA RTX 4090 GPU and implemented using PyTorch. To present the actual contributions of each link in the model more intuitively, at three different stages of style embedding extraction, depth reconstruction and adaptive enhancement, the time consumption and output image quality indexes were recorded, and qualitative analysis was performed by visual comparison.

As shown in Figure 3, the size of the style features is controlled by selecting a VGG feature layer from which to compute the Gram matrix. With the increase in feature dimension, the time required for style embedding extraction rose from 82ms to 146ms on average, but the style consistency of reconstructed images significantly improved and LPIPS-style decreased from 0.31 to 0.19. Fine-grained style statistical modelling can effectively learn the complex rules of colour matching and pattern repetition in ICH, but it increases the computational cost by as much as 78 per cent. Therefore, according to the actual deployment environment, either efficiency or accuracy at the feature level can be selected flexibly.

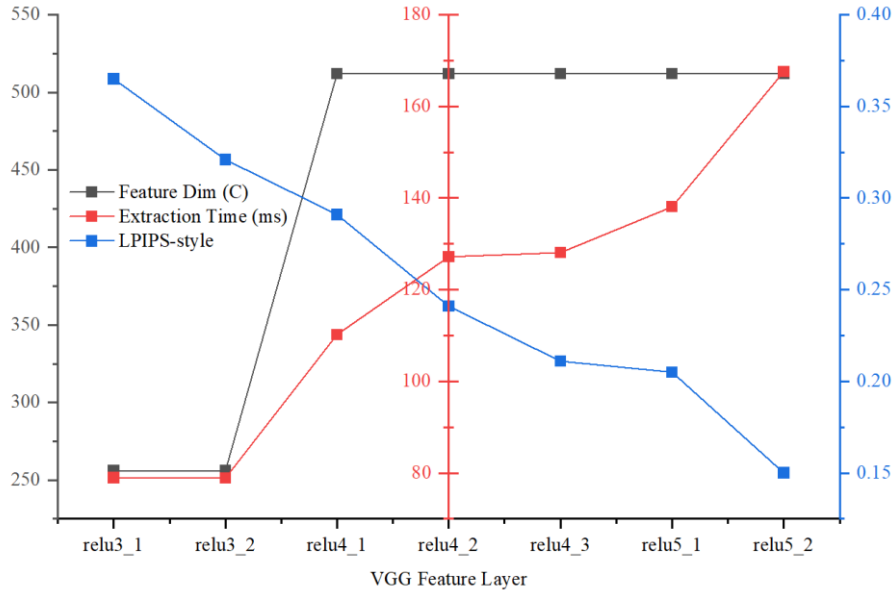


Figure 3: Relationship between Style Embedding Extraction Time and Reconstruction Quality

Figure 4 shows the two reconstruction methods: the baseline model using only L1 content loss and the full reconstruction module that adds style-aware loss. The results show that although the baseline model can restore the rough structure of the composition, the colours are generally grey and the patterns are unclear; after adding ICH AdaIN, the reconstructed pictures look much more like the originals. Therefore, ICH-style elements can be introduced in the form of affine parameters for skip connections to guide the network in generating visual expressions that comply with traditional aesthetics and maintain the structural integrity of the content.

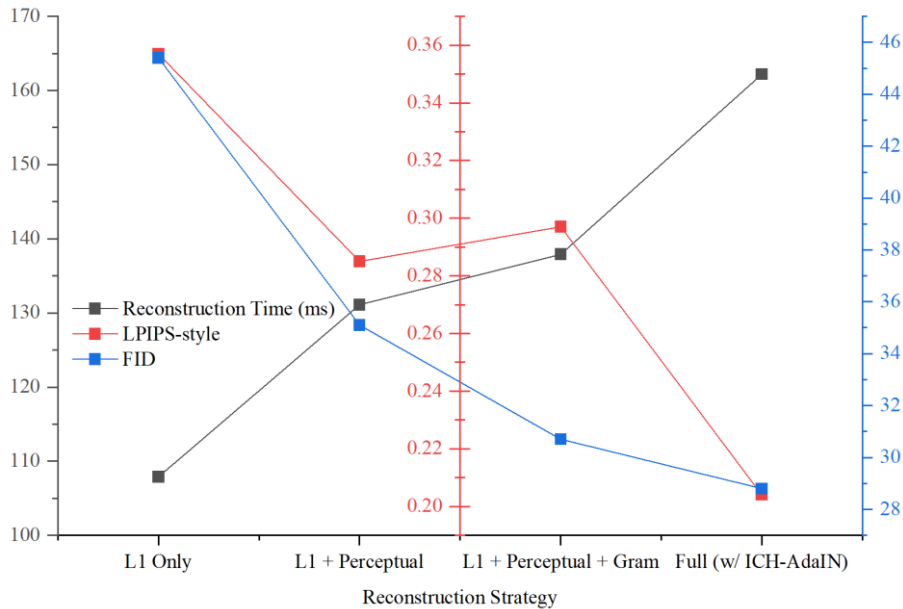


Figure 4: Influence of Depth Reconstruction Strategy on Image Quality

Also, determine how AVEM affects people's sense of sight finally. As shown in Figure 5, after applying AVEM based on the reconstructed output, the local contrast and edge sharpness of key cultural areas have been significantly enhanced; at the same time, the background curtain or blank areas remain soft and there is no excessive sharpening noise. Quantitatively, the local

