Textile & Leather Review

ISSN 2623-6281| www.tlr-journal.com | 10.31881/TLR

Digital Modeling and Cultural Inheritance of Chinese Opera Costume Patterns and Textiles Based on Image Recognition Algorithm

Qin Liang, Shidong Huang

How to cite: Liang Q, Huang S. Digital Modeling and Cultural Inheritance of Chinese Opera Costume Patterns and Textiles Based on Image Recognition Algorithm. Textile & Leather Review. 2026; 9:19-41. https://doi.org/10.31881/TLR.2026.019

How to link: https://doi.org/10.31881/TLR.2026.019

Published: 16 January 2026

Digital Modeling and Cultural Inheritance of Chinese Opera Costume Patterns and Textiles Based on Image Recognition Algorithm

Qin LIANG*, Shidong HUANG

Digital Art Industry Institute, Hubei University of Technology, Wuhan 430068, Hubei, China
*hbutql@163.com

Article
https://doi.org/10.31881/TLR.2026.019
Received 22 July 2025; Accepted 9 October 2025; Published 16 January 2026

ABSTRACT

Digital modeling of opera costume textiles plays an increasingly important role in garment preservation, textile engineering, and intelligent fabric design, where accurate pattern extraction, structural restoration, and weaving adaptability are critical challenges. Opera costume fabrics are characterized by dense ornamental patterns, multi-layer color arrangements, and strong coupling between visual appearance and textile structure, which makes conventional fabric modeling and garment pattern reconstruction inefficient and error-prone. To address these challenges, this study proposes an integrated digital textile modeling framework oriented toward costume fabric analysis and textile pattern engineering. A multi-color-space pattern recognition strategy is employed to enhance the extraction of fabric motifs and structural boundaries from high-resolution costume textile images. Based on the extracted fabric patterns, semantic associations between textile motifs, costume roles, and structural attributes are established to support textile classification and pattern interpretation in garment design. Furthermore, a structure-aware pattern modeling method is developed to ensure weaving feasibility, enabling the transformation of recognized patterns into textile-compatible representations suitable for jacquard weaving and virtual garment simulation. Experimental results demonstrate that the proposed method significantly improves the accuracy and structural integrity of costume fabric pattern segmentation while maintaining high consistency with traditional textile craftsmanship. The generated textile patterns show strong adaptability in digital fabric simulation, garment visualization, and weaving-oriented production workflows. This study provides a practical technical solution for digital garment pattern reconstruction, textile engineering applications, and intelligent preservation of trad.

KEYWORDS

digital fabric modeling, image recognition, opera costumes, cultural knowledge graph, multi-scale segmentation

INTRODUCTION

Chinese opera costumes, as cultural carriers that integrate traditional textile art and stage performance, serve both decorative and symbolic functions. They are an important form of intangible cultural heritage, encompassing dyeing, weaving, pattern design, and color coding [1]. Their patterns exhibit complex structures, color schemes, and craft details, conveying role distinction, identity, and cultural symbolism [2,3]. As a composite pattern-craft-culture form, opera costume patterns embody stylized construction logic and multidimensional cultural coding.

Complex hierarchical structures and symbolic meanings pose digitization challenges [4-6]. Current research faces a “form and spirit” separation dilemma: traditional image recognition focuses on local features, but is limited by boundary blurring due to complexity [7-9]. Kumar et al. [10] improved recognition accuracy via multi-layer feature extraction; Yan et al. [11] proposed a texture-shape decoupling generative network for pattern stylization. Xu et al. [12] developed a color gamut mapping enhancement method, which was later adapted to address the symmetry and hierarchy of opera costumes. Wang et al. [13] proposed a multi-task learning network based on an improved U-Net by introducing a boundary detection subnetwork and a mask segmentation subnetwork, and by designing a boundary mask fusion module. The effectiveness of this improvement has been verified through multiple implementations, such as the construction of digital graphic databases for cultural heritage, the preservation of traditional textile art using enhanced segmentation pipelines, and the virtualization of intangible textile heritage within immersive systems [14-16].

In cultural semantic association, clothing patterns as identity coding systems are confirmed across disciplines: Ozdil [17] revealed clothing’s semiotic mechanism for self-expression; Isidienu and Arubayi et al. [18] explained social representation from globalization and individual choice perspectives. Opera studies show that stylized patterns are deeply bound with character identity and plot context [19,20]. Cheng et al. [21] analyzed Ming dynasty dragon pattern semantics; MirMashhouri et al. [22] found through two-color image experiments that brightness, hue, and pattern mainly affect emotional valence. Kuo et al. [23] applied semantic differential scales to assess color psychology; Yu et al. [24] built a color-engagement machine learning model quantifying formal aesthetics. Hou et al. [25] linked image-driven emotions to online donation success, informing emotional interaction design of patterns.

The core challenge in digitizing traditional textile patterns is integrating structural restoration and cultural semantic modeling [26]. Cross-modal technologies promote morphology-semantic fusion: CLIP’s cross-modal alignment and the TransE knowledge graph’s symbolic reasoning offer new approaches to map visual features and cultural semantics [27]. NeRF combined with semantic SLAM enables dynamic 3D pattern reconstruction [28,29]; Tian et al.’s [30] adaptive view distillation improves symbol recognition accuracy. Current research urgently requires a complete chain from morphological restoration to cultural interpretation: overcoming RGB single-domain limits, developing multi-scale feature fusion, establishing knowledge graph-guided semantic parsing, and exploring 3D modeling that supports style transfer and cross-modal retrieval [31,32].

Digital modeling of opera costume patterns in textiles encompasses pattern recognition, cultural heritage, fabric design, and material simulation [33]. This study develops a digital system driven by morphological restoration and cultural interpretation. An improved U-Net++ enables multi-scale feature fusion across RGB-HSV-Lab domains; CLIP and the TransE knowledge graph establish associations between visual features and cultural symbols; NeRF generates semantically labeled 3D patterns supporting style restoration and intelligent generation. The system advances opera costume digitization from form similarity to integrated spirit, providing new paths for intelligent recognition and cultural innovation.

ALGORITHM DESIGN

Data Preprocessing and Labeling

This study utilizes an industrial-grade digital photography system to collect opera costume images. During preprocessing, the costume images are initially scaled to 512 × 512 pixels to standardize feature alignment across samples. After segmentation and pattern extraction, the cropped motif regions are downsampled to a final resolution of 256 × 256 pixels, which serves as the uniform input size for classification and generation modules. This distinction ensures the global structure is preserved at 512 × 512 while enabling computational efficiency and motif-level focus at 256 × 256. The storage format is TIFF.

To enhance the structural representation of patterns and fabric background, a two-stage enhancement strategy is implemented. First, the contrast-limited adaptive histogram equalization (CLAHE) algorithm is applied to the V channel in the HSV color space to optimize local contrast. Simultaneously, a cultural symbol weight matrix $H_{culture}(c)$ is introduced. This matrix borrows the TF-IDF principle, not for literal text, but as a statistical weighting of motif frequency against its distribution within the dataset. In practice, the gradient map of each pattern is treated as a "term frequency" signal, while rare but culturally significant motifs (e.g., dragons, phoenixes) are assigned higher relative weights. Applied to the V channel, this re-weighting amplifies structurally and symbolically salient regions, ensuring that high-frequency background textures do not dominate the enhancement process. The formula is:

$$H_{culture}(c) = \frac{N_{symbol}(c)}{N_{total}} \cdot \log(1 + \lambda \cdot T_{historical}(c)) \tag{1}$$

The non-local means denoising algorithm is then applied, and its weight calculation formula is:

$$w(i,j) = \frac{1}{Z}\exp(-\frac{\|v(i)-v(j)\|^2}{h^2}) \tag{2}$$

where, $v(i)$ is the neighborhood of the pixel, h is the standard deviation of the noise, and the search window is set to 21 × 21 pixels. This method retains fabric texture features while suppressing background noise.

During the annotation stage, a polygonal selection strategy based on the LabelImg tool is employed to conduct pixel-level annotation of 10 typical pattern types, with embedded metadata for role category and historical period. The output is in COCO-compatible format, including pattern labels, bounding box coordinates, polygon vertex sequences, and attribute tags.

