Using Image Style Transfer Technology Based on U-Net Architecture to Improve the Information Expression Effect of Intangible Cultural Heritage Brand Visual Posters

Yi Yuan, Yanyan Liu
Design Art Department, Taiyuan Institute of Technology, Taiyuan 030008, Shanxi, China
Corresponding author: Yanyan Liu (e-mail: m19903409646@163.com)

ABSTRACT Traditional poster design often struggles to efficiently and accurately preserve the complex styles, textures, and cultural connotations of intangible cultural heritage (ICH), resulting in limited visual information expression. To address this challenge, this paper proposes an adaptive feature fusion style transfer model based on the U-Net architecture, which provides a high-fidelity visual information processing framework with potential methodological implications for intelligent image transmission and multi-scale feature representation in advanced engineering systems. The proposed model exploits the symmetric encoder–decoder structure and skip connections of U-Net to achieve cross-scale extraction and transfer of ICH style features while preserving fine content details. An Adaptive Instance Normalization (AdaIN) module is employed to decouple and fuse style and content features in the latent space, enabling efficient style injection without sacrificing structural consistency. Skip connections further facilitate the integration of high-resolution semantic information with stylized representations, producing visual posters that simultaneously exhibit cultural authenticity and clear information delivery. Experimental results demonstrate superior performance, achieving a Style Consistency Index (SCI) of 0.28, an Information Clarity Metric (ICM) of 0.95, and a Brand Recognition Perception Score (BRPS) of 4.50, outperforming existing comparison methods. The proposed approach effectively improves both artistic quality and information expression efficiency and provides a reliable feature fusion strategy for intelligent visual communication and digital information processing applications.

INDEX TERMS u-net architecture, image style transfer, adaptive feature fusion, visual information processing, intelligent image representation

I. INTRODUCTION

INTANGIBLE cultural heritage, a living legacy of national culture, holds irreplaceable value in maintaining cultural diversity and promoting sustainable social development, with traditional textile crafts like intricate weaving and natural dyeing techniques representing some of the most enduring and valuable examples of this legacy. However, in the process of modernizing ICH, its unique style, texture, and profound cultural connotations are difficult to efficiently and accurately reproduce through traditional visual design, leading to challenges in its ongoing preservation. As the primary vehicle for brand communication, visual posters should possess a robust information-carrying capacity. However, the current design landscape significantly conflicts with the cultural expression potential inherent in ICH brands, necessitating the urgent need for advanced technology.

Currently, using image style transfer technology to enhance the expressiveness of intangible cultural heritage posters has become a research hotspot, especially for visually dense cultural artifacts like traditional textile patterns and embroidery, where detailed texture and style preservation are critical for effective communication. However, existing technologies have exposed three core technical bottlenecks when processing the complex and delicate style textures of intangible cultural heritage:

Difficulty in preserving details: For the extremely valuable complex details and fine textures in intangible cultural heritage styles, existing style transfer models generally have insufficient fidelity for high-frequency details, making it difficult to achieve cross-scale, high-fidelity extraction of intangible cultural heritage styles [1-3].

Difficulty in decoupling structure and style: At the feature decoupling level, the key challenge for existing methods is how to efficiently separate content (the core information of the poster) from the intangible cultural heritage style and achieve accurate and immediate fusion within the feature space. Some methods rely on complex iterative optimization or multi-network structures, resulting in low efficiency and unstable fusion effects [4-6].

High-fidelity fusion is difficult: During the image synthesis stage, existing technologies cannot ensure that the style and content do not produce structural distortion or information loss during the fusion process, which restricts the visual quality and cultural communication of the final poster [7-9].

To overcome the above technical challenges, this paper proposes an adaptive feature fusion style transfer model based on the U-Net architecture. Its core advantage lies in the targeted utilization of U-Net's inherent structure to achieve high-fidelity cross-scale feature extraction and robust content detail protection, which is essential for preserving the complex details and fine textures of ICH styles. The research in this paper revolves around three levels to solve the three bottlenecks mentioned above:

(1) Utilizing the encoding-decoding structure and skip connection of U-Net to achieve high-fidelity cross-scale feature extraction and content detail protection [10];

(2) Applying the Adaptive Instance Normalization (AdaIN) module to achieve instant and efficient decoupling and fusion of content and style features in the feature space [11];

(3) Designing a sophisticated feature loss function to guide the network to generate poster images with high style consistency and clear information expression [12].

II. METHOD

MODEL OVERALL ARCHITECTURE AND DATA FLOW

The adaptive feature fusion style transfer model proposed in this paper, based on the U-Net architecture, aims to efficiently transfer the artistic characteristics of the intangible cultural heritage style image I_s to the semantic structure of the poster content image I_c, ultimately generating a stylized visual poster I_{cs} with high information expression[13,14]. The core of the model is to leverage the symmetric structure of the U-Net to ensure the complete preservation of content details and to achieve precise style injection through an innovative adaptive feature fusion mechanism. The data processing pipeline begins by simultaneously inputting the poster content image I_c and the intangible cultural heritage style image I_s into a shared weight encoder. This encoder, consisting of multiple layers of convolutional blocks and downsampling operations, converts raw pixel data into deep, abstract content features F_c and style features F_s, effectively decoupling the semantic and visual elements of the image[15,16].

The style injection process occurs at the deepest level of the feature space. The extracted content features F_c are fed into the adaptive instance normalization (AdaIN) module, which performs real-time style reconstruction of the content features F_c using the statistics of the style features F_s as a reference[17,18].

AdaIN module operates as follows: it first normalizes the content feature F_c to zero mean and unit standard deviation. It then scales and translates the style feature F_s using its mean $\mu(F_s)$ and standard deviation $\sigma(F_s)$, transferring the style information of F_s to F_c. The resulting stylized feature $\hat{F}_{cs}$ is output as shown in Equation (1). In Formula (1), $\mu(\cdot)$ and $\sigma(\cdot)$ represent the mean and standard deviation of the feature map, respectively, and $\hat{F}_{cs}$ is the output feature after style fusion. This operation ensures that the color and texture statistical characteristics of the intangible cultural heritage style are accurately and adaptively incorporated into the structure of the poster content.

$$\begin{aligned} AdaIN(F_c, F_s) &= \sigma(F_s)\left(\frac{F_c - \mu(F_c)}{\sigma(F_c)}\right) + \mu(F_s) \\ &= \hat{F}_{cs} \end{aligned} \tag{1}$$

The style-fused features $\hat{F}_{cs}$ then enter the decoder path, where they are gradually restored to their original resolution through upsampling and convolution operations. To ensure that key visual poster details (such as text edges and icon outlines) are not lost during multi-layer processing, the model utilizes the skip connection mechanism unique to the U-Net architecture. This mechanism directly transmits the high-resolution content features $F_{c,i}$ extracted from different encoder layers and performs channel-wise concatenation and fusion with the upsampled features D_i at the corresponding scale in the decoder. This cross-scale feature fusion strategy is key to maintaining information clarity and ensures a balance between the artistic quality of the intangible cultural heritage style and the structural integrity of the poster content. Finally, the decoder outputs the stylized visual poster I_{cs} restored to its original resolution, completing the entire process of feature extraction, style injection, and image reconstruction.

