Visual Communication Strategies of Textile Patterns Driven by Digital Media

Xueling Zhang
Department of Visual Communication Design, Zhengzhou Academy of Fine Arts, Zhengzhou 451450, Henan, China
Corresponding author: Xueling Zhang (e-mail: 13513895873@163.com)

ABSTRACT With the rapid development of digital media and intelligent information transmission technologies, maintaining semantic consistency and stable multi-scale feature propagation has become increasingly important for visual communication systems and data-driven engineering applications, including future electromagnetic sensing and communication scenarios. However, textile pattern communication often suffers from fragmented information representation and disrupted transmission pathways, resulting in reduced visual coherence and communication effectiveness. To address these issues, this study proposes a visual communication optimization framework based on Swin Transformer (Swin-T) by integrating a Multi-Scale Attention (MSA) module and a path regularization strategy. The MSA module reorganizes hierarchical feature representations to achieve semantic alignment across different scales, while the path regularization constraint stabilizes similarity distributions during cross-layer propagation and suppresses feature dispersion. Experimental results demonstrate that the proposed framework achieves a maximum cosine similarity of 0.918 and a minimum Jensen–Shannon divergence of 0.183 during cross-scale transfer, while reaching an SSIM of 0.911 in geometric pattern scenarios and 0.871 in complex composite scenarios. The collaborative optimization strategy significantly improves visual consistency, information aggregation, attention concentration, and cross-scale transmission stability compared with existing approaches. The proposed framework provides an effective solution for intelligent textile pattern communication and offers valuable insights for multi-scale information representation and reliable feature propagation in future electromagnetic information processing and communication-oriented engineering systems.

INDEX TERMS swin transformer, multi-scale attention, visual communication, path regularization, cross-scale information propagation

I. INTRODUCTION

WITH digital media's pervasive penetration in society, expressing textile patterns through images is encountering significant disruptions amidst the transition of audiences' aesthetic needs from twodimensional viewing to multi-mode experiencing. Due to issues like broken information continuity and interrupted communication routes, pattern expression based on textile fabrics' physical characteristics gradually reveals problems, for instance, design and visual effects becoming two separate parts and complete communication routes being split apart [1,2]. Information discontinuity in textile design represents the gap between a pattern's design and its final visual effect, where digital translation weakens the continuity of texture, gloss, and structure. Fragmented communication routes emerge as images are transferred among different media, where each damages visual data and scatters symbols. These problems impair semantic consistency and prevent designers and audiences from reaching a shared meaning. Thus, an integrated mechanism is urgently needed to guarantee coherent information and meaning in different media. The validity of the proposed strategy is verified by quantitative indicators to maintain semantic consistency in the study. One the one hand, the lack of visual consistency between screens and textiles, caused by differences in color reproduction and detail rendering, hinders the accurate transmission of design intent [3,4]. On the other hand, the lack of tactility and material feedback in digital communication also causes deviations in audiences when understanding textile design [5,6]. There have been some studies discussing these problems from the viewpoints of big data trend analysis, digital printing color gamut management, light and shadow perception optimization, and even cultural interpretation. However, in general, there are still some problems, such as fragmented technology application scenario, lack of cross-scale information integration, and lack of systematic visual communication framework [7,8]. Therefore, achieving stable, efficient, and semantically consistent visual communication of textile patterns driven by digital media has become a key challenge that needs to be addressed urgently. This study uses visual consistency and cross-scale semantic stability as training objectives to ensure that pattern structure and

details remain continuous in the propagation chain, and uses this as the basic task constraint for the model. Driven by digital media, the visual communication of textile patterns is undergoing a profound transformation, and different scholars have explored this from multiple dimensions, including technology and perception. Lee and Suh employed big data methods to identify four major design trend clusters driven by technology, offering a macro perspective for understanding the overall direction [9]. Moon and Chae found that the color gamut in digital printing differed significantly, and the color reproducibility between the display and the fabric decreased significantly [10]. Focusing on sensory experience, Haghzare et al. demonstrated that light and shadow details can significantly alter the perception of online fabrics by manipulating light reflection and three-dimensional dynamic display [11]. In response, Jang and Ha's research showed that although digital media has expanded the channels of visual expression, the lack of tactile information has led to deviations in design understanding [12]. Tkalec et al. compared the differences in detail and color reproduction between handmade and digital printing, emphasizing the important role of material properties in visual communication [13]. Vargas et al. pointed out from a cultural perspective that current research suffers from the problem of data monopoly and the separation of national cultural interpretation [14]. Overall, these studies reveal that the digital media environment has changed the audience's aesthetic habits and the way they receive information, but there is a phenomenon of disconnected information expression and scattered communication paths in the visual communication of textile patterns, which hinders the expression effect and communication optimization. In recent years, technologies based on deep learning and multi-scale modeling have provided new methods and perspectives for optimizing the visual communication of textile patterns. Xu et al. significantly improved the detail quality and semantic consistency of textile patterns through multi-scale optimization and attention mechanisms [15]. At the same time, Liu et al. reviewed 101 articles from the past decade. They summarized the application trends and performance of intelligent technologies and optimization algorithms in the four major subfields of textile color management [16]. In addition, Xu et al. proposed a fiber recognition framework based on multi-head attention, which significantly improved the accuracy of fine-grained fiber recognition on a textile surface image dataset [17]. Skliarenko et al. reconstructed the multi-level application mechanism of symmetry in dynamic visual communication, providing a structured design strategy for enhancing the communication efficiency of visual information [18]. Although these studies have made progress in pattern detail optimization, fiber recognition accuracy, and visual aesthetic strategies, they still have problems such as limited application scenarios, insufficient cross-scale integration, and the lack of a systematic optimization framework. To address this dilemma, this article constructs a visual communication strategy centered around Swin Transformer (Swin-T). Specifically, Swin-T's hierarchical window partitioning structure is first utilized to perform a multi-level representation of the local texture and overall layout of textile patterns, thereby balancing details and global relationships in the initial modeling phase. Secondly, a multi-scale attention mechanism is applied during feature interaction to dynamically weight texture morphology and color distribution at different scales, ensuring consistency of semantic information across resolutions. Thirdly, a path regularization term is introduced to restrict deviations in the similarity-based transition pattern during cross-layer propagation and reduce unnecessary scattering of feature relations. Its role is limited to regulating the statistical structure of cross-layer relationships and does not involve the description of perception paths.