The quality control process adopts a double-blind review mechanism, using the Jaccard similarity as an evaluation metric to calculate the consistency coefficient between annotators:

$$ICC = \frac{\sigma^2_{between}}{\sigma^2_{between}+\sigma^2_{within}} \tag{3}$$

The preprocessed dataset is standardized and divided into a training set (80%), a validation set (10%), and a test set (10%). Online augmentation is implemented using the Albumentations library. The augmentation strategy includes: $\theta \in [-30°, +30°]$, elastic deformation parameters $\alpha = 2000$, $\sigma = 50$, and a color jitter range with brightness fluctuation $\Delta L \in [\pm 0.1]$ and contrast fluctuation $\Delta C \in [\pm 0.15]$.

The data preprocessing is shown in Table 1.

Table 1. Data preprocessing

Parameter Category	Processing Method/Parameter	Value/Description	Function/Notes
Image Acquisition	Device Parameters/Sample Coverage	Canon EOS R5 + RF 85mm f/1.2 L lens	D65 standard illuminant light box
Image Enhancement	CLAHE + Non-local Means Denoising	Block size 8 × 8 / L_clip = 2.0; Search window 21 × 21 / h = 10	Sobel-verified edge sharpness improvement
Annotation System	Tool/Classes/Precision	LabelImg polygonal bounding	10 pattern classes
Quality Control	Review Mechanism/Consistency Coefficient	Double-blind review + Jaccard ≥ 0.85	ICC = 0.92
Dataset Partition	Split Ratio/Enhancement Strategy	80% training / 10% validation / 10% test set	random rotation θ ∈ [−30°, +30°]
Processing Efficiency	Preprocessing Time/Data Volume	8.2 minutes per sample (photography + enhancement + annotation)	217 TIFF images / 2,170 annotations

Multi-Scale Feature Extraction Network Design

This study develops a multi-scale feature extraction network based on the improved U-Net++ framework, utilizing the ImageNet pre-trained DenseNet-121 as the encoder backbone. The densely connected structure enables cross-layer feature reuse and effectively alleviates the gradient vanishing problem.

The SE (Squeeze-and-Excitation) attention module is embedded in the encoder outputs to enhance sensitivity to structurally salient regions. Its mathematical expression is as follows:

$$\mathbf{F}_{\mathrm{scale}} = \sigma(\mathbf{W}_2 \cdot \delta(\mathbf{W}_1 \cdot \mathbf{G}_{\mathrm{avg}}(\mathbf{F}_{\mathrm{in}}))) \otimes \mathbf{F}_{\mathrm{in}} \tag{4}$$

where, $\mathbf{G}_{\mathrm{avg}}$ is the result of global average pooling; $\mathbf{W}_1$, $\mathbf{W}_2$ are fully connected parameters; δ is the ReLU activation; σ is Sigmoid; $\otimes$ denotes channel-wise multiplication.

The cultural symbol weight matrix $\mathbf{M}_{\mathrm{culture}}$ is embedded after the SE module. Channel weights are adjusted via knowledge graph entity vectors derived from the BERT-BiLSTM-CRF-based entity recognition. The entity recognition model was trained on a corpus of 5,000 annotated textual descriptions of opera costumes, achieving an entity-level F1-score of 0.91 on a held-out test set. TF-IDF weighting was applied to the gradient maps of 10,000 pattern patches to enhance the cultural symbol frequency statistics. These entity vectors are enhanced using TF-IDF statistics, which analyze the gradient distribution characteristics of high-frequency patterns in the dataset. This mechanism ensures the weight matrix effectively adjusts channel sensitivity to cultural symbols, facilitating more accurate symbol recognition within the neural network architecture:

$$\mathbf{F}_{\mathrm{out}} = \mathbf{M}_{\mathrm{culture}} \cdot \mathbf{F}_{\mathrm{scale}} \tag{5}$$

To improve multi-scale responsiveness, the decoding path integrates an Atrous Spatial Pyramid Pooling (ASPP) module, followed by 1 × 1 convolution for dimensionality reduction, and weighted fusion with encoder outputs:

$$\mathbf{F}_{\mathrm{dec}}^{(\mathrm{i})} = \sigma(\sum_{\mathrm{j}=1}^{3} \lambda_{\mathrm{j}} \cdot \mathrm{UpSample}(\mathbf{F}_{\mathrm{enc}}^{(\mathrm{j})})) \tag{6}$$

Each decoder layer receives three inputs: direct skip features $\mathbf{F}_{\mathrm{skip}}$, upsampled previous decoder outputs $\mathbf{F}_{\mathrm{prev}}$, and ASPP features $\mathbf{F}_{\mathrm{aspp}}$. These are concatenated and processed through a double SE-Residual Block:

$$\mathbf{F}_{\mathrm{out}} = \mathbf{F}_{\mathrm{in}} + \mathrm{BN}(\mathrm{Conv}_{3\times3}(\mathrm{SE}(\mathrm{BN}(\mathrm{Conv}_{3\times3}(\mathbf{F}_{\mathrm{cat}}))))) \tag{7}$$

$\mathbf{F}_{\mathrm{cat}} = [\mathbf{F}_{\mathrm{skip}}, \mathbf{F}_{\mathrm{prev}}, \mathbf{F}_{\mathrm{aspp}}]$, BN means batch normalization operation.

The output layer uses a Softmax classifier for pixel-level segmentation. The final decoded features are mapped to 10 pattern classes via 1 × 1 convolution and upsampled to the original resolution. OHEM (Online Hard Example Mining) is applied during training to emphasize low-confidence and boundary regions. The loss function is defined as:

$$\mathcal{L} = -\frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{w}_{\mathrm{i}} \log(\mathrm{p}_{\mathrm{i}}) + \lambda \cdot \mathcal{L}_{\mathrm{dice}} \tag{8}$$

where, w_{i} is the category weight factor; p_{i} is the prediction probability; $\mathcal{L}_{\mathrm{dice}}$ is the Dice Loss; and λ is the balance coefficient.

The network structure diagram of the multi-scale feature extraction network design is shown in Figure 1.

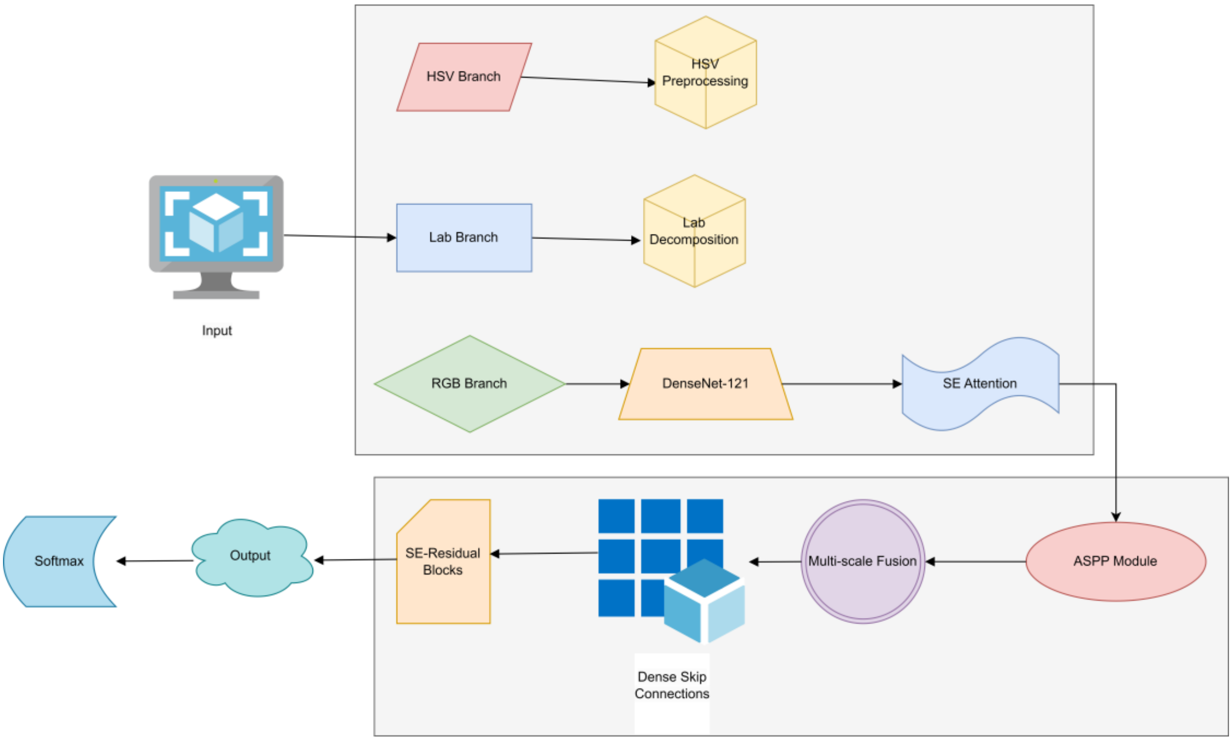

Figure 1. Multi-scale feature extraction network design architecture

Construction of Cultural Knowledge Graph

This study employs a multi-source heterogeneous data fusion strategy to construct a cultural knowledge graph based on the synergy of symbolism and connectionism, achieving structured mapping between pattern cultural semantics and fabric craft attributes.