SSIM of these areas rose from 0.82 to 0.91, and the standard deviation of noise in the background areas dropped by 12 per cent. This kind of addition is not a general enhancement but rather a "specific focus on the cultural part" to help the audience get to know the story in the picture more easily.

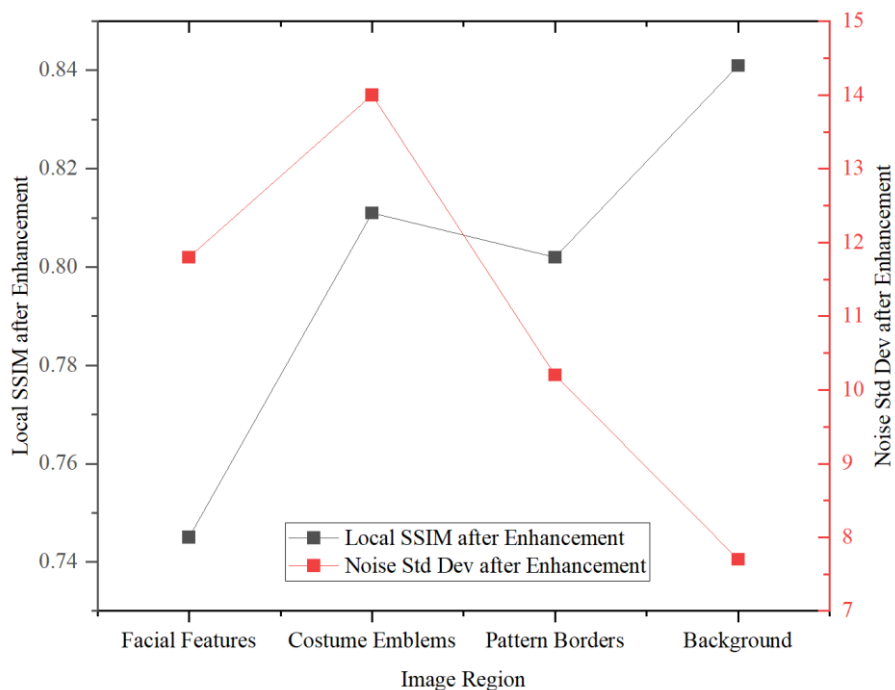


Figure 5: Influence of adaptive enhancement link

Finally, the whole of the model's advantages has been verified by ablation studies. Figure 6 shows the reconstruction results using only the standard U-Net: colour imbalance, broken lines and simplified patterns, and the general appearance is "technical but uncultural". Figure 7 is the output of the entire model in this paper. It has good colour coordination, is line-oriented and fully realised patterns, thus meeting the aesthetic expectations of traditional crafts. According to the user survey, eight ICH researchers gave the "style credibility" of the full model a score of 4.3/5 on average and, compared with the 2.8/5 for the baseline model, it was significantly higher.



Figure 6: ICH Modeling Effect Before Optimisation



Figure 7: ICH Modeling Effect After Optimisation

Experiments have verified the technical feasibility of all modules and shown that only by integrating style decoupling, cultural attention and a multi-scale enhancement mechanism organically can the digital reconstruction of ICH images be realised. It provides a solid basis for the following popularisation and application in an expanded ICH database.

6 Discussion

The depth reconstruction and adaptive enhancement model proposed in this paper has shown good results for the digital reproduction of ICH images, but its scope of application and cultural suitability still need to be optimized. Currently, the methods use high-quality reference images for style embedding extraction. Therefore, when there is only a small, local part of a severely damaged ICH image, the style reconstruction may be inaccurate due to a lack of full context. In the future, we can explore a modeling method based on weak supervision that integrates ICH knowledge maps and text descriptions; thus, the reconstruction results will be closer to the traditional style visually and more accurately reflect the technological specifications and cultural characteristics.

AVEM module uses weak supervised attention guidance, and its performance is constrained by the labeling coverage of the key areas in the training set. Currently, most of the general visual focus, such as faces and central patterns, are used as proxy tags in the model, but the visual focus of ICH is highly context-dependent. If the model fails to learn this difference in culture, the improvement may be incorrect. Therefore, to improve the cultural sensitivity of attention further, richer cultural semantic annotations or cross-modal alignments are being applied.

