



## OPEN Traditional clothing pattern extraction considering attention mechanism and image data enhancement processing

Dandan He✉, Bing Xia & Hong Li

The three core challenges faced in traditional clothing pattern extraction are poor image quality, complex texture, and scarcity of labeled data. To solve these problems, the study proposes a multi-scale data augmentation with polarized self-attention for traditional costume pattern extraction. The study designs a lightweight segmentation network mainly by fusing the multi-scale enhancement strategy of wavelet transform and generative adversarial network, while incorporating the channel-space concurrent attention module. To address the three core challenges of poor image quality, complex textures, and scarce labeled data in traditional clothing pattern extraction, the proposed MSDA-PSA method introduces three targeted innovations: (1) a multi-scale data augmentation (MSDA) strategy that combats image degradation by integrating wavelet transforms and generative adversarial networks to enhance contrast and preserve critical texture details; (2) a dual attention mechanism employing polarized self-attention (PSA) to resolve complex texture ambiguities by jointly focusing on discriminative color channels and spatial details, thereby improving pattern localization and reducing intra-class variation; and (3) a lightweight network design that mitigates data scarcity through depth-separable convolutions and adaptive dropout, ensuring high parameter efficiency and adaptability to few-shot learning scenarios. The results of the study showed that the conventional clothing pattern extraction method proposed by the study performed well in several key metrics. The method achieved 89.7% mean intersection over union (mIoU) and 0.87 Boundary F-score on the standard test set. This performance significantly outperforms mainstream models such as DeepLabv3+, with the mIoU exceeding it by over 10%. The method's distinct advantage is further demonstrated by its exceptional balance between small-sample adaptability and cross-cultural generalization. It maintained a 72.4% IoU with only 10% of the training data, while achieving 83% accuracy in cross-cultural generalization tests. These results confirm the method's robust capability to deliver high-precision segmentation in both data-scarce and culturally diverse scenarios, a balance that existing methods have not achieved. The real-time processing performance reached 100FPS, and the 16-channel concurrent latency was only 35.4ms. Under extreme degradation conditions ( $\sigma=0.2$  Gaussian noise), noise robustness IoU still reached 0.82, and color retention ability under extreme degradation conditions was controlled below 15%. The experiments also confirmed that the four-scale pyramid architecture improved the feature consistency score to 0.85, and the dual attention module increased the critical region intersection and merger ratio to 83%, while keeping the computational overhead at 2.9 GFLOPS. This research establishes a high-precision and interpretable clothing pattern segmentation framework through hierarchical feature learning and robust enhancement techniques. It provides a solution for digitizing cultural heritage that balances computational efficiency and cultural specificity.

**Keywords** Traditional dress pattern, Attention mechanism, Image data enhancement, Feature extraction, Deep learning

As an important intangible cultural heritage, the unique cultural value of traditional clothing patterns is facing an urgent need for digital protection<sup>1,2</sup>. However, due to the rich colors and complex textures of the patterns

School of Information Engineering, Pingdingshan University, Pingdingshan 467000, China. ✉email: 2650@pdsu.edu.cn