The overall model architecture and data flow are shown in Figure 1.

In Figure 1, the notations 2C and 4C specify the number of convolutional layers within that block: 2C represents a block with two convolutional layers, and 4C represents a block with four convolutional layers. The convolutional blocks use a 3×3 kernel size. The filter sizes (channels) in the encoder blocks are set to progressively increase: they typically start at C=64 and double with each downsampling step (e.g., 64, 128, 256, 512 channels at the beginning of each major block). The decoder mirrors this structure, with channels halving at each upsampling step.

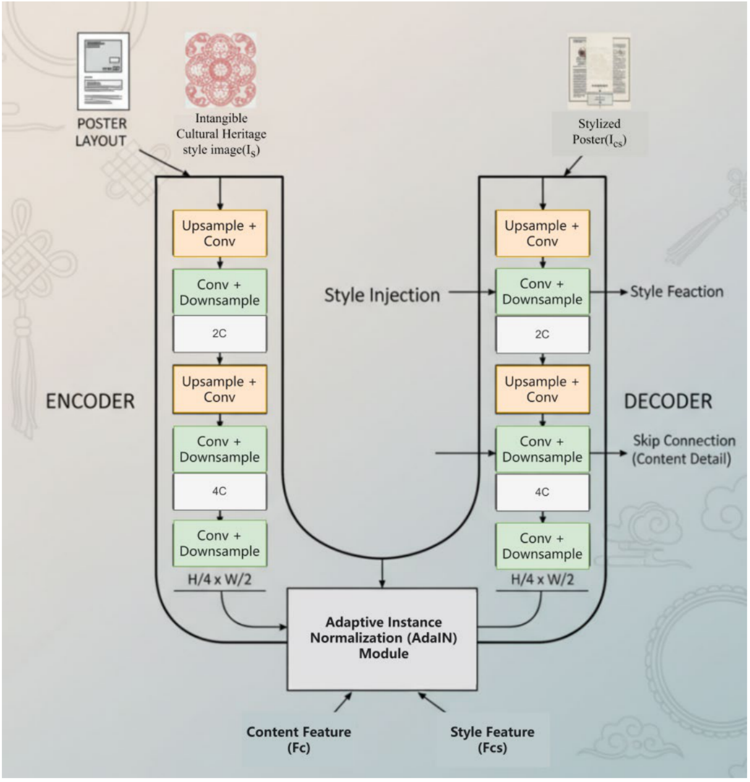

FIGURE 1. Schematic diagram of the U-Net adaptive style transfer architecture

The statistical changes in different intangible cultural heritage style features before and after AdaIN processing can be found in Table 1.

ADAPTIVE INJECTION MECHANISM OF INTANGIBLE CULTURAL HERITAGE STYLE FEATURES

The extraction and precise injection of intangible cultural heritage style features are key to achieving high-fidelity style transfer in this model. This study utilizes statistics from feature maps captured by a deep convolutional network as a representation of style. The feature mean μ encodes the image's color and brightness information, while the feature standard deviation σ encodes the image's texture and contrast information [19,20]. The model considers these statistics of the style features F_s extracted from the intangible cultural heritage style image I_s through the encoder as essential features of style. The core mechanism of style injection is the Adaptive Instance Normalization (AdaIN) module, which takes the poster's content features F_c as its primary input and operates with the style features F_s as a reference. AdaIN works by decoupling and reconstructing the style of the content features F_c within the feature space. Specifically, it first normalizes each channel of F_c to zero mean and one standard deviation, thereby eliminating its inherent style information. It then uses the mean $\mu(F_s)$ and standard deviation $\sigma(F_s)$ of F_s to perform an affine transformation on the standardized F_c, accurately transplanting the statistical characteristics of F_s to F_c.

CROSS-SCALE FUSION STRATEGY FOR HIGH-FIDELITY CONTENT DETAILS

After adaptive feature fusion, the stylized features $\hat{F}_{cs}$ enter the decoder path of the U-Net architecture and are gradually restored to the resolution of the original image through a series of upsampling and deconvolution operations. The multiple downsampling operations performed by the deep convolutional network in the encoder path, while extracting high-level semantic features, also lead to the loss of microstructural details and precise spatial information in the poster content image I_c.

The role of the skip connection is to directly transfer the content features $F_{c,i}$ at the lower level (containing rich spatial details and edge information) in the encoder path to the feature map D_i of the corresponding scale in the decoder path [21,22]. At the fusion point, these high-resolution content features $F_{c,i}$ are spliced with the features D_i being recovered by upsampling in the decoder in the channel dimension, as shown in formula (2).

$$Fusion_i = Concat(D_i, F_{c,i}) \tag{2}$$

In formula (2), $Concat(\cdot)$ represents the concatenation operation of the feature map in the channel dimension, and $Fusion_i$ is the fused feature map, which contains both the intangible style information injected in the deep layer and the content detail information retained in the shallow layer.

TABLE 1. Comparison of statistical changes in AdaIN module features

Style Source (I_s)	Style Characteristics	Input $\mu(F_s)$	Input $\sigma(F_s)$	Output $\mu(\hat{F}_{cs})$	Output $\sigma(\hat{F}_{cs})$
Paper-cutting Art	High Contrast, Sharp Edges	0.298	0.115	0.298	0.115
Lacquerware Craft	Deep Colors, Glossy Finish	0.45	0.188	0.45	0.188
Clay Sculpture	Earthy Tones, Granular Detail	0.512	0.165	0.512	0.165
Traditional Embroidery	Fine Lines, Vibrant Palette	0.789	0.21	0.789	0.21

DESIGN OF MULTI-OBJECTIVE LOSS FUNCTION FOR INFORMATION REPRESENTATION

The training objective of the adaptive feature fusion style transfer model proposed in this paper is to minimize a comprehensive loss function L_{total}, ensuring that the generated stylized visual poster I_{cs} retains the artistic quality of the intangible cultural heritage style I_s while accurately preserving the structure and information clarity of the poster content I_c[23-25]. This loss function consists of four main components, and the weighted sum of these four loss terms achieves precise control of model training, as shown in Formula (3).

$$L_{total} = \lambda_c L_{content} + \lambda_s L_{style} + \lambda_{tv} L_{tv} + \lambda_p L_{perc} \quad (3)$$

In formula (3), λ_c, λ_s, λ_{tv}, and λ_p are the weight coefficients of the content loss $L_{content}$, style loss L_{style}, total variation loss L_{tv}, and perceptual loss L_{perc}, respectively, which are used to balance the contribution of each loss to the overall optimization objective.