Technology intervention in ICH protection should be mindful of the risk of "digitalisation of standardisation". The model learns style commonness through a Gram matrix, and thus strengthens the "typical style" of a certain type of ICH while disregarding regional differences or the creative variations among individual inheritors. Future research should continue to explore the space for style variation in model design, and technology should be used as a supportive tool for the work of inheritors, rather than determining the aesthetics of that work. This paper is limited to static pictures and does not cover dynamic ICH forms. Among the above types, the living essence of action rhythm, light and shadow variations, and audience interaction cannot be conveyed in a single-frame image. In the future, we can explore the combination of video generation models and sequential attention mechanisms to extend this method for digital recording and reproduction of dynamic visual intangible cultural heritage.

7 Conclusions

Given the problem that it is difficult to have both "form" and "connotation" in the digital reproduction of ICH artistic images, this study builds an end-to-end model that combines deep style reconstruction and adaptive visual feature enhancement. This paper is not about general image generation but focuses on the properties of ICH ontology to make technology serve the expression of culture through style decoupling and cultural sensitivity. According to the experiment results, only the reconstruction of structure cannot express the aesthetic essence of ICH. After introducing the ICH-AdaIN modulation mechanism, the style consistency index LPIPS-style has been improved significantly; and with the addition of the AVEM module, the local SSIM of key areas such as the face of the idol and the coat of arms has increased from 0.82 to 0.91, and background noise has been suppressed.

There are still some deficiencies in the current research: model training uses paired image data, its generalization ability for ICH categories with few samples is still weak, and it cannot cover dynamic ICH projects. In the future, a weak supervised learning method can be developed by combining the ICH knowledge map, and the reconstruction task can be extended to video sequences. Technology should always be employed in line with the goals of preserving and promoting culture, and accurately and reliably support the development of ICH.

Acknowledgements

This work was supported by the 2025 Provincial Philosophy and Social Science Research General Project: Research on the Symbolic Narration and Innovative Design of Multi-ethnic Dance Costumes in the Wuling Mountain Area (No. 25Y133).

Author details

Lili Zhang was born in Chengdu, Sichuan Province, China, in 1984. Doctor of Philosophy, lecturer. She graduated from De La Salle University-Dasmariñas. She currently works at the College of Arts of Jingchu University of Technology. Her research interests include the visual expression and dynamic inheritance of intangible cultural heritage, and aesthetic education, etc.

Yu Fang was born in Jingmen City, Hubei Province, China in 1984. She obtained her doctorate from De La Salle-Dasmariñas. Currently, she teaches at the College of Arts of Jingchu University of Technology, specializing in Visual Communication Design. Her research focuses on traditional culture and digital illustration.

References

- [1] Alkishriwo, O. A. S. (2020). Image compression using adaptive multiresolution image decomposition algorithm. *IET Image Processing*, 14(14), 3572-3578. <https://doi.org/10.1049/iet-ipr.2019.1699>
- [2] Cao, J., Wang, H., & Zhang, W. (2025). An open-set image classification algorithm based on hardness-aware angular margin loss and label noise filtering. *Applied Intelligence*, 55(12), 857. <https://doi.org/10.1007/s10489-025-06718-5>
- [3] Cao, J. F., Zhang, Z. B., & Zhao, A. D. (2021). Super-resolution reconstruction for