II. OPTIMIZATION STRATEGIES FOR VISUAL COMMUNICATION OF TEXTILE PATTERNS IN DIGITAL MEDIA ENVIRONMENTS

A. VISUAL FEATURE MODELING METHOD BASED ON SWIN-T

The visual communication of textile patterns requires hierarchical modeling of local textures and overall structures to ensure the stability and integrity of visual features during the communication process [19,20]. Swin-T achieves this requirement through a hierarchical window attention mechanism in pattern feature modeling. Its method is to divide the image into fixed-size windows, apply multi-head self-attention within the window to capture local information, and then expand global dependencies by moving the window and stacking the layers, thereby achieving progressive modeling from local to global. Assuming that the input textile pattern is an image I of size $H \times W$ with C channels. After linear mapping, the initial feature tensor $F_0 \in \mathbb{R}^{H \times W \times C}$ is obtained. This feature tensor is divided into windows of size $M \times M$. Each window M^2 contains image blocks. The features within the window are represented as $X \in \mathbb{R}^{M^2 \times d}$, where d is the embedding dimension. The calculation formula of the window attention mechanism is:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}} + B\right)V \quad (1)$$

In Equation (1), $Q = XW_q$, $K = XW_k$, $V = XW_v$, where W_q, W_k, and W_v are learnable linear transformation matrices; d_k is the dimension of K; B is the relative position bias matrix. The scaling factor is expressed as $1/\sqrt{d_k}$ to ensure numerical stability when calculating similarity. After extracting local features within the window, the window translation strategy is used to enhance cross-window information interaction. The translation amount is defined as:

$$\Delta = \left(\frac{M}{2}, \frac{M}{2}\right) \quad (2)$$

The translation is calculated by shifting the window halfway horizontally and vertically. The shifted window is then re-divided into features X', and the attention calculation process is repeated to fully represent cross-region features. In

the layer-by-layer network, the feature update is formalized as:

$$Z_i = MSA(LN(X_{i-1})) + X_{i-1} \tag{3}$$

$$Y_i = MLP(LN(Z_i)) + Z_i \tag{4}$$

In Equations (3) and (4), MSA defines multi-head window self-attention module; LN specifies layer normalization function; MLP refers to neural network layer (feedforward network); X_{i-1} specifies the input representation fed into the previous layer (input feature set); Z_i specifies the feature representation following the application of self-attention; Y_i defines output (prediction) from feedforward network in conjunction with residual connection mechanism employed. This enables the stepwise increasing of a feature's representation from a local context to a global context and provides for stable gradient dissemination via residual pathways. After multiple layers are stacked, the network forms a pyramidal feature representation through hierarchical downsampling, resulting in a multi-scale feature set $\{F_1, F_2, F_3, F_4\}$, where each feature tensor satisfies $F_i \in \mathbb{R}^{h_i \times w_i \times c_i}$, $h_i = H/2^i$, $w_i = W/2^i$, and the number of channels c_i increases with the number of layers. This set serves as input in the subsequent multi-scale information alignment and path regularization stages, providing structured feature support for optimizing the visual communication of textile patterns [21,22]. To more intuitively demonstrate the process of visual feature modeling of textile patterns based on the Swin-T, this article constructs the overall framework of this method, as shown in Figure 1. In Figure 1, the framework takes a textile pattern image as input, first implementing local modeling of the initial features through linear mapping and window partitioning; then, in the local modeling layer, the multihead self-attention mechanism and position bias are used to complete local feature extraction and updating; feature fusion between different regions is achieved through the cross-window interaction layer; residual connections and normalization operations are applied in the hierarchical stacking layer to ensure training stability; finally, progressive modeling from local to global is completed in the global modeling layer, and a multi-scale feature set is formed through downsampling, laying the feature foundation for subsequent crossscale semantic alignment and path regularization. The Multi-Scale Attention module processes these outputs without altering the window self-attention structure of the Swing Transformer.

B. INTEGRATION OF MSA MODULES AND SEMANTIC ALIGNMENT OF CROSS-SCALE INFORMATION

The Swin-T's attention mechanism operates within a single-resolution window range to update embedding representations within the same scale. Our study's multi-scale attention operates on the hierarchical output of the Swin-T, handling feature relationships between different resolutions, and is not a repetitive stacking of internal self-attention. This module establishes a semantic alignment structure between multi-scale results, maintaining a clear distinction from the Swin-T's window attention in terms of input source and functional purpose. All subsequent mathematical expressions in this section consistently describe this single cross-scale attention module applied to Swin-T outputs. After encoding through the Swin-T, a multi-scale feature set is obtained. The design goal of the MSA module is to map these features into a unified semantic space and achieve information interaction through cross-scale attention, thereby improving the alignment and transfer stability of pattern features at different levels. Then, to align the features at different scales to the same resolution, we upsample or downsample the features via upsampling or downsampling, and to align the number of channels to be the same dimension, we apply linear projection. To align features across different scales, low-resolution features are upsampled to the target spatial resolution using bilinear interpolation, ensuring continuous spatial correspondence during scale expansion. High-resolution features are transformed to the target resolution through stride convolutions, which perform controlled downsampling while preserving dominant structural responses. After spatial alignment, a standard convolution without stride is applied to refine local continuity and stabilize feature distribution across scales. This design enforces spatial consistency during scale transformation and establishes a unified structural basis before channel projection. While the high-dimensional feature is first convolved with a kernel of size 3 and stride 2 to output a downscaled feature, to ensure that the spatial structure is retained during mapping, we apply a convolution without any stride. Finally, after spatial alignment, one linear mapping is applied to perform channel transformation, ensuring that all scales share a stable distribution within the same embedding dimension. Finally, the features with the same dimension are transferred to Q, K, and V via a linear transformation across all dimensions, making sure that all scales have a stable distribution in the same dimension before feeding into the attention computation. Specifically, Q is extracted from the current scale, while K and V are output from all scales, making sure that all scales have a complete mapping across cross-scale relationships in the attention matrix. For the input of the i-th scale, the query, key, and value are defined as $Q_i = \hat{F}_i W_q^i$, $K_i = \hat{F}_i W_k^i$, and $V_i = \hat{F}_i W_v^i$, where W_q^i, W_k^i, and W_v^i are trainable parameter matrices. The following formulation gives the unified and explicit definition of cross-scale attention used throughout this work. All symbols in this section refer to the cross-scale interaction module operating on the hierarchical outputs of the Swin-T, rather than to the internal window self-attention of Swin-T. The same symbol system is consistently adopted in the subsequent descriptions to avoid ambiguity between module-level definition and its concrete implementation. The core calculation formula for cross-scale semantic alignment is:

$$\tilde{F}_i = \sum_j \mathrm{Softmax}\left(\frac{Q_i K_j^T}{\sqrt{d}} + B_{ij}\right) V_j \tag{5}$$