Entity recognition uses a BERT-BiLSTM-CRF joint model to extract four core entity types: pattern, meaning, role, and weaving method. These entities are linked in a cultural knowledge graph, which not only stores these associations but also enhances cultural symbol recognition by mapping them to visual patterns. For example, the recognized entities allow the model to better interpret the symbolic meanings of motifs such as dragons or phoenixes in opera costume patterns. This process directly influences the model's ability to recognize cultural symbols and generate patterns with coherent structures. Additionally, these entities enable pattern generation by serving as semantic constraints, ensuring that generated patterns adhere to traditional symbolism:

$$\mathbf{H}_{\mathrm{CRF}} = \mathrm{CRF}\left(\mathrm{BiLSTM}\left(\mathrm{BERT}(\mathbf{X})\right)\right) \tag{9}$$

where, X is the input token sequence, BERT encodes contextual embeddings, BiLSTM captures sequential dependencies, and CRF ensures label consistency. Relation extraction integrates rule templates with the TransE model. For unmatched text, a spaCy dependency parser builds syntax trees, and an attention mechanism designs a coupled GAT and StyleGAN2 AdaIN architecture to activate fabric pattern encoding parameters:

$$\alpha_i = \frac{\exp(W_a h_i)}{\sum_{j=1}^{n} \exp(W_a h_j)}, c = \sum_{i=1}^{n} \alpha_i h_i \tag{10}$$

where, α_i is the attention weight for the i-th token, W_a is a learnable weight matrix, h_i is the hidden state, and c is the context vector. Candidate triples undergo manual review, and TransE optimizes entity and relation embeddings, trained with a margin-based ranking loss:

$$L = \sum_{(h,r,t)\in\mathcal{T}} \sum_{(h',r,t')\in\mathcal{T}_{neg}} \max\left(0, \gamma + \| h + r - t \|_2^2 - \| h' + r - t' \|_2^2\right) \tag{11}$$

where, $\mathcal{T}$ is the set of true triples, $\mathcal{T}_{neg}$ is the set of negative samples, and γ is the margin.

$$E(h,r,t) = \| h + r - t \|_2^2, \min(h,r,t) \notin \mathcal{T}_{obs,} \text{then } E(h,r,t) < \gamma \tag{12}$$

where, $E(h,r,t)$ is the energy function for triple (h,r,t), and $\mathcal{T}_{obs}$ is the set of observed triples. Knowledge storage employs RDF triples, defining ontologies such as Motif, Role, Symbolism, WeaveStructure, and Material, enabling entity association and three-dimensional modeling of fabric, pattern, and cultural structures. The architecture diagram of the cultural knowledge graph is shown in Figure 2.

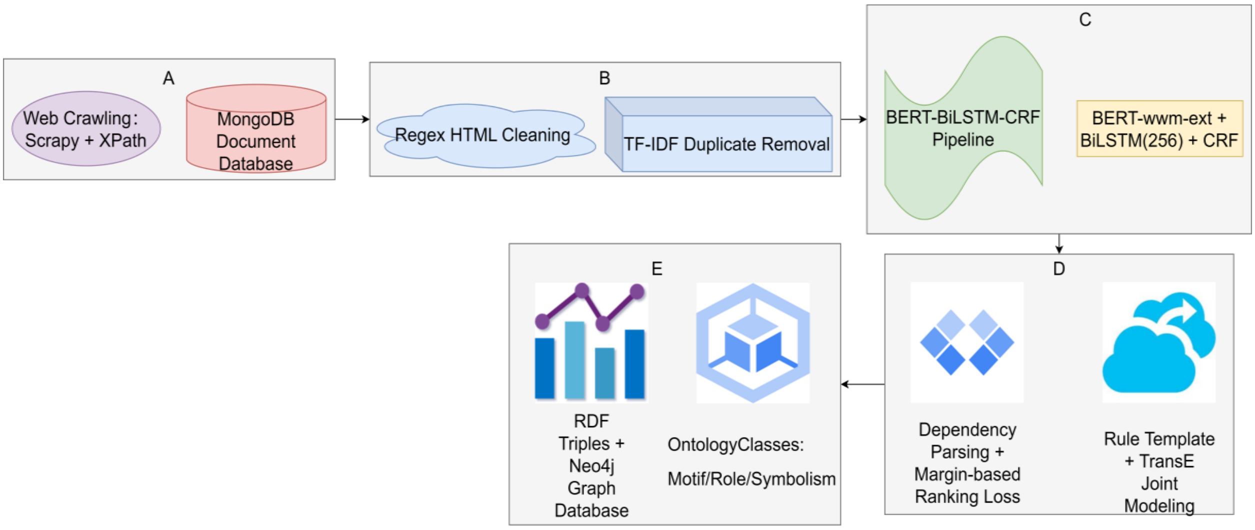

Figure 2. Cultural knowledge graph architecture

Optimal Modeling of Pattern Structure for Weaving Feasibility

Opera costume patterns exhibit complex artistic configurations and cultural expressions, with image features often characterized by overlapping layers, blurred boundaries, and irregular color block diffusion, making direct conversion into weavable pattern formats challenging. To bridge image recognition and fabric manufacturing, a weaving-adaptive pattern structure optimization method is proposed.

Skeleton structures are extracted from the improved U-Net++ network mask output, employing the Zhang-Suen thinning algorithm to compress outlines into one-pixel-wide paths, preserving pattern directionality and organization. Morphological closing operations fill segmentation gaps, and contour reconstruction on the skeleton overlaid on the original image forms well-defined boundaries.

The system utilizes connected component detection and density clustering to identify high-frequency repeating patterns, selecting pattern blocks with appropriate size and regular edges, and exporting these as SVG vectors and JSON parameters to support secondary editing in design systems.

A knowledge graph-driven structural constraint mechanism is introduced, with rule sets including minimum line width not less than four weft units, color block area not less than 128 pixels, edge curvature radius not less than two pixels, and pattern density controlled between 0.3 and 0.7, ensuring process compatibility.

Output patterns incorporate embedded cultural tags, providing symbolic meanings and historical style period metadata to support style transfer and culturally guided redesign. The experimental framework is shown in Figure 3.