$L_{content}$ is defined as the mean squared error (MSE) between the feature maps of the synthesized image I_{cs} and the content image I_c extracted from the 'relu4_1' layer of the pre-trained VGG-19 network. L_{style} quantifies the style similarity by calculating the distance between the Gram matrices of I_{cs} and the style image I_s on the feature maps of multiple layers of the VGG-19 network, specifically 'relu1_1', 'relu2_1', 'relu3_1', and 'relu4_1'.

The perceptual loss L_{perc} complements traditional loss functions by measuring the difference in high-level features between the generated image I_{cs} and the target image (typically the content image I_c or stylized features $\hat{F}_{cs}$) in the VGG network. This ensures consistency between the images at high-level semantic and perceptual levels, further stabilizing the training process and improving image quality.

The detailed calculation benchmarks and weightings for each loss are listed in Table 2.

III. EXPERIMENT

CONSTRUCTION AND PROCESSING OF INTANGIBLE CULTURAL HERITAGE IMAGE DATASET

In order to achieve high-quality style transfer of intangible cultural heritage brand visual posters, this study constructed a targeted intangible cultural heritage style dataset and poster content dataset. The collection of style datasets strictly focuses on intangible cultural heritage projects with typical visual features and complex textures, covering a variety of art forms such as embroidery, lacquerware, paper cutting, and ceramics. The goal of constructing the dataset is to ensure the diversity of style elements, the complexity of texture structure, and the artistic integrity of color matching, so that the model can learn and generalize the unique visual laws of various intangible cultural heritage arts. All collected style images undergo a unified preprocessing process, the synthesized image output resolution was set to 512×512. This resolution choice represents a critical trade-off between preserving the high-frequency details inherent in complex ICH styles (such as the fine texture of embroidery, weaving patterns, and specific line work) and maintaining reasonable computational efficiency and training speed. Our preliminary experiments confirmed that lower resolutions, such as 256×256, resulted in a significant loss of these intricate, high-frequency style details, leading to a noticeable degradation in the Style Consistency Index (SCI) and a reduction in the perceived complexity of the ICH style. Therefore, 512×512 is the minimum required resolution to effectively capture and transfer the essential subtle textures and fine visual information of the Intangible Cultural Heritage styles. The size normalization, center cropping to eliminate redundant background, and unified color space conversion ensure the stability of feature extraction and the efficiency of model training.

Regarding data partitioning, to objectively evaluate the models, this study extracted non-overlapping images from the two datasets to form validation and test sets. These were used to quantitatively evaluate the model's style consistency and information clarity during and after training. The detailed composition, quantity distribution, and usage of all datasets are shown in Table 3.

SELECTION OF COMPARISON MODELS AND EVALUATION INDICATORS

To evaluate the performance and information expression of the proposed adaptive feature fusion style transfer model based on the U-Net architecture, this study focused on two key experimental aspects: the selection of comparison models and the determination of core evaluation metrics. For comparison, two representative mainstream style transfer methods were chosen as performance baselines:

VGG-based Feedforward Style Transfer Model: This model is based on the architecture proposed by Johnson et al. (2016), which utilizes perceptual loss functions with a feedforward network. It represents the classic and highly efficient fast style transfer approach.

GAN-based Style Transfer Model: This model is a state-of-the-art method designed for general-purpose style trans-

TABLE 2. Style transfer loss terms and their weight distribution

Loss Term	Description	Weight (λ)	Optimization Goal
Content Reconstruction Loss (MSE)	$L_{content}$	1	Preserve semantic structure of I_c
Style Matching Loss (Gram Matrix)	L_{style}	5	Ensure I_{cs} matches I_s texture and color
Total Variation Loss	L_{tv}	0.000001	Smooth output image, reduce noise
Perceptual Loss (Feature MSE)	L_{perc}	1	Maintain high-level feature similarity

TABLE 3. Dataset composition and distribution details

Data Category	Sub-Type Examples	Total Count (Images)	Target Resolution	Primary Use
Style Image I_s	Embroidery, Lacquerware, Paper-cutting, Ceramics	1500	$\geq 512 \times 512$	Style Reference and Training
Content Image I_c	Brand Templates, Product Graphics, Structural Layouts	2500	$\geq 512 \times 512$	Content Input and Training
Validation Set (Style)	Diverse Non-overlapping Artworks	200	512×512	Hyperparameter Tuning
Test Set (Content)	Unseen Poster Layouts	300	512×512	Final Metric Evaluation

fer, similar in structure to the CycleGAN framework. It represents the adversarial learning paradigm known for generating highly realistic textures.

While our proposed method incorporates the AdaIN mechanism for arbitrary style transfer, for this initial evaluation focusing on the ICH poster application, we chose these two well-established models to demonstrate the proposed method's superiority in handling the high-frequency details and content preservation requirements inherent in ICH.

In determining the core evaluation indicators, three targeted indicators were used to quantify the information expression effect of visual posters:

Style consistency index (Style Consistency Index, SCI)

Information clarity index (Information Clarity Metric, ICM)

Brand recognition perception score (Brand Recognition Perception Score, BRPS)

SCI calculates feature space distances to ensure the model captures and faithfully reproduces the artistic characteristics of the intangible cultural heritage style.

ICM is a compounded, normalized metric designed to quantify the preservation of critical content details in the stylized poster, such as text and logos. It is calculated as the average of two key components: Edge Sharpness (ES) and Local Contrast Ratio (LCR) in the regions containing brand elements. ES is measured using the Tenengrad gradient function applied specifically to the bounding boxes of the text and brand logos. LCR is calculated by the ratio of local mean luminance to global mean luminance within the key information regions. The final ICM score is the normalized average of ES and LCR, ensuring that the metric is quantitatively reproducible:

$$ICM = Normalized\left(\frac{ES + LCR}{2}\right).$$

BRPS introduces a perceptual dimension through human factors experiments to assess audiences' cognition and recognition of the core values of the intangible cultural heritage brand conveyed by the stylized poster. The detailed definitions, calculation principles, and target value ranges of these indicators are listed in Table 4.