themselves, computer vision methods face significant difficulties in segmentation and extraction<sup>3,4</sup>. The current research faces three core challenges: difficulty in ensuring feature extraction quality for low definition historical images, significant intra class differences in similar patterns that increase recognition difficulty, and scarcity of professional annotation data that restricts the application of supervised learning methods<sup>5,6</sup>. To address these challenges, the proposed MSDA-PSA method is designed with a three-part solution directly targeting each problem: First, to overcome poor image quality, a multi-scale data augmentation (MSDA) strategy is introduced, integrating wavelet transforms and generative adversarial networks to specifically enhance image clarity and preserve texture in degraded historical images. Second, to handle complex textures and significant intra-class differences, a dual attention mechanism utilizing polarized self-attention (PSA) is developed, focusing concurrently on color channels and spatial details to precisely localize embroidery patterns and improve feature discrimination. Third, to circumvent the scarcity of professional annotation data, a lightweight network design is employed, leveraging depth-separable convolutions and adaptive dropout to reduce parameter count and enable effective few-shot learning, thus overcoming the limitations of scarce annotated data. The practical urgency of addressing the three core challenges is profoundly evident in the digitization of specific cultural heritage items, such as Miao embroidery and Dong brocade. For Miao embroidery, historical images often suffer from severe degradation including fading and blurring due to aging and improper storage conditions. This degradation directly compromises the reliability of feature extraction for intricate embroidery stitches and the characteristic indigo dyes, which are central to its cultural identity. Concurrently, the intricate and highly repetitive geometric textures of Dong brocade introduce significant intra-class variations, posing a substantial challenge for pattern recognition algorithms and increasing the risk of misclassification. Furthermore, the acute scarcity of high-quality, expert-annotated datasets for these specific ethnic patterns severely constrains the application of data-driven deep learning methods, ultimately hindering the effective preservation of cultural diversity. These concrete cases underscore the critical necessity and innovation of the proposed MSDA-PSA method. In response to these image quality issues, researchers have begun to explore solutions for image enhancement techniques, improving input data quality through improved preprocessing methods<sup>7,8</sup>.

Lu F et al. proposed an adaptive enhancement method based on filtering and variational decomposition to address the issues of low contrast and random noise in computed tomography images. This study decomposed images into noise layers, texture layers, and structural layers using median filtering and total variation models. Guided filtering was used to extract residual details and fuse texture features, combined with adaptive enhancement factors and local mean gamma correction to optimize contrast. Experiments revealed that this method could effectively balance the relationship between denoising, contrast enhancement, and detail preservation<sup>9</sup>. Li Y et al. proposed an enhanced algorithm combining Retinex theory and guided filtering to address the issue of low contrast in low light environments. This study optimized brightness channel weights using a bimodal energy function. It extracted the illumination component via edge-preserving guided filtering and estimated the reflection component in hyperbolic tangent space. These steps were taken to achieve adaptive brightness and contrast adjustment. Experiments revealed that this method could effectively enhance edges and suppress dark area noise<sup>10</sup>. Jebadass et al. proposed an enhancement algorithm based on intuitionistic fuzzy sets to address the issues of low contrast and blurry details in low light color images. This study used the Yager generation function to calculate non membership degrees and combined entropy optimization to improve image quality. Experiments revealed that this method outperformed traditional methods, such as histogram equalization and contrast-limited adaptive histogram equalization, in terms of entropy and structural similarity metrics. This method could also effectively preserve complex texture details<sup>11</sup>. Chen et al. proposed an enhancement algorithm that combined Retinex with weighted illumination guided filtering to address the issues of halo artifacts, loss of edge details, and noise amplification in low light image enhancement. This study calculated atmospheric light values and transmittance by combining light and dark channels. It also optimized transmittance estimation using weighted guided filtering in the illumination gradient domain. This approach effectively solved the problems of local depth differences and overexposure caused by traditional dark channel prior methods. The experimental results revealed that this method had significant improvements in denoising, halo elimination, brightness adjustment, and edge preservation<sup>12</sup>. However, these traditional enhancement techniques have limitations in traditional clothing pattern extraction scenarios. For example, Retinex-based methods can easily distort the characteristic dye colors of ethnic clothing, such as Miao indigo, due to their reliance on artificial parameters and poor adaptability to complex lighting conditions, often leading to color distortion and texture loss. Similarly, although deep learning-based segmentation models like U-Net variants improve accuracy, they struggle to balance multi-scale texture capture and computational efficiency for complex embroidery patterns, often resulting in detail loss or high computational costs due to fixed receptive fields, which limits their practicality. The proposed MSDA-PSA method directly addresses these limitations. To overcome color distortion caused by the global adjustments of Retinex-based methods in RGB space, it employs HSV color space fine-tuning. This allows for independent manipulation of color attributes—specifically enhancing hue (H:1.2) and saturation (S:1.1) to intensify characteristic dyes like Miao indigo, while separately controlling brightness (V:0.9) to prevent overexposure and preserve fidelity. To tackle the limited receptive fields of U-Net variants, it introduces a four-level pyramid architecture. This structure provides multi-scale feature maps (via 2x, 4x, 8x downsampling), effectively expanding the network's receptive field to simultaneously capture microscopic embroidery stitches and macroscopic pattern layouts, thus resolving the detail loss inherent in single-scale approaches. Furthermore, the integration of depth-separable convolutions and an adaptive dropout strategy ensures computational efficiency and model stability, maintaining a real-time performance of 100 FPS while achieving high accuracy.