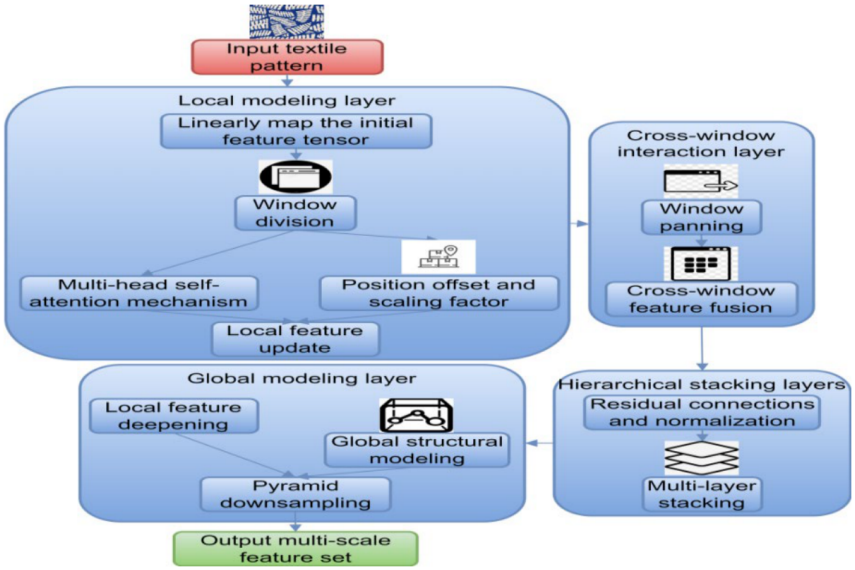

FIGURE 1. Textile Pattern Visual Feature Modeling Framework Based on Swin-T

In Equation (5), the similarity between the query at scale i and the keys at all scales j is measured and updated using the corresponding values. Bij represents the cross-scale position bias, which is used to maintain the spatial structure relationship between different scales. The bias term is generated based on the spatial offset between scales at a uniform resolution. It transforms the spatial correspondence between different scales into discrete offsets and records them with a learnable matrix, so that the bias continuously expresses the spatial relationship between scales in the attention weights. The processed features contain both detail level information and structural level information. The updated results are stabilized through layer normalization and residual connection:

$$F_i' = LN(\tilde{F}_i) + F_i \tag{6}$$

This design guarantees that the cross-scale feature can be unified without destroying the original feature. Meanwhile, the numerical stability of propagation process is also well preserved. The obtained fused tensor is always spatially and channel-wise consistent with the residual path of uniform dimension. The normalization process is order by scale, which keeps the stability of distribution variation after cross-scale fusion. Moreover, the MSA module allows the local texture and global composition of textile pattern to be uniformly constrained in the same modeling framework, which offers a semantically consistent input basis for the following path regularization constraint. The optimization of cross-scale alignment is denoted by L_{msa}. Let F_i and F_i' denote the input and output features of the cross-scale attention defined in Equation (5), respectively, and their definitions are:

$$L_{msa} = 1 - \frac{1}{N}\sum_{i=1}^{N} \cos(F_i, F_i') \tag{7}$$

This item is used to limit the structural offset during the alignment process, keeping the semantic mapping stable.

C. IMPLEMENTATION AND OPTIMIZATION EFFECT OF PATH REGULARIZATION CONSTRAINTS

In the process of visual communication driven by digital media, information dissemination between different levels often suffers from path fragmentation and redundancy, resulting in disjointed representation of pattern features at the audience end [23,24]. To address this issue, a regularization term is applied to stabilize the similarity distribution among adjacent layers during propagation. The term addresses the fragmentation of relational consistency within the computational hierarchy rather than any perceptual pathway. During the transmission of pattern information, uncontrolled dispersion of multi-scale features often leads to interruptions in perceptual linkage between visual elements. By regulating the propagation direction and intensity of feature transfer, the model forms a continuous and centralized trajectory of information flow, which corresponds to the audience's natural tendency to follow coherent visual routes when interpreting textile imagery. The purpose of this regularization is to suppress redundant and scattered inter-layer transitions while maintaining structured differentiation among layers, so that information propagation is guided toward a controlled central tendency rather than driven toward uniform diffusion or premature path collapse. This adjustment forms a smoother progression of inter-layer relations by reducing abrupt distributional fluctuations during propagation. The core idea is to regulate the probability distribution governing inter-layer influence so that the transition process avoids disproportionate dispersion and forms a centralized yet non-degenerate information trajectory. The target distribution p^* is defined as a directional reference distribution constructed from stabilized cross-layer similarity statistics, rather than directly adopting the empirical path frequencies observed

in early unregularized training. The balance coefficient is set between 0.1 and 1.0 and is updated according to the reduction of path divergence on the validation set. The visual information transmission process is represented as a directed acyclic graph G constructed at the feature-subspace level. Each layer i is decomposed into M feature subspaces produced by multi-head and multi-scale representations, and each node (i, m) corresponds to the m-th feature subspace within layer i. Directed edges are established only between subspaces of adjacent layers, forming transitions from (i, m) to $(i + 1, m')$, which preserves the ordered propagation structure while introducing multiple concurrent transition candidates between neighboring layers. The graph is therefore organized as a set of structured inter-layer subspace transitions rather than a single one-to-one layer mapping. All transition probabilities originate from the similarity values defined earlier and are computed between feature subspaces across adjacent layers. The transmission intensity between two connected subspace nodes is described by a probability variable $p((i, m) \to (i+1, m'))$, and the overall flow of the communication process is expressed as:

$$T = \sum_i \sum_j p_{ij} f_{ij} \tag{8}$$

Among them, f_{ij} represents the transferred feature vector. Path probability set $P = \{p_{ij}\}$ defines the regularization term already demonstrated in Equation (8), representing the divergence loss between real distribution P and target distribution P^*. In the process of gradient-based learning, the derivative $\frac{\partial L_{\text{path}}}{\partial p_{ij}}$ can adaptively update the propagation strength, which can help the information flow concentrate on the most stable path and alleviate the unnecessary dispersion in the whole process of visual information propagation. The feature representation after updating by the MSA module is denoted as F_i'. The path probability distribution $p(i, m) \to (i + 1, m')$ is defined during cross-layer propagation, representing the transition weight from a feature subspace in layer i to a feature subspace in the next layer. The KL-based term stabilizes the distributional form of similarity across layers. When these distributions fluctuate sharply, multi-scale features lose relational coherence. Restricting their divergence yields a continuous transfer pattern within the model. The effect is confined to statistical regulation and does not imply perceptual equivalence. To avoid excessive path divergence, a regularization term is applied:

$$L_{path} = \sum_i \sum_j p_{i\to j} \log \frac{p_{i\to j}}{\pi_j} \tag{9}$$

In Equation (9), π_j represents a directional target distribution that encodes a centralized transition preference across subspaces of adjacent layers. The shape of π_j is derived from normalized cross-scale similarity profiles averaged over stabilized training stages, which provides a convergent orientation for information flow without forcing it to replicate any single early trajectory. It is constructed to emphasize structurally stable and semantically consistent paths, rather than assigning equal probability to all transitions or directly copying empirical frequencies. Path weights are determined by feature similarity and structural constraints. The similarity between layer i and layer j is calculated as:

$$S_{i\to j} = \frac{F_i' \cdot F_j'}{\|F_i'\| \|F_j'\|} \tag{10}$$

Before being used in the path computation, the similarity term is scaled by a temperature τ as $S_{i\to j}/\tau$. The path probabilities are obtained through a softmax normalization using $\exp(S_{i\to j}/\tau)$, which yields a continuous and fully differentiable distribution during backpropagation. The softmax normalization is applied over all subspaces in layer $i+1$ that are connected to a given subspace in layer i, forming a valid local transition distribution under the adjacent-layer constraint. The path probability is defined as:

$$p_{i\to j} = \frac{\exp(\lambda S_{i\to j})}{\sum_k \exp(\lambda S_{i\to k})} \tag{11}$$

In Equation (11), λ is the balancing coefficient, which is used to adjust the influence of similarity on the path distribution. Then, via this normalization, the sum of transition probabilities from one subspace in layer i to all candidate subspaces in layer $i + 1$ equals 1, ensuring that the information transfer is distributed only within the next layer. In the training part, what will be optimized in total includes a task loss and a path regularization loss. The task loss L_{task} gives the supervisory signal and we design it to keep structural consistency across different scales. It is defined as:

$$L_{task} = 1 - SSIM(\hat{x}, x) \tag{12}$$

where x denotes the reference pattern and $\hat{x}$ denotes the predicted representation. This formulation maintains spatial coherence and stabilizes the main optimization target. During the training phase, the overall optimization objective is composed of the task loss and the path regularization loss.

$$L_{total} = L_{task} + \beta L_{path} \tag{13}$$

In Equation (13), L_{task} is the term representing the loss for the visual communication task (consistency measurement or attention concentration optimization) and β is the regularization coefficient used to control the influence of path regularization on the main objective. During the training procedure, we perform a grid search over [0.1,1.0] to choose β. Finally, the value of β will keep the ratio between the task objective and the regularization term around a constant value. Thus, the regularization does not dominate or neglect the main objective. Using path regularization constraints improves how textile patterns are visually communicated. Improved visual communication includes improved consistency between the local features and the overall structure of the pattern; improved methods for aggregating cross-scale information; and improved audience focus on their visual communication design. Additionally, experimental results show that the application of path regularization reduces

redundancy in the features contained in the communicated patterns, as well as improves the consistency of patterns presented across multiple terminals and media environments by establishing a reliable method for optimizing the visual communication of textile patterns.

D. IMPLEMENTATION AND EFFECT IMPROVEMENT OF COLLABORATIVE OPTIMIZATION STRATEGY

Swin-T has built a hierarchical model, which creates an array of features at different levels. Multi-Scale Attention creates a "consistent semantic space" for those different levels of features. Path Regularization provides a way to impose a directional central tendency on cross-layer subspace information transfer, stabilizing propagation routes while suppressing redundant and unstructured dispersion. A collaborative optimization strategy utilizes all three of these components and combines them into a single Optimization Framework. In this framework, the input pattern first extracts multi-scale features through Swin-T, and then performs cross-scale fusion through the MSA module to obtain the updated features F_i'. During the cross-layer propagation process, path regularization constraints control the information distribution and avoid the divergence of the transmission chain. L_{msa} and L_{task} together constitute a unified training constraint, ensuring that feature updates follow a consistent optimization objective. The overall loss function is defined as:

$$\hat{L}_{total} = L_{task} + \alpha L_{msa} + \beta L_{path} \tag{14}$$

In Equation (14), L_{msa} represents the cross-scale alignment constraint; L_{path} represents the path regularization constraint; α and β represent weight parameters. This paper employ AdamW optimizer with an initial learning rate of 0.0001 and cosine annealing scheduling method. The number of epochs in total is 300. The batch size is fixed to 16 according to the capacity of GPU memory. The final loss function is the weighted average of task loss, multi-scale alignment loss and path regularization loss as shown in Equation (14). We calibrate the hyperparameters α and β by grid search and find their optimal values are 0.5 and 0.3, respectively. The training process is to optimize feature semantic alignment and the stability of information propagation paths simultaneously. This objective form enables semantic alignment and path adjustment to form a closed structure, maintaining a stable relationship for cross-scale propagation. After collaborative optimization, textile patterns exhibit higher consistency and convergence in visual communication, more stable cross-scale information transfer, more focused attention distribution, and overall communication effectiveness superior to that of a single strategy.