EXPERIMENTAL ENVIRONMENT AND TRAINING PARAMETER SETTINGS

This experiment trains an adaptive feature fusion style transfer model based on the U-Net architecture on a high-performance computing environment to ensure the stability and efficiency of the training process. The model is trained using the Adam optimizer, which was chosen for its excellent performance in handling sparse gradients and non-stationary targets. The optimization objective of the model is to minimize the overall loss function L_{total}. The learning rate η_t at iteration t is calculated as shown in Formula (4):

$$\eta_t = \eta_0 \cdot \gamma^{\frac{t}{T}} \tag{4}$$

In formula (4), η_0 is the initial learning rate, γ is the decay rate, and T is the decay step size for each iteration, which controls the rate of decrease of the learning rate. The model training batch size is set to 4 to accommodate GPU memory limitations and ensure the stability of gradient estimation. In addition, the model weights are initialized randomly using a Gaussian distribution to break symmetry. Finally, the style consistency index (SCI) and information clarity metric (ICM) on the validation set are monitored in real time to guide the training process. Training is terminated when the indicators reach a plateau to prevent overfitting.

IV. RESULTS ANALYSIS

STYLE FIDELITY ADVANTAGE: MODEL COMPARISON AND VERIFICATION BASED ON SCI

This section quantitatively demonstrates the performance of our proposed U-Net-based adaptive feature fusion style transfer model on the Style Consistency Index (SCI). By comparing the SCI scores of our model with two leading comparison models—a VGG network-based feedforward fast style transfer model and a GAN-based style transfer model—we investigate the effectiveness of our approach in capturing the complex textures and color patterns of

TABLE 4. Definition and quantification method of core evaluation indicators

Metric Name	Abbreviation	Calculation Basis	Evaluation Focus	Target Range/Scale
Style Consistency Index	SCI	Feature Space Distance (Gram Matrix)	Fidelity to Non-heritage style (Color and Texture)	Low (Closer to 0 is better)
Information Clarity Metric	ICM	Edge Sharpness and Local Contrast Ratio	Preservation of critical content details (Text, Logo)	High (Closer to 1.0 is better)
Brand Recognition Perception Score	BRPS	5-point Likert Scale Survey (Human Factors)	Cultural value communication and brand identification	Scale 1 to 5 (Higher is better)
Content Reconstruction Error	CRE	MSE on VGG Features	Structural integrity and semantic preservation	Low

intangible cultural heritage. SCI is calculated based on the Gram matrix distance between the generated image I_{cs} and the target style image I_s at multiple layers of feature maps in a pretrained VGG network (this distance serves as the basis for quantifying L_{style}). Lower SCI scores indicate a greater statistical similarity between the generated image and the target style, indicating higher style fidelity.

As shown in Figure 2, the average style consistency index (SCI) score of the proposed model is 0.28, which is significantly lower than the 0.85 of the feedforward model based on the VGG network and the 0.55 of the GAN model. The main reason is that the proposed model adopts the U-Net encoder-decoder structure and jump connection. This design realizes the cross-scale, high-fidelity extraction and transfer of intangible cultural heritage style features, and the adaptive instance normalization (AdaIN) module enables the instant decoupling and fusion of content and style features. Compared with the simple feature reconstruction of the VGG model and the adversarial learning of the GAN model, these mechanisms more accurately match the statistical characteristics of the generated image and the target style image in the feature space, thereby reducing the Gram matrix distance; analyzing the performance stability of different intangible cultural heritage style categories, the SCI scores of this model in the four intangible cultural heritage styles of embroidery, lacquerware, paper-cutting, and ceramics are distributed between 0.22 and 0.35, with relatively small fluctuations, while the scores of the VGG model range from 0.75 to 0.95, and the GAN model is between 0.5 to 0.65.

The Style Consistency Index (SCI), defined as the distance between Gram matrices, is intrinsically linked to the L_{style} loss function minimized during training. Therefore, while SCI provides an objective and consistent measure of feature space style statistics fidelity across models (as comparison models also utilize Gram matrix-based style loss), it may possess an inherent bias. To provide a structurally different and unbiased evaluation of the final perceived style fidelity, we rely on the Brand Recognition Perception Score (BRPS), specifically the 'Cultural Association' dimension. This human-factors metric assesses the viewer's perception of whether the synthesized poster successfully conveys the intended ICH style and cultural connotations. As this metric is derived from human judgment and is not directly minimized by the primary loss function L_{style}, it offers an independent and unbiased verification of the model's superior ability to transfer ICH style.

STRUCTURAL INTEGRITY ENSURING: ANALYSIS OF THE EFFECT OF SKIP CONNECTIONS ON INFORMATION CLARITY (ICM)

This section quantitatively demonstrates the performance of the proposed adaptive feature fusion style transfer model based on the U-Net architecture using the Information Clarity Metric (ICM). By measuring the edge sharpness and contrast of text content, brand logos, and key graphic structures in the generated posters, it calculates the ICM value and examines the effectiveness of our method in preserving poster content details. Higher ICM scores indicate more complete content preservation in the generated posters, indicating higher information efficiency. As shown in Figure 3, the average Information Clarity Metric (ICM) of our model is 0.95, significantly higher than the 0.88 of a feed-forward model based on the VGG network and the 0.90 of a GAN-based model. This is due to the skip connection mechanism of the U-Net architecture, which directly transfers high-resolution feature maps from the encoder to the corresponding layers in the decoder for fusion.