The innovations of the proposed MSDA-PSA method are: (1) a multi-scale data augmentation (MSDA) strategy integrating wavelet transforms and generative adversarial networks to enhance image clarity while preserving critical texture details; (2) a dual attention mechanism employing polarized self-attention (PSA) to

jointly focus on discriminative color channels and spatial details, improving pattern localization and reducing intra-class variation; and (3) a lightweight network design using depth-separable convolutions and adaptive dropout, ensuring high parameter efficiency and adaptability to few-shot learning scenarios. These innovations collectively enable high-precision segmentation with real-time performance (100 FPS), robust noise resistance, and strong cross-cultural generalization, effectively addressing challenges such as color distortion, limited receptive fields, and data scarcity in traditional clothing pattern extraction.

Traditional image enhancement techniques have several drawbacks. They are not adaptable to complex lighting, they depend heavily on artificial parameters, and they have difficulty simultaneously optimizing denoising and detail preservation. These issues can easily lead to texture loss or color distortion when processing clothing images<sup>13,14</sup>. With the development of deep learning, image segmentation methods based on convolutional neural networks (CNNs) and attention mechanisms have significantly improved the accuracy of clothing pattern extraction. In recent years, the development of lightweight CNNs, such as the efficient architecture presented in<sup>15,16</sup>, has achieved high-efficiency feature extraction on mobile terminals and embedded devices, thereby making the real-time processing of clothing patterns a practical possibility. Ning et al. proposed a hybrid multi-scale Transformer U-shaped network (U-Net) model to address the issue of insufficient accuracy in semantic segmentation of ethnic clothing patterns caused by complex textures, details, and background interference. The experiment revealed that the model achieved an average Dice score of 89.80% on the ethnic clothing pattern dataset, with more complete edge preservation and fewer misclassifications. Compared with mainstream models such as deep lab version 3 Plus (DeepLabv3+) and residual U-Net, it improved by 0.67%–7.72%, significantly optimizing the fine segmentation effect of complex patterned clothing<sup>17</sup>. Chen R et al. proposed an improved U-Net model based on a visual geometry group 16 layer deep convolutional neural network (VGG16 based U-Net, VGG16 UNet) to address the issues of insufficient semantic labels, poor local segmentation accuracy, and rough edges in clothing effect image segmentation. Experiments revealed that the efficient feature U-Net improved mean intersection over union (mIoU) and average pixel accuracy metrics by 4.91% and 4.98% respectively compared to the original VGG16 UNet, outperforming mainstream models such as fully convolutional network (FCN) and segmentation networks. This method significantly improved the fine segmentation effect of clothing images, especially in edge processing<sup>18</sup>. Liu et al. proposed a comprehensive solution to the problem of restoring clothing sewing patterns from daily photos. The experiment proved that this method could accurately restore the sewing structure of clothing from randomly taken photos. The authenticity of the synthesized dataset and the generalization ability of the model provided reliable support for applications such as fashion design and virtual try on. The relevant code and dataset could be open sourced and shared<sup>19</sup>. Yu F et al. proposed a phase contour enhancement network to address the difficulties in identifying fine-grained widgets and the confusion of similar categories in clothing analysis tasks. Experiments on mainstream datasets, such as fashion analysis and segmentation, demonstrated that this method effectively distinguished between small clothing components and similar categories. It surpassed existing optimal methods in clothing and extended human body analysis tasks while achieving a balance between accuracy and efficiency<sup>20</sup>.