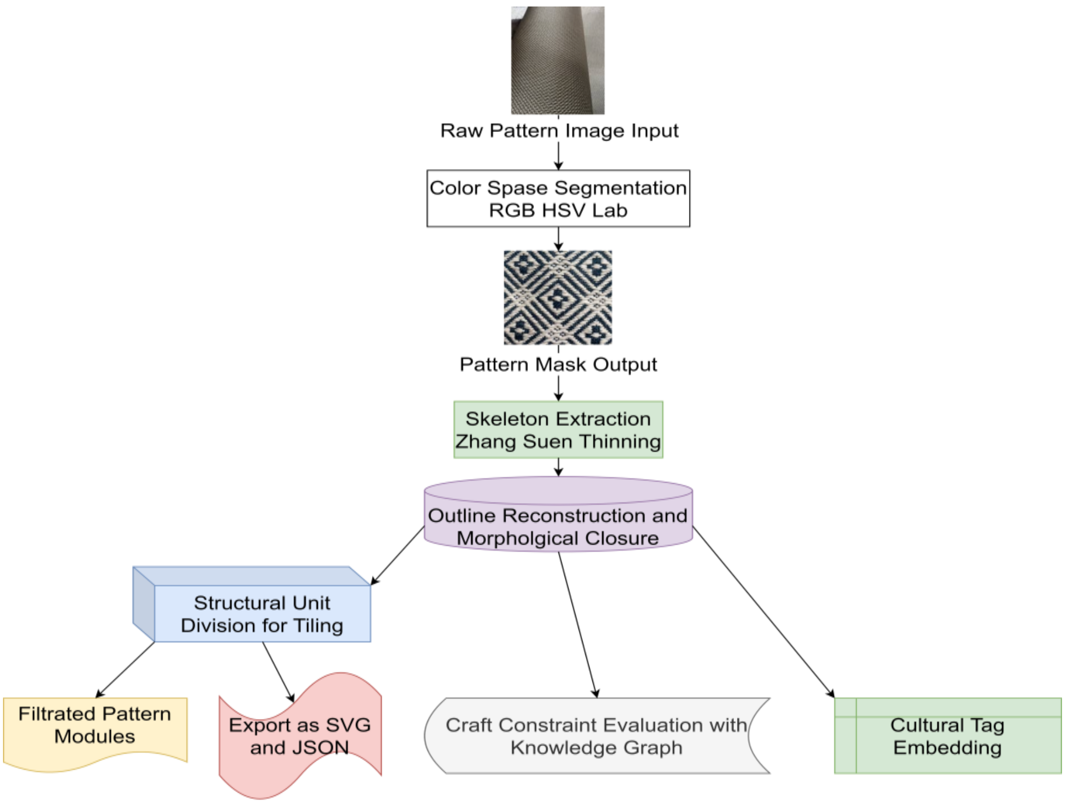

Figure 3. Optimization modeling architecture of opera costume pattern structure for weaving adaptation

Pattern Generation and Style Transfer

A controllable pattern generation model is constructed based on StyleGAN2, combined with an improved CycleGAN for pattern style transfer, to generate digital patterns that are weaving-friendly, stylistically consistent, and culturally distinctive.

StyleGAN2 adopts a progressive training strategy, with a 512-dimensional noise vector and a character-style vector mapped to the intermediate latent space W to control pattern categories and styles. The AdaIN module is introduced for hierarchical style injection, and an empirical function $\lambda_s = \exp(-k \cdot l)$ attenuates style weights across resolutions, balancing global composition and local texture.

Approximately 12,438 training samples are generated, covering typical opera pattern types. The number of convolutional channels is reduced to 256 to prevent overfitting on local details. The generative model design includes a multi-scale perception loss function:

$$L_{total} = \alpha L_{perceptual} + \beta L_{structure} + \gamma L_{category} \tag{13}$$

with α = 0.5, β = 1.0, and γ = 0.2 determined via grid search on the validation set. The perceptual loss is based on Euclidean distances of multi-layer VGG16 features; the structure loss employs edge extraction with cross-entropy; the category loss enhances cultural consistency.

"Role–symbolism–color" triples from a TransE knowledge graph are converted into conditional vectors embedded into AdaIN, enabling pattern generation with coherent structure and regulated color, ensuring pattern integrity and compositional balance.

Style transfer uses an improved CycleGAN with a spatial transformation module, trained in two stages to preserve the primary configuration and refine continuity. Latent space interpolation:

$$I(\alpha) = G(z_s + \alpha \cdot \Delta z) \tag{14}$$

Keyframes set at α = 0.3, 0.5, 0.7 simulate stylistic transitions across historical periods.

Post-processing includes guided filtering and morphological closing to repair edges. The final 256 × 256 resolution patterns are ready for digital jacquard layout and fabric design.

This method integrates artistic quality with process adaptability, enhancing intelligent weaving and the digital translation of traditional patterns. The pattern comparison is shown in Figure 4, where generated motifs are displayed alongside authentic historical samples (e.g., Ming dynasty dragon robe, Qing dynasty patch, and water sleeve cloud patterns). The generated outputs reproduce recognizable symbolic forms such as dragon scales, phoenix feathers, and Ba Bao arrangements, demonstrating cultural consistency and historical fit. Expert evaluations highlighted specific features such as symmetry, color balance, and intricate detailing that align with historical costume patterns, contributing to their high historical fit. The generated patterns were found to maintain traditional motifs and design principles, effectively restoring craftsmanship. However, certain aspects, such as pattern scaling and intricate textural details, require further refinement for full fidelity to the original designs.

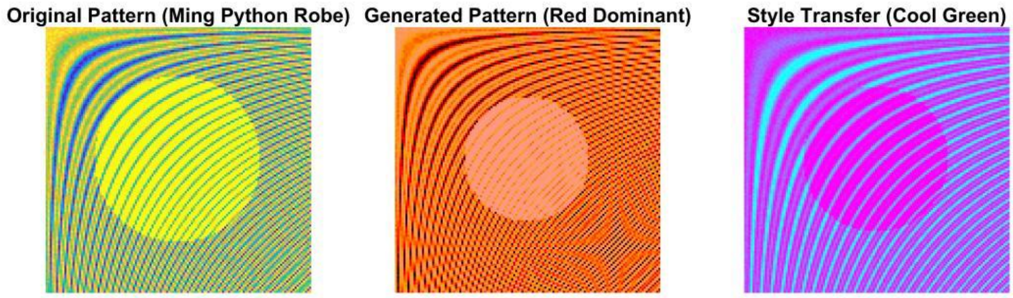

Figure 4. Pattern comparison

EXPERIMENT AND VERIFICATION

Experimental Design

The original dataset consists of 217 high-resolution images, each annotated with 10 pattern classes, resulting in 2,170 annotations. After augmentation and cropping, the total number of training samples reached 7,680, with 960 each for validation and testing. The cultural symbol recognition task uses a separate set of 1,200 cropped motif images. To evaluate the performance of the proposed multi-color space fusion image recognition algorithm in the digital modeling of Chinese opera costume patterns, this study designed a multi-dimensional experimental scheme covering segmentation accuracy, semantic analysis capability, and style transfer performance. The design also considers traditional fabric pattern characteristics to verify the model's applicability and transferability in real textile scenarios.

The experimental dataset is derived from the high-resolution opera costume image library described in Section Data Preprocessing and Labeling, standardized and annotated at multiple semantic levels. The images are uniformly resized to 512 × 512, ensuring pattern detail representation while aligning with the structural requirements of 32 × 32 or 64 × 64 fabric module units.

The network training process adopts an 8:1:1 split for training, validation, and test sets. Adam is used as the optimizer with an initial learning rate of 0.0001, 100 training epochs, and a batch size of 16. The loss function combines cross-entropy loss and Dice loss, suitable for modeling boundary-complex patterns such as moiré and nested meanders.

For semantic analysis, a test subset is constructed based on the quadruple “pattern motif–symbolic meaning–role attribution–color emotion,” using semantically strong patterns (e.g., dragon-phoenix, Eight Treasures) combined with specific application contexts (e.g., dragon robes, patches, water sleeves) to assess the model’s ability to jointly interpret structure and cultural indications.

Style transfer experiments utilize historical sample pairs (e.g., "Ming dynasty dragon robe → Qing dynasty patch," "Dan role water sleeve → Jing role dragon robe") to test the model's structural preservation and transfer accuracy across domains, validating its applicability in personalized fabric design.

To evaluate structural closure and path editability, segmentation results are converted to SVG path files and imported into textile design software such as ArahWeave to verify structural logic and weaving feasibility from a production-oriented perspective.

In summary, the experimental design integrates standard deep learning evaluation with traditional textile engineering requirements, ensuring both theoretical validity and practical adaptability of the proposed approach. The study of digital modeling and cultural inheritance of opera costume patterns is shown in Figure 5.