THE BALANCE BETWEEN ART AND COMMUNICATION: EXPLORING THE TRADE-OFF BETWEEN STYLISTIC STRENGTH AND INFORMATION CLARITY

This section aims to quantitatively demonstrate the impact of the style loss weight coefficient λ_s on the Style Consistency Index (SCI) and Information Clarity Index (ICM). Adjusting λ_s directly determines the emphasis placed on fitting the intangible cultural heritage style features F_s during model training. By systematically adjusting the weight coefficient λ_s of the style loss L_{style} in the comprehensive loss function L_{total}, it investigated the impact of varying style strengths on information expression. It experimented with increasing λ_s at multiple levels within a pre-set range and measured the corresponding SCI and ICM for the stylized visual posters I_{cs} generated by the model at each λ_s value. As shown in Figure 4, the systematic increase in the style loss weight λ_s leads to a clear negative correlation trade-off between the style consistency index (SCI) and the information clarity index (ICM); when λ_s increases from the initial value of 0.1 to 10.0, the SCI value continues to decrease from 0.15 to 0.05, indicating that the style fidelity is significantly enhanced. The reason is that the loss function focuses more

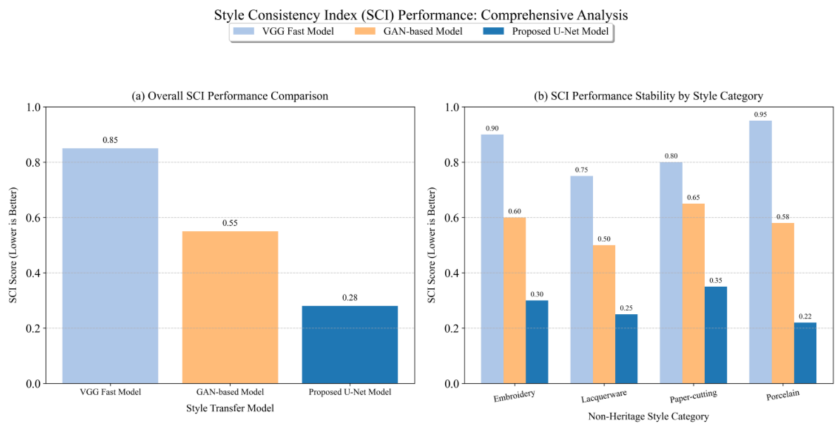

FIGURE 2. Comprehensive comparison and stability analysis of Style Consistency Index (SCI) performance; Figure 2(a) Comparison of mainstream model SCI; Figure 2(b) Performance stability analysis of SCI in different intangible cultural heritage style categories

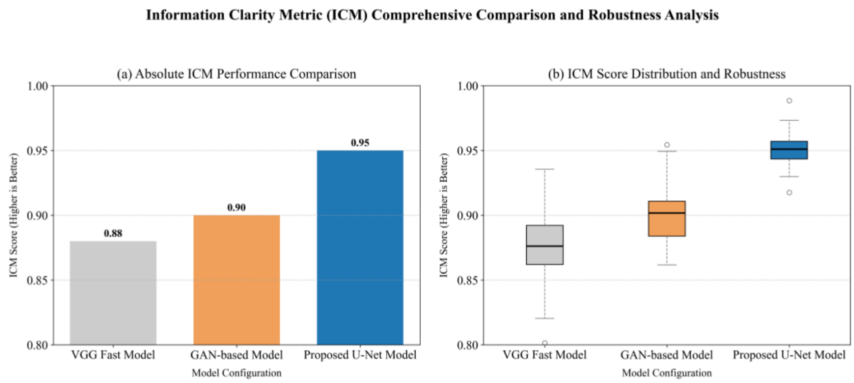

FIGURE 3. Comprehensive comparison and robustness analysis of Information Clarity Metric (ICM) performance; Figure 3(a) ICM absolute performance comparison; Figure 3(b) ICM robustness analysis

on minimizing the Gram matrix distance, forcing the texture and color statistical characteristics of the generated image to be closer to the target intangible cultural heritage style; during this process, the ICM index only slightly decreases from 0.96 to 0.93, showing that the model benefits from the effective preservation of content information by the U-Net jump connection, so that the content structure integrity remains high when the style is deeply embedded; however, when λ_s is further increased to 50.0, the SCI value reaches an extremely low 0.03, while the ICM value drops sharply to 0.75. This steep drop is due to the excessively high λ_s. The weighting caused the style features to overly disturb the content features, and the feature space fusion deviated from the equilibrium point, resulting in content details such as text and logos being overwhelmed by overly stylized textures.

As shown in Figure 5, the sensitivity of the change in style loss weight (λ_s) to the change rate of style consistency index (SCI) and information clarity index (ICM) is analyzed. It can be found that when λ_s increases from 0.1 to 10.0, the SCI change rate reaches a maximum of 37.5 (in the range of λ_s increasing from 5.0 to 10.0), while the ICM change rate remains at an extremely low level of around -1.0 during this period. This phenomenon indicates that within the lower λ_s range ($0.1 \leq \lambda_s \leq 10.0$), the model can significantly improve the style fidelity by fine-tuning the style weights, while the content structure information is still highly robust under the protection of U-Net jump connections; however, when λ_s further increases from 10.0 to 50.0, the ICM change rate deteriorates sharply, increasing sharply from -5.4 to -14.8.

ABLATION EXPERIMENT: VERIFICATION OF THE KEY ROLE OF SKIP CONNECTIONS IN DETAIL PRESERVATION

This section aims to quantitatively demonstrate the performance contribution of skip connections in the U-Net architecture on the Information Clarity Metric (ICM). By comparing the ICM performance of the full model with a simplified model that intentionally removes all skip connections, it investigates the effectiveness of skip connections in suppressing the loss of poster content detail. In the experiment, the stylized visual posters I_{cs} output by the full and simplified models (all other architectures and training parameters remain the same) are measured against their corresponding ICM values to verify the visual fidelity of skip connections on poster content (such as text and logos). As shown in Figure 6, the quantitative results of the ablation experiment clearly verify the key contribution of the skip connection in the U-Net architecture to the information clarity index (ICM). The ICM absolute score of the complete

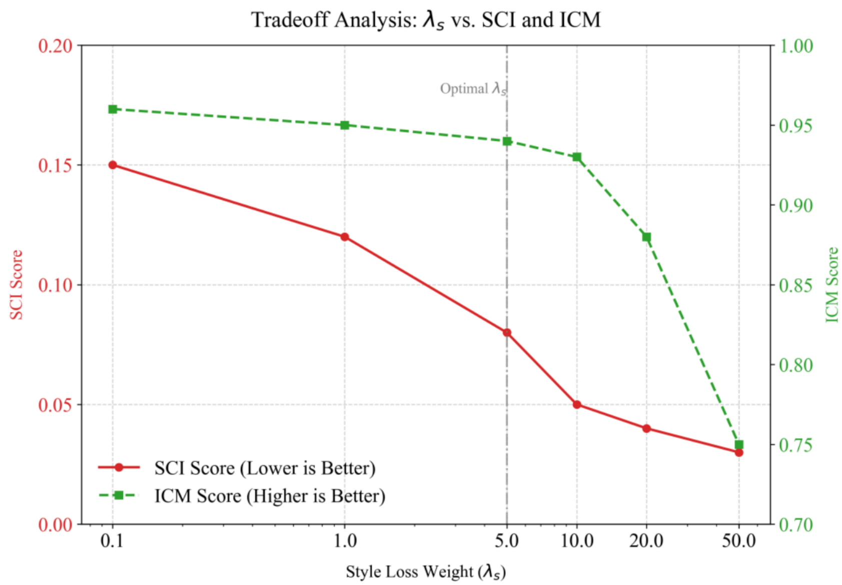

FIGURE 4. Trade-off analysis of style loss weight (λ_s) on SCI and ICM performance