III. VERIFICATION OF THE VISUAL COMMUNICATION EFFECT OF TEXTILE PATTERNS UNDER OPTIMIZATION STRATEGY

A. VISUAL CONSISTENCY ANALYSIS AND VERIFICATION

In the study of visual communication of textile patterns, evaluating a method's performance at different training stages is crucial for verifying model effectiveness. To comprehensively demonstrate the differences among methods in terms of structural preservation and visual consistency, this experiment uses the Structural Similarity Index (SSIM) as a metric. The original textile pattern image serves as the reference, and the output from each method is the test image. A higher SSIM value indicates better preservation of the original structure, directly reflecting superior visual consistency in the digital communication process. The SSIM evaluation was performed on pixel-wise aligned image pairs. All output images maintained the original spatial structure without any post-processing warping, ensuring the metric's validity in structural comparison. This experiment compares the performance of FabricDiffusion, the Hierarchical Visual Sensibility Network (HVS-Net), the Inversion-based style transfer method (InST), and the method of this article across four scenarios. All baseline methods are implemented using their official codebases and are trained under identical settings, including the use of the same TextileNet and FID datasets, a unified input resolution of 512 $\times$ 512 pixels, and consistent training hyperparameters encompassing 300 epochs, a batch size of 16, and the AdamW optimizer. Identical image preprocessing is applied throughout. The performance trends as the number of iterations increases reveal the convergence characteristics and stability of the methods under diverse pattern features. The SSIM index is computed with an 11 $\times$ 11 pixels Gaussian window, following the standard implementation. Figure 2 shows the SSIM curves for each method under multiple scenario conditions. In Figure 2, in the geometric pattern scene, FabricDiffusion obtains 0.822 SSIM in the 100th iteration, while HVS-Net and InST get 0.851 and 0.862, respectively. The proposed method reaches 0.911, demonstrating a stronger capability for structural preservation. The overall SSIM value of natural texture scene is relatively low. When HVS-Net and InST are approaching to convergence, they get 0.831 and 0.846, respectively, while the method of this article is stabilized at 0.892, which shows that it has better ability to preserve texture details. Abstract representation scenes have higher requirements for semantic feature extraction. When reaching to the later stage, FabricDiffusion achieves 0.777, while HVS-Net and InST are improving gradually to 0.808 and 0.827, respectively. The method of this article remains at 0.881 and still has significant advantage. Complex composite scenes bring the greatest challenge overall. When reaching to the final, FabricDiffusion achieves SSIM 0.760, while HVS-Net and InST are improved to 0.789 and 0.807, respectively. The method of this article achieves 0.871, which shows that when there are multiple patterns overlaying, it still can keep certain information consistency. Overall, the method of this article has a smoother and

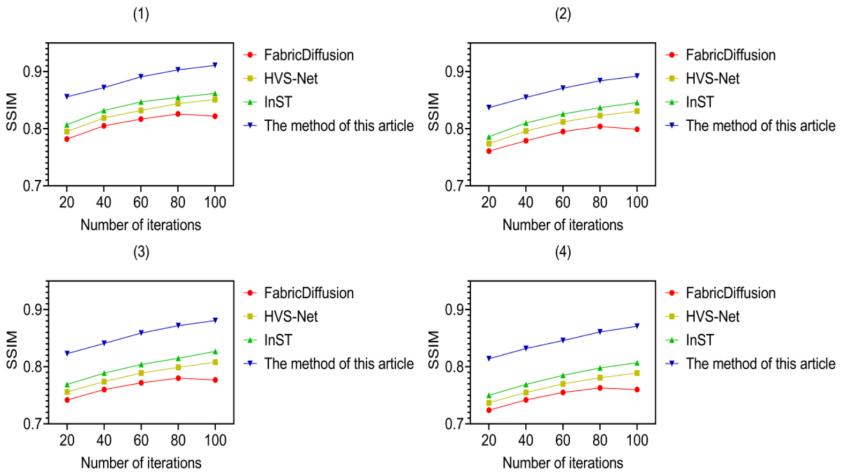

FIGURE 2. SSIM Curves of Various Methods under Multiple Scene Conditions: (1) Geometric Pattern; (2) Natural Texture; (3) Abstract Expression; (4) Complex Composite

more consistent SSIM enhancement curve in improving SSIM in all above scenarios, which shows that it has greater adaptability and stability in conveying various textile patterns visually.

B. ANALYSIS OF THE EFFECTIVENESS OF INFORMATION AGGREGATION MEASUREMENT

In image processing and style transfer, the comparison of image feature extraction and style transfer at different scales is of great significance. In order to evaluate the performance of these methods at different scales, the experiment chooses four different algorithms, i.e., the method of this article, FabricDiffusion, HVS-Net and InST. The experimental design is based on four evaluation metrics, including Feature Consistency (FC), Structural Entropy (SE), Edge Density (ED) and Visual Semantic Density (VSD). The FC is used to evaluate the feature consistency, which is calculated by mean cosine similarity between multi-scale feature vectors. SE is used to evaluate the texture complexity, which is calculated based on gray-level co-occurrence matrix. ED extracts the edges, and then calculate the ratio of edge pixels. VSD uses the pre-trained visual encoder to extract features, and then calculate the L2 norm to characterize the semantic information density. These four metrics characterize the information aggregation effect from the view of feature stability, texture complexity, structural clarity and semantic richness, respectively. These four metrics can reflect the performance of these algorithms at different scales and provide a basis for analyzing the performance of these algorithms at different scales and finding their advantages and disadvantages in multi-scale image processing. As shown in Figure 3, the performance of the four methods on these metrics at three repeating unit scales. As shown in Figure 3, while the method of this article achieves higher SE and ED values (for example, SE is 0.88, and ED is 0.85 at the large scale; the comparison methods achieve the highest values of 0.87 and 0.84, respectively), the difference is only 0.01. However, in terms of positive metrics, the method of this article maintains a significant advantage in both FC and VSD, achieving FC of 0.87 and VSD of 0.95 at the large scale, while the comparison methods achieve the highest values of only 0.81 and 0.89. This indicates that, under multi-scale conditions, the proposed method achieves significant improvements in FC and VSD, while also maintaining improvements in the SE and ED metrics, resulting in superior overall performance.