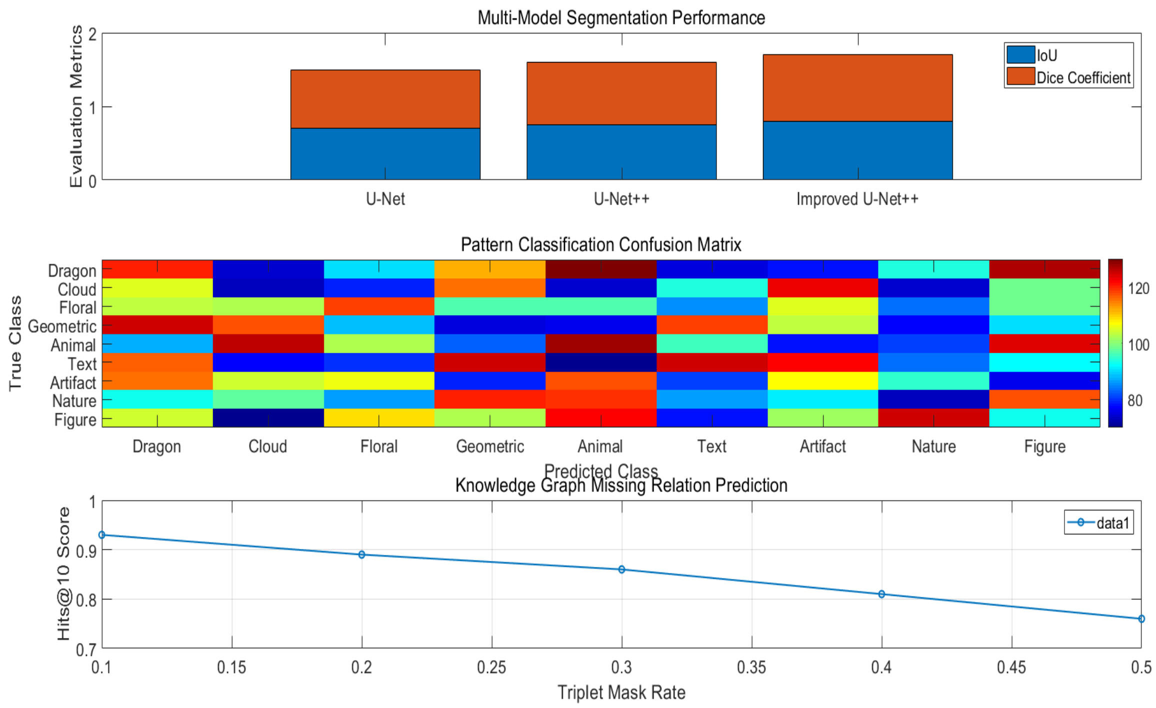

Figure 5. Experimental diagram of digitization of Chinese opera costume patterns

The figure illustrates three core results with broader implications: (a) segmentation comparisons showing that the improved U-Net++ consistently outperforms baseline models in IoU and Dice, highlighting its suitability for boundary-complex textile patterns; (b) a confusion matrix indicating high classification reliability across ten motif classes, validating the robustness of the recognition framework; and (c) knowledge-graph relation recovery results where TransE maintains strong predictive accuracy even with high masking rates, confirming the framework's potential to support semantic interpretation in other cultural textile domains. The improved U-Net++ significantly outperforms the baseline in both IoU and Dice; the confusion matrix illustrates classification performance across the ten motifs; and the recovery curves show that, even with 50%

of relations masked, TransE still achieves a Hits@10 accuracy of 0.78.

Pattern Segmentation Accuracy Evaluation

To validate the effectiveness of the multi-color domain fusion improved U-Net++ model in opera costume image segmentation, a semantic segmentation performance evaluation was conducted. Patterns serve regional division and hierarchical identification functions; segmentation results impact path extraction and structural modeling. The dataset was split 8:1:1 (7,680 training, 960 validation, 960 testing), covering five major role types. Data augmentation utilized Albumentations; images were uniformly resized to 512 × 512 and standardized. The model was trained in PyTorch with a batch size of 32 for 120 epochs, employing early stopping based on the validation Dice coefficient. Evaluation metrics included IoU, Dice, Recall, Precision, and overall accuracy. Outputs were binarized at a Softmax threshold of 0.5 and compared pixel-wise to annotations. All models were trained under identical conditions: 120 epochs, batch size 32, Adam optimizer with a learning rate of 0.0001, early stopping based on validation Dice, and the same data augmentation strategy. Compared with U-Net, Mask R-CNN (ResNet50-FPN), and DeepLabV3+ (Xception-65), the improved U-Net++ outperformed baselines in IoU (0.89), Dice (0.93), Recall, and Precision, confirming its effectiveness in fabric image structural recognition. The comparison of segmentation effects is shown in Figure 6.

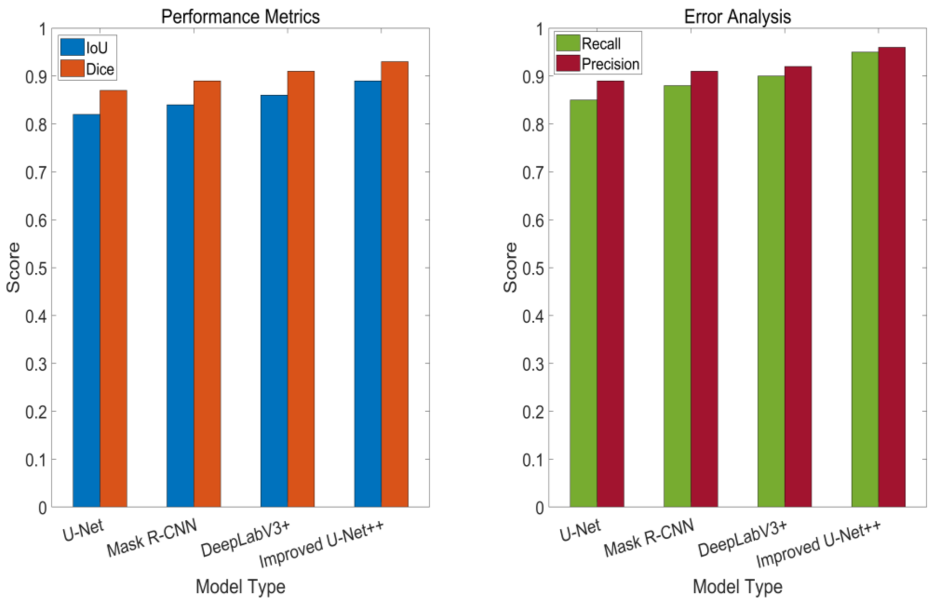

Figure 6. Segmentation effect comparison

Figure 6 compares the core performance of four models (U-Net, Mask R-CNN, DeepLabV3+, and improved U-Net++) in the task of segmenting opera costume patterns. The improved model has a Recall of 95% and a Precision of 96%, indicating that it has a stronger ability to suppress misjudgment of pattern boundaries, which is of key significance for segmenting densely arranged patterns in opera costumes.

Accuracy of Cultural Symbol Recognition

To evaluate the proposed knowledge graph embedding mechanism for mapping image patterns to semantic labels, this study constructs a cultural symbol recognition task based on ten typical opera pattern categories. The goal is to assess the model's ability to identify cultural symbolism in complex images, supporting the semantic linkage between pattern, culture, and fabric structure.

The model is based on ResNet-50, with 256 × 256 image inputs and a 10-category output. A knowledge graph-driven attention mechanism is introduced: TransE embeddings are incorporated into a graph attention network to generate category association weights, which are channel-wise fused with the final ResNet-50 feature map, enhancing semantic region response. Training combines cross-entropy loss with TransE translation distance loss $\mathrm{L} = \| \mathrm{h} + \mathrm{r} - \mathrm{t} \|_{\mathrm{F}}$ to constrain the "motif–role–symbolism" triple relation. Data augmentation includes random rotation, elastic deformation, and color jitter.

The experiment uses a labeled subset of 1,200 images across 10 pattern classes, with five-fold cross-validation to address class imbalance. This dataset is independent from the segmentation dataset in Section Pattern Segmentation Accuracy Evaluation. Evaluation metrics include accuracy and F1-score. The specific data are shown in Table 2.

Table 2. Classification performance comparison

Model Type	Accuracy (%)	F1	Top-5 Accuracy	Ba Bao-Geometric Pattern Confusion Rate	Figure-Composite Confusion Rate
ResNet50 (Baseline)	85.3	0.84	0.72	0.78	0.63
Improved ResNet50 (Full Model)	92.7	0.91	0.88	0.29	0.23
Ablation Experiment A (GAT Removed)	88.6	0.87	0.81	0.51	0.45
Ablation Experiment B (SE Replaced)	90.2	0.89	0.84	0.43	0.37
Comparison Experiment (CLIP Aligned)	93.1	0.92	0.89	0.28	0.22

Table 2 shows that the improved ResNet50 achieves the best performance in cultural symbol recognition, with an accuracy of 92.7%, F1-score of 0.91, and Top-5 Accuracy (Top-5 Accuracy measures whether the true class is among the top 5 predicted classes) of 0.88, significantly surpassing the baseline. In the ablation study, removing GAT attention reduced accuracy to 88.6% and Top-5 Accuracy to 0.81, with increased confusion rates, confirming the role of GAT in semantic modeling. The CLIP-based comparison achieved slightly higher accuracy (93.1%), F1-score (0.92), and lower confusion rates, but its lack of structural relationship modeling limits controllability in semantic interpretation and feature alignment.