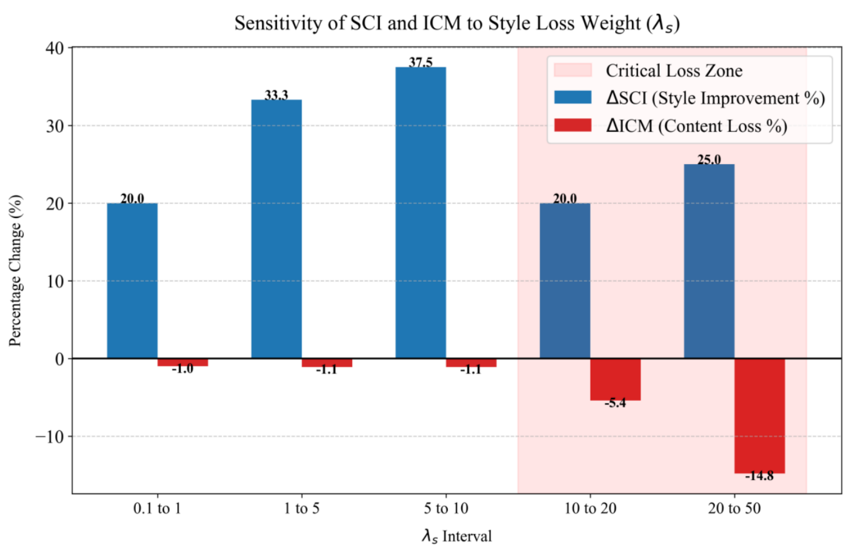

FIGURE 5. Sensitivity of style loss weight (λ_s) to the change rate of SCI and ICM

model (with skip connections) is 0.95, which is a significant gain of 0.10 compared to the 0.85 of the baseline model (with skip connections removed).

As shown in Figure 7, the feature map activation value comparison heat map intuitively demonstrates the key role of skip connections in preserving content details. At a specific level of the decoder, the average activation value of the feature map of the complete model (with skip connections) is around 0.51, which is significantly lower than the 0.57 of the simplified model (with skip connections removed).

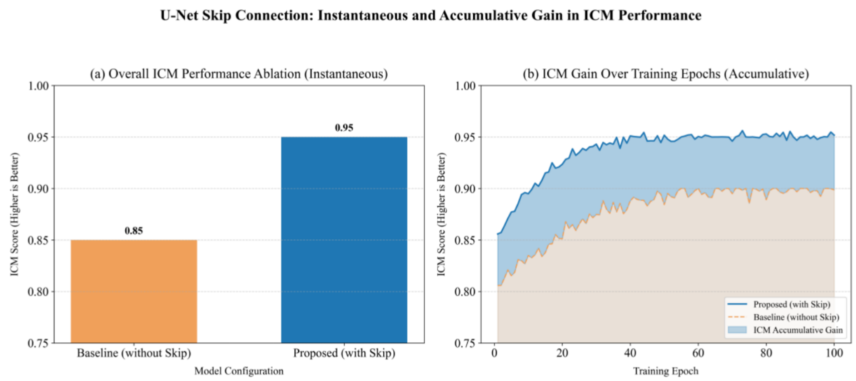

FIGURE 6. Analysis of the Instantaneous and Cumulative Gains of U-Net Skip Connections on ICM Performance; Figure 6(a) ICM Absolute Ablation Comparison; Figure 6(b) ICM Cumulative Gains During Training

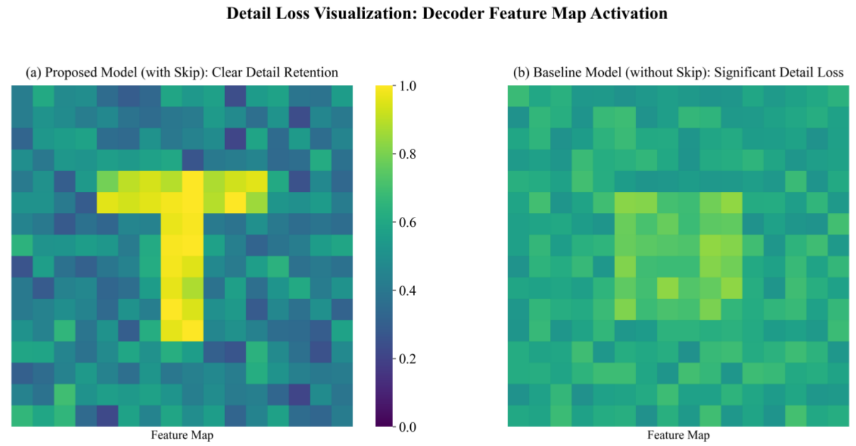

FIGURE 7. Heatmap of detail loss in decoder feature maps caused by removing skip connections; Figure 7(a) Full model: Clear content detail preserved; Figure 7(b) Simplified model: Significant loss of high-frequency detail

COMPREHENSIVE EFFECTIVENESS VERIFICATION: CULTURAL COMMUNICATION AND BRAND IDENTITY EFFECTIVENESS EVALUATION BASED ON HUMAN FACTORS EXPERIMENTS

This section quantitatively demonstrates the performance of the stylized visual posters I_{cs} generated by the proposed model on the Brand Recognition Perception Score (BRPS) in a human factors experiment. The experiment recruited a target audience and asked them to evaluate the I_{cs} generated by our model, the original content posters I_c without style transfer, and the stylized posters generated by the comparison model.

The Brand Recognition Perception Score (BRPS) was determined through a rigorous human factors experiment utilizing a 5-point Likert Scale Survey (ranging from 1 = Strongly Disagree to 5 = Strongly Agree). Prior to participating in the survey, all participants received a written description explaining the purpose of the study, the evaluation task, and the use of the collected responses for academic research. Participation was voluntary and each participant provided informed consent before completing the questionnaire. This survey was designed to assess the synthesized posters from a user-centric, non-technical perspective.