C. STABILITY VERIFICATION OF CROSS-SCALE INFORMATION TRANSFER

To avoid ambiguity in the interpretation of Figure 4, the four evaluation stages correspond to four fixed crossscale propagation points in the hierarchical architecture. Stage 1 denotes the feature transition from the highest-resolution Swin-T output to the first aligned multi-scale representation after the MSA module. Stage 2 denotes the subsequent propagation from this aligned representation to the next lower-resolution scale after cross-scale attention fusion. Stage 3 denotes the further transition between intermediate scales under the joint influence of cross-scale attention and path regularization. Stage 4 denotes the final propagation from the last intermediate scale to the deepest feature representation used for output consistency evaluation. Each stage therefore reflects a concrete step of hierarchical information transfer rather than independent training phases, and all measurements are computed on features extracted immediately after each defined propagation step. During cross-scale feature transfer, whether the model can maintain stable representation consistency directly affects the reliability of the final generated results. To this end, this experiment applies two metrics: cosine similarity and Jensen–Shannon (JS) divergence. From the perspective of similarity and distribution difference, the performance of FabricDiffusion, HVS-Net, InST, and the method of this article at each scale level is evaluated, verifying their stability during multi-scale transfer. Experiments are conducted using TextileNet, a publicly accessible dataset developed for textile

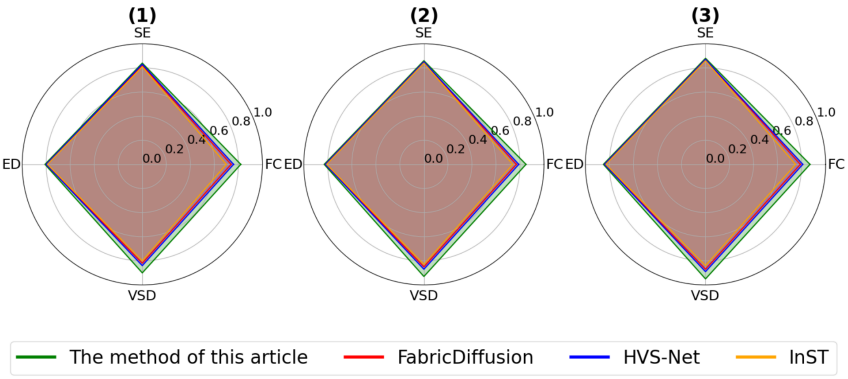

FIGURE 3. Performance Comparison of Four Methods at Different Repeating Unit Scales: (1) Small Repeating Unit Scale; (2) Medium Repeating Unit Scale; (3) Large Repeating Unit Scale

material recognition and fabric analysis, containing over 760,000 images labeled with fibre and fabric taxonomies. The Fabric-Image-Data (FID) dataset is additionally incorporated, comprising 12,181 wool fabric images categorized into lattice, pattern, solid, and stripe classes. From GitHub TextileNet, a balanced subset of 60,000 images covering representative categories in fiber and fabric classification is selected. The subset is constructed by stratified random sampling across ten major fabric categories to ensure class balance. Potential near-duplicate images are removed using perceptual hashing with a similarity threshold of 0.95. The complete TextileNet dataset is publicly available at https://github.com/hahashu/TextileNet. Verification confirms no image source overlap exists between TextileNet and the FID dataset. From FID, all 12,181 images are used. The combined dataset is randomly split into training (70%), validation (15%), and testing (15%) subsets at the image level, ensuring class balance in all splits. Before feature extraction, preprocessing is applied: all images are resized to 512 $\times$ 512 pixels; pixel intensities are normalized to [0,1]; color normalization is performed for each dataset where applicable to mitigate domain shift between datasets. Such preprocessing guarantees the consistency of the statistical profiles of scale and intensity across different datasets. All the experiments are implemented on a workstation with NVIDIA GeForce RTX 4090 GPU. The total training time of our model is around 72 hours for all stages. This dataset composition guarantees that the cosine similarity and JS divergence results are validated on robust cross-dataset generalization and multi-scale stability. The cosine similarity is evaluated on L2-normalized feature vectors extracted from the final model layer. Thus, the metric compares the angular difference in feature space. Figure 4 shows the evaluation results corresponding to Stage 1 through Stage 4, which represent successive hierarchical cross-scale propagation steps defined in this section. The results in Figure 4 show that the method of this article consistently maintains a high level of cosine similarity, reaching 0.918 in Stage 1 and 0.868 in Stage 4. This is higher than FabricDiffusion's 0.689 in Stage 4, demonstrating that the method of this article can reduce feature attenuation during multi-level information transfer and ensure cross-scale consistency. Regarding the JS divergence metric, the method of this article achieves the lowest level at all four scale levels, with a Stage 1 score of 0.183 and a FabricDiffusion score of 0.271. This indicates that the method of this article effectively reduces the offset between feature distributions during the step-by-step transfer, making cross-scale feature propagation closer to the true distribution. The overall results demonstrate that the method of this article has a clear advantage in the stability of cross-scale information transfer. In terms of model complexity, compared with Swin-T backbone without MSA module and path regularization term, the whole network architecture contains in total 87.5 M parameters. For the input resolution of 512 $\times$ 512 pixels, the number of GFLOPs for one forward pass of the model is around 245, which is moderate in both performance and cost on the hardware platform we used.