- enhancing the artistic quality of mural images. *Computer Engineering and Design*, 42(08), 2291-2298. <https://doi.org/10.16208/j.issn1000-7024.2021.08.027>
- [4] Chigurupati Navyasree, D. D. V. R. (2020). Combined Robust Algorithm for Image Retrieval of Minority Inheritance. *Solid-State Technology*, 63(5), 4630-4643.
- [5] Ding, H., Zhang, H., & Fu, G. (2024). Towards high-quality photorealistic image style transfer. *IEEE Transactions on Multimedia*, 26, 9892-9905. <https://doi.org/10.1109/TMM.2024.3394733>
- [6] Dong, Y., Fan, S., & Zheng, L. (2025). AttenStyler: Text-image style transfer based on attention mechanism. *Neurocomputing*, 638, 130075. <https://doi.org/10.1016/j.neucom.2025.130075>
- [7] El-Hoseny, H. M., Farahat, M. A., & El-Hag, N. A. (2024). An efficient Stego-OptDehaz algorithm for image dehazing and metadata concealment. *Journal of Optics*, 53(3), 2441-2451. <https://doi.org/10.1007/s12596-023-01364-x>
- [8] Gao, X., & Zhang, Y. (2025). SRAGAN: Saliency regularized and attended generative adversarial network for Chinese ink-wash painting style transfer. *Pattern Recognition*, 162, 111344. <https://doi.org/10.1016/j.patcog.2025.111344>
- [9] Iseringhausen, J., Weinmann, M., & Huang, W. (2020). Computational parquetry: Fabricated style transfer with wood pixels. *ACM Transactions on Graphics (TOG)*, 39(2), 1-14. <https://doi.org/10.1145/3378541>
- [10] Liang, Y., Huang, H., & Cai, Z. (2020). PSO-ACSC: A large-scale evolutionary algorithm for image matting. *Frontiers of Computer Science*, 14(6), 146321.
- [11] Li, Q. S., Ni, T., & Luo, X. (2025). Research progress on digitalization of cultural heritage and visualization service technology for digital cultural tourism. *Journal of Image and Graphics*, 30(06), 2304-2324.
- [12] Li, Z., Lv, H., & Cheng, L. (2025). Image deblocking algorithm based on GC and SSR. *The Visual Computer*, 41(1), 53-66. <https://doi.org/10.1007/s00371-024-03309-2>
- [13] Paul, A. (2023). Adaptive tri-plateau limit tri-histogram equalization algorithm for digital image enhancement. *The Visual Computer*, 39(1), 297-318. <https://doi.org/10.1007/s00371-021-02330-z>
- [14] Song, B. (2020). Optimization of the progressive image mosaicing algorithm in fine art image fusion for virtual reality. *IEEE Access*, 9, 69559-69572. <https://doi.org/10.1109/ACCESS.2020.3022484>
- [15] Su, Q., Zhang, X., & Wang, H. (2022). A blind color image watermarking algorithm combined spatial domain and SVD. *International Journal of Intelligent Systems*, 37(8), 4747-4771. <https://doi.org/10.1002/int.22738>
- [16] Tan, G., & Liu, Y. (2023). Picture processing optimization technology based on Mask R-CNN algorithm. *Procedia Computer Science*, 228, 647-654. <https://doi.org/10.1016/>

[j.procs.2023.11.075](#)

- [17] Wang, C. X., Xu, L. L., & Yu, T. X. (2023). A multi-label crowdsourcing answer aggregation method for digital cultural heritage based on adaptive graph regularization and joint low-rank matrix factorization. *Application Research of Computers*, 40(04), 1119-1129. <https://doi.org/10.19734/j.issn.1001-3695.2022.09.0442>
- [18] Wang, Z., Zhao, L., & Lin, S. (2020). GLStyleNet: Exquisite style transfer combining global and local pyramid features. *IET Computer Vision*, 14(8), 575-586. <https://doi.org/10.1049/iet-cvi.2019.0844>
- [19] Xia, H., Meng, X., & Liu, Y. (2025). Non-parallel story author-style transfer with disentangled representation learning. *ACM Transactions on Knowledge Discovery from Data*, 19(4), 1-26. <https://doi.org/10.1145/3726870>
- [20] Xiang, S., Fu, Y., & Liu, T. (2020). Progressive learning with style transfer for distant domain adaptation. *IET Image Processing*, 14(14), 3527-3535. <https://doi.org/10.1049/iet-ipr.2020.0166>
- [21] Xiao, J., Lyu, S. W., & Huang, Y. D. (2024). Multimodal measurement and conservation-renewal application of heritage value equivalents in historical districts. *Planners*, 40(12), 48-57. <https://doi.org/10.3969/j.issn.1006-0022.2024.12.007>
- [22] Zhang, J. R. (2024). Experimental analysis of style transfer and target detection in interactive art on smartphones. *IETE Journal of Research*, 70(8), 6920-6931. <https://doi.org/10.1080/03772063.2024.2315190>
- [23] Zhang, Z. F., Lu, S., & Guo, Q. (2025). CaVIT: An integrated method for image style transfer using parallel CNN and vision transformer. *Applied Intelligence*, 55(5), 306. <https://doi.org/10.1007/s10489-024-06114-5>
- [24] Zhao, S. (2025). Digital media art style conversion algorithm based on deep neural network. *Procedia Computer Science*, 261, 906-911. <https://doi.org/10.1016/j.procs.2025.04.420>
- [25] Zhou, Y., Ren, Y., & Wu, C. (2024). Style transfer of Chinese Wuhu iron paintings using hierarchical visual transformer. *Sensors*, 24(24), 8103. <https://doi.org/10.3390/s24248103>
- [26] Zuo, J., Zhao, W., & Chen, L. (2022). Multi-focus image fusion algorithm based on random features embedding and ensemble learning. *Optics Express*, 30(5), 8234-8247. <https://doi.org/10.1364/OE.452081>