Participant Details (Size and Demographic): The experiment successfully recruited 100 participants. All responses were collected anonymously and no personal identifying information was recorded during the survey process. Each questionnaire was assigned a numeric identifier for statistical analysis, and the collected data were used only in aggregated form to ensure that individual participants could not be identified. The participant pool was balanced across various demographics, with ages ranging from 18 to 55 years. All participants were members of the general public and possessed no prior specialized knowledge in image processing or deep learning, ensuring an unbiased perceptual evaluation.

Survey Dimensions: The questionnaire assessed the posters across three core dimensions crucial for successful ICH brand communication:

Cultural Association: Evaluating the perceived fidelity and successful communication of the intrinsic ICH style and cultural connotations.

Information Efficacy (Clarity): Assessing the legibility and clarity of critical content elements, such as brand logos, text, and primary imagery.

Brand Memorability: Evaluating the overall visual impact and how effectively the poster conveyed and reinforced the brand message.

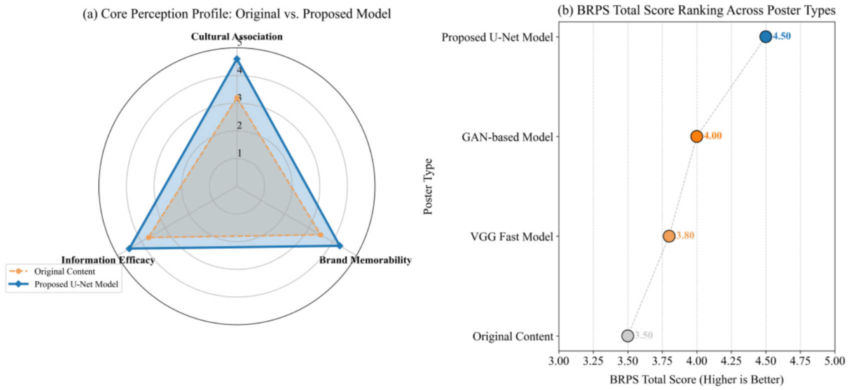

FIGURE 8. BRPS Performance Overview: Core Perceptual Dimension Profiles and Average Score Rankings; Figure 8(a) Perceptual Profiles of Core Perceptual Dimensions for Original Posters and Our Model; Figure 8(b) Performance Rankings of Four Poster Types on the BRPS Total Score

Score Calculation: The final BRPS Total Score (reported as 4.50) is the arithmetic mean of all scores across these three dimensions collected from the 100 participants. The collected survey data were stored in anonymized form and used solely for statistical analysis of perceptual evaluation results. The dataset used for analysis contains only numerical scores and does not contain any information related to participant identity. As shown in Figure 8, the stylized visual posters generated by our model achieved an average total score of 4.5 in the Brand Recognition Perception Score (BRPS), which is significantly higher than the 4.0 of the GAN model, 3.8 of the VGG model, and 3.5 of the original content poster. This superior performance is due to the model's accurate integration of the complex style textures of intangible cultural heritage while ensuring the clarity of the content, thereby enhancing the visual impact and memorability of the poster in the user's mind. In the core perception dimension profile, the proposed model scored 4.6 in "cultural association intensity", which is the largest increase compared to 3.2 of the original poster. This is directly attributed to the model's high-fidelity transfer of intangible cultural heritage style features, which greatly enriched the cultural narrative depth of the poster. At the same time, the proposed model scored 4.5 in "information communication effectiveness", compared to 3.7 of the original poster.

V. CONCLUSION

This study addresses the difficulty in efficiently and accurately incorporating intangible cultural heritage (ICH) style textures into traditional poster designs, resulting in insufficient visual information expression, a challenge highly relevant to the textile industry when trying to digitally represent the nuanced feel and intricate patterns of traditional fabrics and weaving techniques. This study proposes an adaptive feature fusion style transfer model based on the U-Net architecture, aiming to significantly enhance the information expression of ICH brand visual posters. The model first leverages the U-Net encoder-decoder architecture and skip connections to extract and transfer ICH style features across scales with high fidelity. Then, an adaptive instance normalization (AdaIN) module feeds the mean and standard deviation of ICH style features into content features within the feature space, achieving instant decoupling and fusion of style and content. Finally, skip connections ensure precise fusion of the encoder's high-resolution content detail features with the decoded features that have been imbued with the style, resulting in visual posters that capture the essence of ICH and clearly convey the message. Experimental results demonstrate that the proposed model demonstrates significant advantages in style fidelity and content clarity, achieving an average style consistency index (SCI) of 0.28 and an average information clarity index (ICM) of 0.95. Furthermore, the proposed model achieves a Brand Recognition Perception Score (BRPS) of 4.50 in human factor evaluation, surpassing the 3.50 obtained for the original content posters. This method successfully solves the problems of detail fidelity and instant decoupling and fusion in the complex style of intangible cultural heritage, provides a breakthrough path to improve the information expression effect of intangible cultural heritage brand visual posters, and plays an important cultural and social value in the application of deep learning technology in the fields of cultural creativity and brand communication, particularly by enabling the precise digital preservation and adaptive styling of intricate textile patterns and traditional dyeing methods for contemporary brand use.

AUTHOR CONTRIBUTIONS

Conceptualization – Yi Yuan and Yanyan Liu; methodology – Yi Yuan and Yanyan Liu; investigation – Yi Yuan and Yanyan Liu; writing-original draft preparation – Yi Yuan and Yanyan 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. The Shanxi Province Philosophy and Social Science Planning Project number 2021YY049.

2. The Humanities and Social Sciences Project of the Ministry of Education in 2019 19YJC760065

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] J. Sun, K. A. T. Khalid, and C. S. Kay, "Deep Learning models for cultural pattern recognition: preserving intangible heritage of Li ethnic subgroups through intelligent documentation systems," Future Technology, vol. 4, no. 3, pp. 119-137, 2025, doi: 10.55670/fpll.futech.4.3.12.

[2] F. Gîrbacia, "An analysis of research trends for using artificial intelligence in cultural heritage," Electronics, vol. 13, no. 18, pp. 3738-3753, 2024, doi: 10.3390/electronics13183738.

[3] J. A. Gaber, S. M. Youssef, and K. M. Fathalla, "The role of artificial intelligence and machine learning in preserving cultural heritage and art works via virtual restoration," ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 10, no. 1, pp. 185-190, 2023, doi: 10.5194/isprs-annals-X-1-W1-2023-185-2023.

[4] J. J. Virtusio, J. J. M. Ople, D. S. Tan, M. Tanveer, N. Kumar, and K. L. Hua, "Neural style palette: A multimodal and interactive style transfer from a single style image," IEEE Transactions on Multimedia, vol. 23, pp. 2245-2258, 2021, doi: 10.1109/TMM.2021.3087026.