D. COMPREHENSIVE OPTIMIZATION STRATEGY IMPROVES COMMUNICATION EFFECTIVENESS

To validate the effectiveness of the comprehensive optimization strategy, this section compares the overall performance of different methods in the visual communication process using multi-dimensional indicators, thereby revealing the patterns in performance differences between single optimization and collaborative optimization. The experiment uses Swin-T as the baseline method, gradually superimposing the MSA module and path regularization constraints to form four groups of solutions. These are compared in terms of visual consistency, information aggregation, attention focus, and cross-scale stability. The relevant results are shown in Figure 5. It integrates the metrics clearly defined in the previous sections: SSIM values of strictly aligned sample pairs, multi-metric aggregation, internal feature entropy, and independently calculated crosslayer similarity. All data sources are transparent and verifiable. As shown in Figure 5, Swin-T can obtain 0.83 on visual consistency and further improve to 0.88 after MSA module directly learns

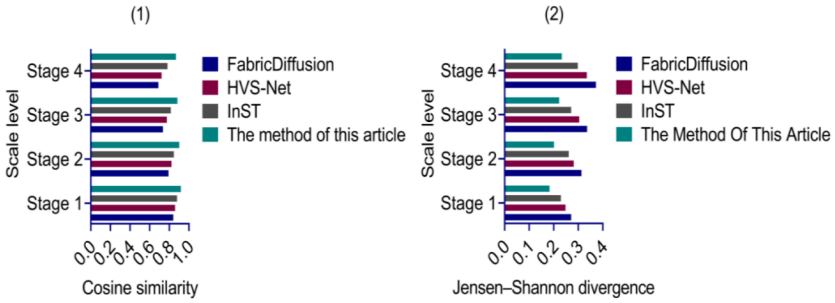

FIGURE 4. Comparison of Multi-Scale Visual Consistency Evaluation Results: (1) Cosine Similarity; (2) JS Divergence

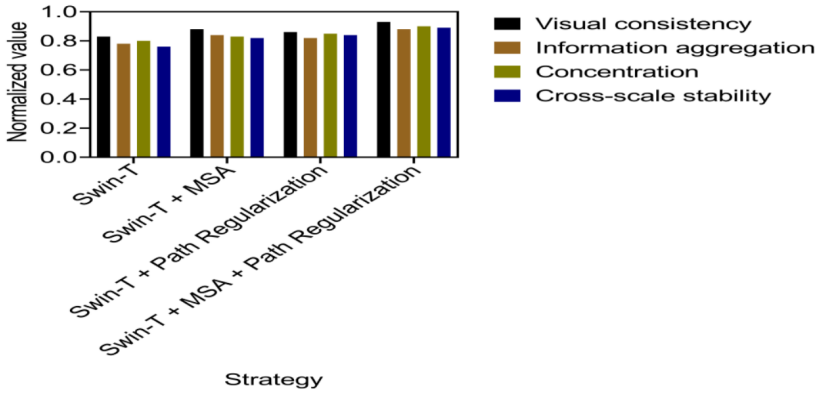

FIGURE 5. Comparison of Different Optimization Strategies Based on Multi-Dimensional Indicators

the multi-scale attention. The effectiveness of multi-scale attention module on feature alignment can be validated by the attention concentration. This paper defines the entropy of channel responses in the channel-wise entropy of the final feature map of the model as attention concentration, which reflects the dispersion of feature activation. Lower channel-wise entropy indicates that feature responses are distributed over a smaller number of dominant channels, meaning that the model allocates its representational capacity to a more centralized set of visual regions and structural cues. Under visual communication settings, this statistical concentration corresponds to a reduced dispersion of salient visual information within the pattern representation. Since the propagated features determine which visual elements are emphasized in the final output, a lower entropy value implies that the generated textile patterns present clearer visual centers and more coherent perceptual organization. This property provides a quantitative basis for interpreting attention concentration as a computational analogue of audience view concentration, because the model output guides the distribution of visually dominant information that audiences are more likely to focus on. When adding path regularization, attention concentration further improves to 0.85, which is obviously better than baseline 0.83 on MSA. This improvement indicates that the collaborative strategy effectively suppresses the dispersion of feature activation and drives the visual representation toward a more centralized distribution, which directly supports the claim that the proposed framework enhances the concentration of visual attention in the communicated textile patterns. The increase trend can also be intuitively reflected by the length of bar in Figure 5 confusion of constrained path transfer on suppressing distraction of information diffusion. Cross-scale stability was independently calculated as the average of cosine similarity of feature transmission of all scale levels on test set, which can be treated as a generalized evaluation indicator. As shown in Figure 5, compared with Swin-T only achieving cross-scale stability of 0.76, model under full collaborative strategy can further improve to 0.89, indicating better generalization performance. The improvement could be attributed to the synergistic effect of path constraint and semantic alignment on cross-scale feature transfer. Overall, Swin-T's combination of MSA and path regularization achieves optimal performance across all four dimensions, demonstrating a balanced and stable optimization trend, and overall superiority of collaborative optimization strategies in visual communication tasks.

IV. CONCLUSION

This article focuses on the visual communication problem of textile pattern in digital media. To optimize the visual communication efficiency, a visual communication optimization framework based on Swin-T is proposed. A multi-scale attention mechanism and path regularization constraints are introduced. The regularization term imposes a directional statistical constraint on cross-layer transfer, guiding information flow toward a centralized and coherent trajectory while suppressing redundant and disordered propagation. The framework can ensure the stable transfer of information across multiple scales and the semantic consistency of textile patterns. The experimental results show that the SSIM in the geometric pattern scene reaches 0.911, compared with 0.822 for FabricDiffusion and 0.851 for HVS-Net; the SSIM in the composite scenario reaches 0.871, compared with 0.807 for InST. The experimental results show that the proposed method preserves pattern details and cross-scale consistency, and significantly increases attention concentration measured by channel-wise entropy reduction, indicating that the optimized visual representations present a more centralized distribution of salient visual information, which supports the conclusion of enhanced audience view concentration in digital textile pattern communication.

AUTHOR CONTRIBUTIONS

Xueling Zhang designed, collected and analyzed the data, and drafted the manuscript. Xueling Zhang conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Xueling Zhang participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] X. Yuan and N. Chuprina, "Ecological Design Concept of Textile Products: A Study on Brocade Weaving Materials of the Chinese Ethnic Minority Yao," Art and Design, vol. (4), pp. 51-61, 2024, doi: 10.30857/2617-0272.2024.4.4.

[2] M. A. Habib and M. S. Alam, "A comparative study of 3D virtual pattern and traditional pattern making," Journal of Textile Science and Technology, vol. 10, no. 1, pp. 1-24, 2023, doi: 10.4236/jtst.2024.101001.