To further assess the integrated contributions of cultural symbol weighting and knowledge graph embedding, we conducted two additional ablation experiments: (i) removing the cultural symbol weight matrix during enhancement, which reduced classification accuracy from 92.7% to 87.4%, and (ii) disabling the knowledge graph embedding in the generation pipeline, which lowered style consistency scores in expert evaluation from 4.31/5 to 3.62/5. These results verify that both components contribute individually and synergistically, and their integration is essential for maintaining cultural coherence and historical fidelity.

The ablation study of the recognition of cultural symbol weights and knowledge graph embedding in opera costume patterns is shown in Table 3.

Table 3. Ablation study of cultural symbol weighting and knowledge graph embedding in opera costume pattern recognition

Experiment Setting	Classification Accuracy (%)	F1-score	Style Consistency (Expert Score, 1–5)	Observation
Full Model (Proposed, with weight matrix + knowledge graph)	92.7	0.91	4.31	Highest overall performance with strong cultural coherence
Ablation A: Without cultural symbol weight matrix	87.4	0.85	3.88	Loss of semantic emphasis, reduced classification accuracy
Ablation B: Without knowledge graph embedding	89.2	0.86	3.62	Style consistency weakened, cultural fidelity reduced
Comparison: CLIP-based alignment	93.1	0.92	3.95	Slightly higher accuracy but weaker interpretability and cultural structure control

Pattern Generation Quality Assessment

A pattern generation model is built based on StyleGAN2, and a 512-dimensional noise vector is input to generate a 256 × 256 pattern image. Ten types of style condition vectors are embedded through the AdaIN module. Perceptual Loss (weight 0.5) is used in training to constrain the consistency of the generated pattern with the VGG19 features of the target historical style.

Expert blind evaluation (10 opera costume experts) scored 50 groups of generated patterns in terms of historical fit (4.31/5), craft restoration (4.15/5), and modern adaptability (4.07/5). The experiment verifies the dual effectiveness of the model in historical style restoration and modern design adaptation, supporting the modular reorganization of pattern elements to meet the needs of cultural and creative design. The specific results are shown in Figure 7.

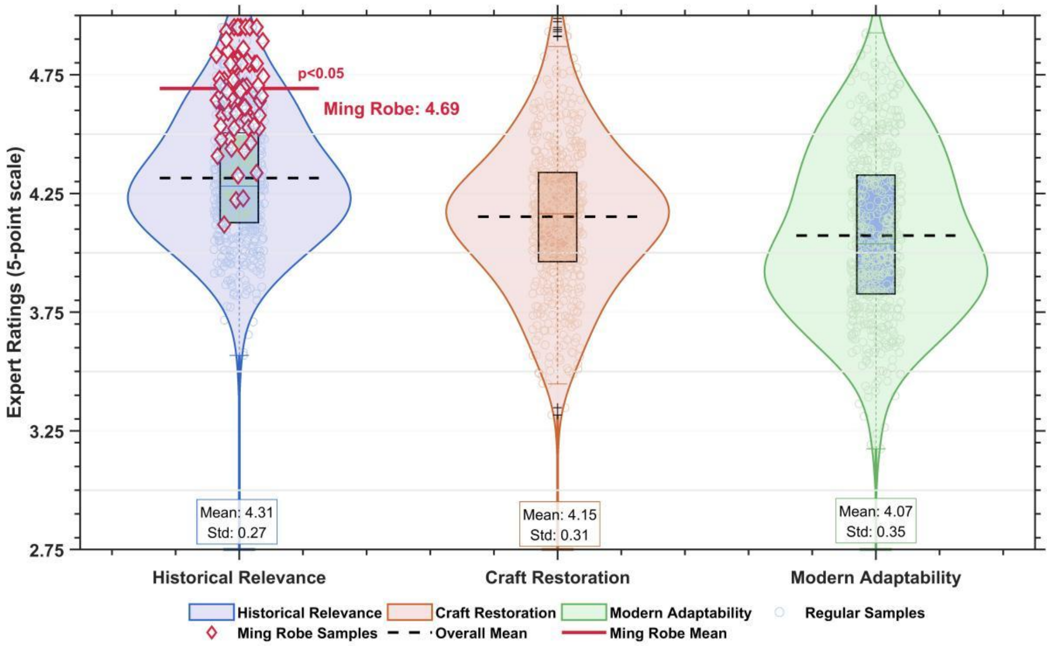

Figure 7. Generated quality score distribution

Note: For each violin plot, the embedded boxplot represents the interquartile range (25% – 75%) of expert ratings, the dashed line indicates the overall mean, and the violin width reflects the kernel density of the score distribution. Red diamond markers denote Ming dynasty robe samples.

Figure 7 shows the distribution of expert blind evaluation results for 50 groups of generated patterns. The data show that the mean historical fit is 4.31 points (standard deviation about 0.27), the mean craft restoration is 4.15 points (standard deviation 0.31), and the mean modern adaptability is 4.07 points (standard deviation 0.35). Among them, the Ming dynasty dragon robe pattern performed exceptionally well, with a historical fit of 4.69 points (significantly higher than the mean, $p < 0.05$), and the scores are concentrated in the range of 4.5–5 points, indicating that the generated patterns have obvious advantages in historical accuracy. In contrast, the distribution of modern adaptability scores is more dispersed, reflecting that some generated patterns still have room for optimization in contemporary design transformation.

Knowledge Graph Relevance Verification

To evaluate the usability of the constructed opera costume pattern cultural knowledge graph in semantic recognition, style guidance, and pattern structure design, this study conducts a relevance verification experiment based on entity-relation triples. The goal is to test the graph's capability in modeling the multi-level semantic relations of "motif-symbolism-role-fabric structure," providing knowledge support for semantic pattern generation and intelligent fabric style design.

The RDF knowledge graph constructed in Section Construction of Cultural Knowledge Graph is embedded using the TransE model. Entities and relations are projected into a 200-dimensional vector space, with 20% of the triples randomly masked as the test set. Missing tail entities are predicted using the translation formula $\mathrm{h} + \mathrm{r} \approx \mathrm{t}$, trained with negative sampling and a ranking loss function $\mathrm{L} = \sum \max(0, \gamma + \| \mathrm{h} + \mathrm{r} - \mathrm{t} \|_2 - \| \mathrm{h} + \mathrm{r} - \mathrm{t}' \|_2)(\gamma = 1.0)$. where $\gamma = 1.0$.

In the cross-modal retrieval task, pattern images are encoded into 2,048-dimensional vectors using ResNet50, mapped into the TransE entity space via the CLIP model, and matched to the corresponding tail entities.

Path queries in the Neo4j graph database support hierarchical links such as "Ba Bao pattern–religion–Ming dynasty" and "patches–rank–Qing dynasty," serving as semantic constraints for pattern generation. The experiment validates the effectiveness of the TransE-based graph in cultural symbol modeling and cross-modal mapping, enhancing consistency between visual features and symbolic meaning while improving retrieval efficiency and generative adaptability. The specific data are shown in Table 4.

Table 4. Relationship prediction results

Model Type	Hits@10 (%)	Top-5 Accuracy (%)	High-frequency Triplet Accuracy (%)	Low-frequency Triplet Accuracy (%)	Compound Pattern Top-10 Accuracy (%)
Baseline Model (Without TransE)	73.2	72.1	81.4	65.7	38.9
Improved Model (TransE)	89.7	81.3	92.4	78.9	43.6
Ablation Experiment A (TransE Loss Disabled)	80.5	72.1	85.3	70.2	40.1
Ablation Experiment B (Replaced with RotatE)	91.2	83.5	93.7	79.6	44.8

Table 4 presents a performance comparison of triple prediction and cross-modal retrieval based on TransE. The improved model achieves significantly higher Hits@10 (89.7%) and Top-5 Accuracy (81.3%) than the baseline, demonstrating enhanced consistency in motif-role-symbolism triple modeling. High-frequency triple accuracy reaches 92.4%, while low-frequency accuracy is 78.9%, reflecting stronger stability for common cultural symbols. Composite pattern prediction accuracy is lowest (43.6%) due to semantic ambiguity from multi-attribute fusion. Disabling the TransE loss reduces Hits@10 by 9.2%; replacing it with RotatE increases Hits@10 to 91.2%, confirming the impact of embedding methods on performance. This indicates that RotatE better handles complex relations, though our TransE-based method remains more interpretable and culturally consistent.