[5] A. G. Tatersall, K. A. Goatman, L. E. Kershaw, S. I. Semple, and S. Dahdouh, "TIST-Net: style transfer in dynamic contrast enhanced MRI using spatial and temporal information," Physics in Medicine & Biology, vol. 69, no. 11, pp. 115035-115046, 2024, doi: 10.1088/1361-6560/ad4193.

[6] P. Krishnan, R. Kovvuri, G. Pang, B. Vassilev, and T. Hassner, "Textstylebrush: transfer of text aesthetics from a single example," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 7, pp. 9122-9134, 2023, doi: 10.1109/TPAMI.2023.3239736.

[7] A. Kumar, I. Memon, A. Sami, Y. Jo, and J. Choi, "Positional Component-Guided Hangul Font Image Generation via Deep Semantic Segmentation and Adversarial Style Transfer," Electronics, vol. 14, no. 13, pp. 2699-2710, 2025, doi: 10.3390/electronics14132699.

[8] A. Almohamade, S. Kammoun, and F. Alsolami, "Local Binary Pattern-Cycle Generative Adversarial Network Transfer: Transforming Image Style from Day to Night," Journal of Imaging, vol. 11, no. 4, pp. 108-122, 2025, doi: 10.3390/jimaging11040108.

[9] U. Jayasinghe, T. Fernando, and A. Fernando, "High-fidelity image transmission in quantum communication with frequency domain multi-qubit techniques," Algorithms, vol. 18, no. 8, pp. 501-517, 2025, doi: 10.3390/a18080501.

[10] G. E. Granados, R. Miorelli, F. Gatt, S. Robert, and D. Clouteau, "Towards a multi-fidelity deep learning framework for a fast and realistic generation of ultrasonic multi-modal Total Focusing Method images in complex geometries," NDT & E International, vol. 139, no. 1, pp. 102906-102919, 2023, doi: 10.1016/j.ndteint.2023.102906.

[11] S. Natarajan, L. Humbert, and M. A. G. Ballester, "Domain adaptation using AdaBN and AdaIN for high-resolution IVD mesh reconstruction from clinical MRI," International Journal of Computer Assisted Radiology and Surgery, vol. 19, no. 10, pp. 2063-2068, 2024, doi: 10.1007/s11548-024-03233-9.

[12] A. Satchidanandam, R. Mohammed, S. A. Ansari, D. S. Rao, D. S. Godla, and D. C. Kaur, "Enhancing style transfer with GANs: Perceptual loss and semantic segmentation," Int. J. Adv. Comput. Sci. Appl, vol. 11, no. 1, pp. 321-329, 2023, doi: 10.14569/IJACSA.2023.0141132.

[13] S. M. Rezaeijo, S. J. Nesheli, M. F. Serj, and M. J. T. Birgani, "Segmentation of the prostate, its zones, anterior fibromuscular stroma, and urethra on the MRIs and multimodality image fusion using U-Net model," Quantitative imaging in medicine and surgery, vol. 12, no. 10, pp. 4786-4798, 2022, doi: 10.21037/qims-22-115.

[14] A. Petsiuk, H. Singh, H. Dadhwal, and J. M. Pearce, "Synthetic-to-real composite semantic segmentation in additive manufacturing," Journal of Manufacturing and Materials Processing, vol. 8, no. 2, pp. 66-81, 2024, doi: 10.3390/jmmp8020066.

[15] A. Dharmalingam and V. K. K. Kolli, "TStarGANv2-VC: Non-parallel Multi-domain Transformer Based StarGANv2 Voice Conversion," Circuits, Systems, and Signal Processing, vol. (1), pp. 1-31, 2025, doi: 10.1007/s00034-025-03238-y.

[16] R. Touati and S. Kadoury, "Bidirectional feature matching based on deep pairwise contrastive learning for multiparametric MRI image synthesis," Physics in Medicine & Biology, vol. 68, no. 12, pp. 125010-125023, 2023, doi: 10.1088/1361-6560/acda78.

[17] S. Mulay, K. Ram, and M. Sivaprakasam, "Attention adaptive instance normalization style transfer for vascular segmentation using deep learning," Applied Intelligence, vol. 53, no. 24, pp. 29638-29655, 2023.

[18] T. Fontanini and C. Ferrari, "Would your clothes look good on me? towards transferring clothing styles with adaptive instance normalization," Sensors, vol. 22, no. 13, pp. 5002-5023, 2022, doi: 10.3390/s22135002.

[19] N. Q. Tuyen, S. T. Nguyen, T. J. Choi, and V. Q. Dinh, "Deep correlation multimodal neural style transfer," Ieee Access, vol. 9, no. 1, pp. 141329-141338, 2021, doi: 10.1109/ACCESS.2021.3120104.

[20] E. Batziou, K. Ioannidis, I. Patras, S. Vrochidis, and I. Kompatsiaris, "Artistic neural style transfer using CycleGAN and FABEMD by adaptive information selection," Pattern Recognition Letters, vol. 165, no. 1, pp. 55-62, 2023, doi: 10.1016/j.patrec.2022.11.026.

[21] A. Kumar, R. Rani, and S. Singh, "Encoder-Decoder Architecture for Image Steganography using Skip Connections," Procedia Computer Science, vol. 218, no. 1, pp. 1122-1131, 2023, doi: 10.1016/j.procs.2023.01.091.

[22] M. Mubashar, H. Ali, C. Grönlund, and S. Azmat, "R2U++: a multiscale recurrent residual U-Net with dense skip connections for medical image segmentation," Neural Computing and Applications, vol. 34, no. 20, pp. 17723-17739, 2022, doi: 10.1007/s00521-022-07419-7.

[23] M. Toshevska and S. Gievska, "A review of text style transfer using deep learning," IEEE Transactions on Artificial Intelligence, vol. 3, no. 5, pp. 669-684, 2021, doi: 10.1109/TAI.2021.3115992.

[24] T. Kausar, Y. Lu, A. Kausar, M. Ali, and A. Yousaf, "SD-GAN: A style distribution transfer generative adversarial network for Covid-19 detection through X-ray images," IEEE Access, vol. 11, no. 1, pp. 24545-24560, 2023, doi: 10.1109/ACCESS.2023.3253282.

[25] G. Bansal, A. Nawal, V. Chamola, and N. Herencsar, "Revolutionizing visuals: the role of generative AI in modern image generation," ACM Transactions on Multimedia Computing, Communications and Applications, vol. 20, no. 11, pp. 1-22, 2024, doi: 10.1145/3689641.