[3] Z. A. Awad, M. A. Eida, H. S. Soliman, M. A. Alkaramani, I. G. EIbadwy, and A. G. Hassabo, "The psychological effect of choosing colors in advertisements on stimulating human interaction," Journal of Textiles. Coloration and Polymer Science, vol. 22, no. 1, pp. 289-298, 2025, doi: 10.21608/jtcps.2024.259790.1323.

[4] J. Liu, "Research on multimedia art expression in visual communication design," Advances in Social Sciences, vol. 13, no. 4, pp. 250-256, 2024, doi: 10.12677/ass.2024.134296.

[5] S. Y. Jang and J. Ha, "The influence of tactile information on the human evaluation of tactile properties," Fashion and Textiles, vol. 8, no. 1, pp. 39, 2021, doi: 10.1186/s40691-020-00242-5.

[6] A. Chen, J. Tan, and P. Henry, "E-textile design through the lens of affordance," Journal of Textile Design Research and Practice, vol. 9, no. 2, pp. 164-183, 2021, doi: 10.1080/20511787.2021.1935110.

[7] M. T. Fischer, F. L. Dennig, D. Seebacher, D. A. Keim, and M. E. Assady, "Communication Analysis through Visual Analytics: Current Practices, Challenges, and New Frontiers," 2022 IEEE Visualization in Data Science (VDS). Oklahoma City, OK, USA, pp. 6-16, 2022, doi: 10.1109/VDS57266.2022.00006.

[8] B. A. Mandour, "Creativity in textile printing design: An integrative framework in design education," International Journal of Technology and Design Education, vol. 35, no. 3, pp. 1031-1062, 2025, doi: 10.1007/s10798-024-09936-z.

[9] N. Lee and S. Suh, "How does digital technology inspire global fashion design trends? Big data analysis on design elements," Applied Sciences, vol. 14, no. 13, pp. 5693, 2024, doi: 10.3390/app14135693.

[10] S. Moon and Y. Chae, "Quantitative analysis of the color accuracy and reproducibility in digital textile printing: Discrepancies within color reproduction media," Textile Research Journal, vol. 95, no. 9-10, pp. 1053-1069, 2025, doi: 10.1177/00405175241288229.

[11] L. Haghzare, X. Ping, M. Arnison, D. Monaghan, D. Karlow, V. Honson, et al., "Digital fabrics for online shopping and fashion design," Frontiers in Virtual Reality, vol. 4, Art. no. 1236095, 2023, doi: 10.3389/frvir.2023.1236095.

[12] S. Y. Jang and J. Ha, "Presenting fabrics in digital environment: fashion designers' perspectives on communicating tactile qualities of the fabrics," Fashion and Textiles, vol. 10, no. 1, pp. 6, 2023, doi: 10.1186/s40691-022-00328-2.

[13] M. Tkalec, M. Glogar, Ž. Penava, F. T. Petra, K. Danjela, and I. Stojanoska, "The Complexity of Colour/Textile Interaction in Digital Printing as an Integral Part of Environmental Design," Arts (2076-0752), vol. 13, no. 1, pp. 29, 2024, doi: 10.3390/arts13010029.

[14] M. X. Vargas, Z. Zhexin, and O. Yoichi, "An exploration in digital techniques to read Ainu textile patterns," Journal of Cultural Heritage Management and Sustainable Development, vol. 15, no. 1, pp. 1-24, 2025, doi: 10.1108/JCHMSD-05-2022-0083.

[15] G. Xu, Y. U. Chen, and J. Hua, "A fast multi-scale textile pattern generation method combining layered loss and convolutional attention," Industria Textila, vol. 76, no. 3, pp. 1-10, 2025, doi: 10.35530/IT.076.03.2024126.

[16] S. Liu, Y. K. Liu, C. Lo, and C. Kan, "Intelligent techniques and optimization algorithms in textile colour management: a systematic review of applications and prediction accuracy," Fashion and Textiles, vol. 11, no. 1, pp. 13, 2024, doi: 10.1186/s40691-024-00375-x.

[17] L. Xu, F. Li, and S. Chang, "A fiber recognition framework based on multi-head attention mechanism," Textile Research Journal, vol. 94, no. 23-24, pp. 2629-2640, 2024, doi: 10.1177/00405175241253307.

[18] N. Skliarenko, I. Gryshchenko, and M. Kolosnichenko, "Symmetry in the visual communication design: Methods of dynamic image construction," Art and Design, vol. 3, no. 4, pp. 9-20, 2021, doi: 10.30857/2617-0272.2021.3.1.

[19] Y. Chen and D. Meng, "Optimization of Visual Communication Design Scheme by Combining Computer-Aided Design and Deep Learning," Computer-Aided Design & Applications, vol. 22, pp. 253-267, 2025, doi: 10.14733/cadaps.2025.S1.253-267.

[20] C. Zhang, "Creation of a system for designing textile patterns using an iterative function system," International Journal of Innovative Research and Scientific Studies, vol. 7, no. 1, pp. 115-26, 2024, doi: 10.53894/ijirss.v7i1.2530.

[21] S. He and J. Lu, "Generation Method of Visual Communication Design Elements Based on Recurrent Neural Network," Computer-Aided Design & Applications, vol. 21, pp. 149-163, 2025, doi: 10.14733/cadaps.2025.S1.149-163.

[22] B. Liu, B. Guo, and L. Jiang, "Image Feature Extraction and Interactive Design of Cultural and Creative Products Based on Deep Learning," Computer-Aided Design & Applications, vol. 21, pp. 148-163, 2024, doi: 10.14733/cadaps.2024.S7.148-163.

[23] J. Bian and Y. Ji, "Research on the teaching of visual communication design based on digital technology," Wireless Communications and

Mobile Computing, vol. 2021, no. 1, Art. no. 8304861, 2021, doi: 10.1155/2021/8304861.

[24] W. Zhu, "A Study of Big-Data-Driven Data Visualization and Visual Communication Design Patterns," Scientific programming, vol. 2021, no. 1, Art. no. 6704937, 2021, doi: 10.1155/2021/6704937.