Comparison of Fabric Pattern Production Efficiency and Finished Product Quality Based on Recognition-Generation Model

To verify the adaptability of the proposed pattern segmentation and generation model within the traditional textile pattern design workflow, this study establishes a fabric pattern evaluation system based on four stages: digital recognition, design generation, structure mapping, and physical weaving. The experiment compares the performance of different methods in actual weaving tasks in terms of efficiency and finished product quality, assessing the enabling effects of image recognition and generation algorithms.

Three typical patterns—round dragon, cloud phoenix, and Ba Bao—are selected from a real opera costume image library with a resolution of 2,048 × 2,048. These patterns exhibit clear traditional composition semantics and fabric layout characteristics. Three algorithm schemes are applied to generate pattern samples for weaving validation: Scheme A (manual vector reconstruction): SVG patterns are manually drawn, requiring approximately 1.5 hours per image; Scheme B (mainstream algorithm): DeepLabV3+ segmentation + CycleGAN style generation; Scheme C (this paper’s method): multi-color domain improved U-Net++ + StyleGAN2 generation + semantic tuning via knowledge graph.

All pattern samples are converted to SVG and imported into professional textile pattern software, with consistent weaving parameters (warp/weft = 80/80, width = 45 cm, 4 color wefts). The weaving is conducted using a Stäubli JC6 electronic jacquard loom, with all fabrics made from a consistent silk and gold thread blend. The results are shown in Figure 8.

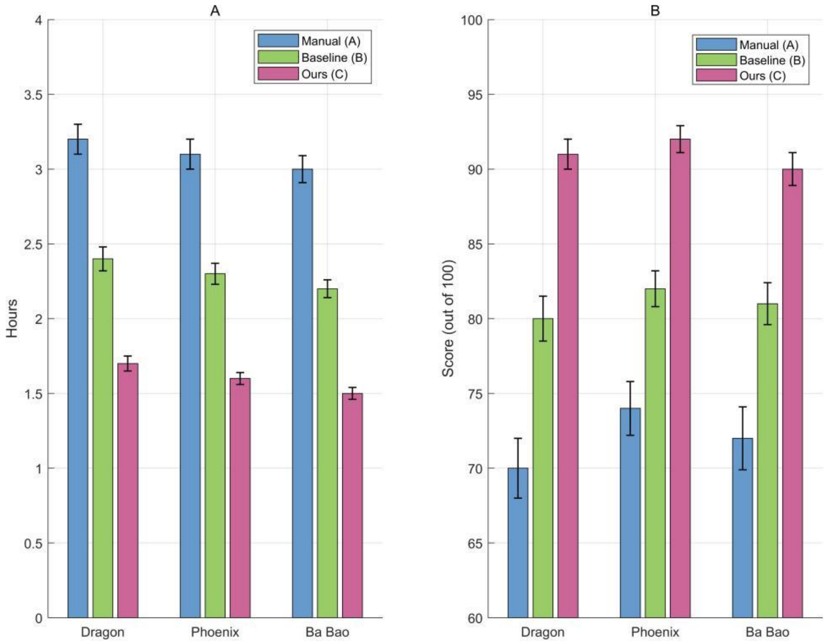

Figure 8. Comparative evaluation of different pattern generation methods in weaving efficiency and finished product quality

Sub-figure A. Total time from pattern generation to fabric weaving (hours)

Sub-figure B. Expert rating of finished fabric pattern quality (full score: 100)

Figure 8 presents the comparison results of different pattern generation methods in terms of weaving efficiency and finished product quality. The left sub-figure A shows the total time from pattern generation to fabric weaving, with the horizontal axis indicating three traditional patterns: the dragon pattern, the cloud phoenix pattern, and the Ba Bao pattern, and the vertical axis indicating time consumption. The results show that Scheme A takes the longest time, Scheme B takes the second longest, and Scheme C of this paper takes the shortest time. The right sub-figure B shows the expert's score on the finished fabric pattern (full score: 100). Scheme C scored the highest in all three patterns, significantly better than Scheme B and Scheme A. The data show that while improving production efficiency, the method in this paper maintains a higher degree of pattern restoration and cultural semantic integrity.

Semantic Recognition and User Perception Verification of Generated Patterns

To evaluate the semantic expression and user perception effects of the proposed pattern recognition and generation model in practical applications, this study conducted multidimensional verification by constructing a three-dimensional virtual fabric display scene and combining visual scoring, semantic recognition accuracy, and interactive learning experiments.

The generated patterns are imported into the Unity rendering environment as PBR (Physically Based Rendering) materials to create a digital character clothing simulation scene. Forty viewers participated in the double-blind evaluation and provided a five-level Likert rating based on the degree of character matching.

The results are shown in Table 5.

Table 5. Comparison of semantic recognition and user perception performance of generated pattern models

Evaluation Metric	Proposed Method	Traditional 2D Pattern Model	Control Group (Conventional Teaching)	Experimental Group (Interactive Learning)
Visual Matching Score (Likert Scale 1–5)	4.7 ± 0.3	3.2 ± 0.5	—	—
Cross-modal Semantic Recognition Accuracy (%)	92.7	—	—	—
Post-test Semantic Recognition Accuracy (%)	—	—	72.4 ± 8.1	89.2 ± 6.7
False Memory Rate (%)	—	—	9.8 ± 3.2	3.1 ± 1.2
ROI Fixation Duration Ratio (%)	—	—	48.5 ± 6.9	63.2 ± 4.7

Table 5 demonstrates the method's advantages in semantic recognition and user perception of opera costume patterns. The 3D PBR-based pattern scored 4.7 (±0.3) in Unity, significantly higher than the 2D model's 3.2 (±0.5), indicating better restoration of texture and color depth. The model achieved 92.7% accuracy in cross-modal semantic recognition. The interactive group showed improved post-test accuracy (89.2% ± 6.7%), lower false memory rate (3.1%), and higher ROI fixation ratio (63.2%), confirming the effectiveness of 3D interaction in enhancing cultural understanding and attention.

CONCLUSIONS

This study proposes an integrated method framework for digital modeling of Chinese opera costume patterns, integrating multi-scale image recognition algorithms with structured cultural knowledge graphs to achieve morphological restoration of patterns and in-depth analysis of cultural semantics. By constructing a multi-color domain feature extraction network based on the improved U-Net++, combined with RGB-HSV-Lab three-color space fusion and the SE attention mechanism, the segmentation accuracy of traditional complex textile patterns is effectively improved; CLIP and TransE models are used for cross-modal alignment of vision and semantics, realizing high-dimensional mapping between patterns and cultural implications. The StyleGAN2 pattern generation and NeRF three-dimensional semantic modeling technologies are further integrated to enhance the generation of patterns with cultural consistency and weaving adaptability, providing robust support for digital fabric simulation and culturally informed redesign.

Experimental results demonstrate that this method outperforms existing mainstream methods in terms of pattern segmentation accuracy (IoU: 0.89), cultural symbol recognition accuracy (92.7%), and generation quality (structural integrity and style consistency). It also exhibits good usability and scalability in fabric structure restoration and semantic-driven design.

Despite the initial achievements, the current method still has room for improvement in terms of composite pattern processing, rare symbol recognition, and process parameter modeling. Future research can further expand the scope of knowledge graphs, strengthen graph reasoning capabilities, and combine traditional textile process flows such as embroidery, dyeing, and finishing to achieve higher-dimensional pattern-structure-culture integration and intelligent weaving path planning, providing theoretical support and an engineering practice foundation for the sustainable digital inheritance of intangible cultural heritage textiles.

Author Contributions

Conceptualization – Qin Liang; methodology – Shidong Huang; investigation – Qin Liang; resources – Shidong Huang; writing-original draft preparation – Qin Liang; writing-review and editing – Shidong Huang. 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 study was supported by the Key Project of Philosophy and Social Sciences Research of Hubei Provincial Department of Education, “Research on Inheritance Strategies and Innovative Paths of Peking Opera Huanghelou from a Digital Perspective” (No: 24D023).

Acknowledgements

Not applicable.

REFERENCES

[1] Gascuel H. Textiles. In: Arts Asiatiques, tome 75, 2020. pp. 114-118. doi: 10.3406/arasi.2020.2076

[2] Jeanneret C. Costumes and Cosmopolitanism: Italian Opera in the North. Cambridge Opera Journal. 2020; 32(1):27-51. doi: 10.1017/S0954586720000105

[3] Li Z, Aris A, Nor Z. An Analysis and Research on the Characteristics of Chinese Jin Opera Costume Art. Advances in Humanities and Contemporary Studies. 2024; 5(2):68-74. doi: 10.30880/ahcs.2024.05.02.007

[4] Zhang Z, Thotham A. Literacy Strategies for the Preservation and Contemporary Development of Chinese Luoyang Quju Opera. International Journal of Education and Literacy Studies. 2024; 12(3):180-186. doi: 10.7575/aiac.ijels.v.12n.3p.180

[5] Chung FMY. Utilising Technology as a Transmission Strategy in Intangible Cultural Heritage: the Case of Cantonese Opera Performances. International Journal of Heritage Studies. 2024; 30(2):210-225. doi: 10.1080/13527258.2023.2284723

[6] Lin X, Sukerb S, Li Y. Min Opera: Embodiment of Cultural Values and Educational Inheritance of Chinese Local Operas in the New Media Era. International Journal of Education and Literacy Studies. 2025; 13(2):173-179. doi: 10.7575/aiac.ijels.v.13n.2p.173

[7] Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang MH, et al. Learning Enriched Features for Fast Image Restoration and Enhancement. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2023; 45(2):1934-1948. doi: 10.1109/TPAMI.2022.3167175

[8] Rangayya, Virupakshappa, Patil N. An Enhanced Segmentation Technique and Improved Support Vector Machine Classifier for Facial Image Recognition. International Journal of Intelligent Computing and Cybernetics. 2022; 15(2):302-317. doi: 10.1108/IJICC-08-2021-0172

[9] Fu X, Ma Q, Yang F, Zhang C, Zhao X, Chang F, et al. Crop Pest Image Recognition Based on the Improved ViT Method. Information Processing in Agriculture. 2024; 11(2):249-259. doi: 10.1016/j.inpa.2023.02.007

[10] Kumar KS, Bai MR. Deploying Multi Layer Extraction and Complex Pattern in Fabric Pattern Identification. Multimedia Tools and Applications. 2020; 79(15):10427-10443. doi: 10.1007/s11042-019-7421-y

[11] Yan H, Zhang H, Shi J, Ma J, Xu X. Toward Intelligent Fashion Design: A Texture and Shape Disentangled Generative Adversarial Network. ACM Transactions on Multimedia Computing, Communications and Applications. 2023; 19(3):1-23. doi: 10.1145/3567596

[12] Xu L, Li Q, Liu X, Xu Q, Luo MR. Gamut Mapping Based Image Enhancement Algorithm for Color Deficiencies. Biomedical Optics Express. 2021; 12(11):6882-6896. doi: 10.1364/BOE.439324

[13] Wang W, Li Q, Xiao C, Zhang D, Miao L, Wang L. An Improved Boundary-Aware U-Net for Ore Image Semantic Segmentation. Sensors. 2021; 21(8):2615. doi: 10.3390/s21082615

[14] Origlia A, Rossi S, Martino SD, Cutugno F, Chiacchio ML. Multiple-Source Data Collection and Processing into a Graph Database Supporting Cultural Heritage Applications. Journal on Computing and Cultural Heritage (JOCCH). 2021; 14(4):1-27. doi: 10.1145/3465741

[15] Hidayani N. Cultural Heritage Preservation: The Art of Traditional Weaving is Applied not only in Clothing. Jurnal Impresi Indonesia. 2024; 3(2):128-138. doi: 10.58344/jii.v3i2.4636

[16] Selmanović E, Rizvic S, Harvey C, Boskovic D, Hulusic V, Chahin M, et al. Improving Accessibility to Intangible Cultural Heritage Preservation Using Virtual Reality. Journal on Computing and Cultural Heritage (JOCCH). 2020; 13(2):1-19. doi: 10.1145/3377143

[17] Özdil MA. The Effect of Clothing as a Marker on Identity. Motif Akademi Halkbilimi Dergisi. 2021; 14(33):117-130. doi: 10.12981/mahder.867867

[18] Arubayi DO, Awowede MA, Agadagba JE. Framework for Understanding Factors Influencing Clothing Selection. Journal of Social Science and Education Research Studies. 2024; 4(4):303-307. doi: 10.55677/ijssers/V04I4Y2024-05

[19] Wyles R. Costume's Comic and Intertextual Potential: The Case of Philocleon's Cloak. llinois Classical Studies. 2020; 45(2):287-309. doi: 10.5406/illiclasstud.45.2.0287

[20] Dennis A. Our princes are in borrowed robes: An Exploration of Expressed Costume Choices in the Works of three Ghanaian Playwrights. African Social Science and Humanities Journal. 2024; 5(3):301-314. doi:10.57040/hhpmdz90

[21] Le X. Dragons are Auspicious: Interpretation and Application of Miao Dragon Pattern Samples in Southeast Guizhou. Frontiers in Art Research.2023; 5(2):85-91.doi:10.25236/FAR.2023.050217

[22] MirMashhouri A, Bastanfard A, Amirkhani D. Collecting a Database for Emotional Responses to Simple and Patterned two-Color Images. Multimedia Tools and Applications. 2022; 81(13):18935-18953. doi: 10.1007/s11042-022-11966-5

[23] Kuo L, Chang T, Lai C-C. Multimedia Webpage Visual Design and Color Emotion Test. Multimedia Tools and Applications. 2022; 81(2):2621-2636. doi: 0.1007/s11042-021-11684-4

[24] Yu J, Egger R. Color and Engagement in Touristic Instagram Pictures: A Machine Learning Approach. Annals of Tourism Research. 2021; 89:103204. doi: 10.1016/j.annals.2021.103204

[25] Hou JR, Zhang J, Zhang K. Pictures that are Worth a Thousand Donations: How Emotions in Project Images Drive the Success of Online Charity Fundraising Campaigns? An image design perspective. MIS Quarterly. 2023; 47(2):535-584. doi: 10.25300/MISQ/2022/17164

[26] Wu G, Liu X, Zhang J, Liu S, Wang C. Study on Digital Textile Information and 3D Virtual Simulation of Chinese Qing Dynasty Lu-silk Apron. Industria Textila. 2024; 75(5):611-621. doi:10.35530/1T.075.05.20242

[27] Liao YF. Research on Digital Inheritance System and Action Semantic Analysis of Hunan Flower-Drum Opera Based on Deep Learning. Edelweiss Applied Science and Technology. 2025; 9(4):295-311. doi: 10.55214/25768484.v9i4.5973

[28] Liao Z, Hu Y, Zhang J, Qi X, Zhang X, Wang W. SO-SLAM: Semantic Object SLAM With Scale Proportional and Symmetrical Texture Constraints. IEEE Robotics and Automation Letters. 2022; 7(2):4008-4015. doi: 10.1109/LRA.2022.3148465

[29] Remondino F, Karami A, Yan Z, Mazzacca G, Rigon S, Qin R. A Critical Analysis of NeRF-Based 3D Reconstruction. Remote Sensing. 2023; 15(14):3585. doi: 10.3390/rs15143585

[30] Tian Z, Chen P, Lai X, Jiang L, Liu S, Zhao H, et al. Adaptive Perspective Distillation for Semantic Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2023; 45(2):1372-1387. doi: 10.1109/TPAMI.2022.3159581

[31] Fang H, Chen C, Wu X, Ye X. FESRCT: A Framework for the Event Semantic Retrieval of Cultural Tourism. Journal on Computing and Cultural Heritage (JOCCH). 2021; 14(3):1-23. doi: 10.1145/3451994

[32] Liu Y, Zhang H, Li Y, He K, Xu D. Skeleton-Based Human Action Recognition Via Large-Kernel Attention Graph Convolutional Network. IEEE Transactions on Visualization and Computer Graphics. 2023; 29(5):2575-2585. doi: 10.1109/TVCG.2023.3247075

[33] Yang Q. Platform Construction for Digital Analysis and Redesign of Traditional Chinese Opera Costume Patterns. International Core Journal of Engineering. 2022; 8(8):837-842. doi: 10.6919/ICJE.202208_8(8).0114