In summary, traditional clothing pattern extraction faces three core challenges: difficulty in feature extraction due to low-quality images, loss of details due to complex textures, and constraints on supervised learning due to scarce annotated data. The practical urgency is underscored by cases such as Miao embroidery and Dong brocade, for which digital preservation is critical to cultural continuity yet is hindered by technical barriers. Traditional enhancement techniques are prone to damaging texture structures, while conventional segmentation networks struggle to balance multi-scale features and long-range dependencies. A research proposal has been put forward to address the above issues multi-scale data augmentation with polarized self-attention for traditional costume pattern extraction (MSDA-PSA). The study first constructs a multi-scale enhancement strategy based on wavelet transform and generative adversarial network, which improves data diversity while maintaining the topological integrity of clothing textures. Based on this, a channel space parallel attention module is designed to improve edge feature extraction capabilities through PSA. It is combined with a convolutional block attention module (CBAM) to optimize feature selection in key regions. The study innovatively integrates wavelet transforms and generative adversarial networks to construct a multi-scale enhancement strategy. It designs a channel-space-parallel PSA to enhance the extraction of edge features while maintaining the integrity of clothing textures. This effectively solves the generalization problem of traditional small-sample clothing pattern segmentation. The research aims to establish an automated extraction framework for ethnic minority costumes. This framework uses hierarchical feature learning and robust enhancement techniques to provide high-precision, interpretable segmentation methods for cultural heritage digitization. The framework also solves the problem of model generalization in scenarios with small samples.

## Methods and materials

### Multi-scale data augmentation and feature extraction methods

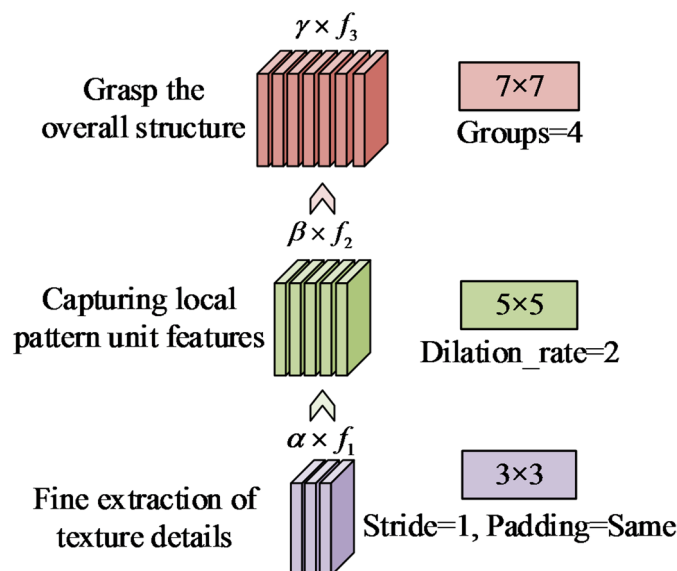
As a precious cultural heritage, the digital protection of traditional clothing patterns faces severe challenges<sup>21,22</sup>. Due to the influence of age and environmental factors, a large number of clothing images have problems such as fading, blurring, and texture degradation. This seriously restricts the subsequent pattern analysis and feature extraction work<sup>23,24</sup>. A systematic data augmentation solution is proposed to overcome the limitations of traditional enhancement methods when balancing texture retention and detail recovery. First, a standardized preprocessing data process is established to ensure consistent input by uniformly adjusting the resolution of each image to  $512 \times 512$ . On this basis, the Keras image data generator is used to realize multi-dimensional geometric transformation enhancement: including  $10^\circ$  random rotation to simulate multi-view shooting,  $\pm 2\%$  horizontal-vertical translation to enhance spatial invariance, and  $\pm 2\%$  shear transformation to improve the adaptability to fabric deformation. Keras is used for data augmentation due to its efficient ImageDataGenerator. However, the preprocessed data is converted to PyTorch tensors to ensure compatibility with the PyTorch-

based network implementation. The study employs a hybrid framework to leverage the distinct advantages of both Keras and PyTorch. Keras, with its highly efficient and user-friendly ImageDataGenerator, is utilized for the data augmentation pipeline to perform multi-dimensional geometric transformations (e.g., rotation, translation, shear). To ensure seamless compatibility with the PyTorch-based network training and inference environment, the augmented image arrays generated by Keras are immediately converted into PyTorch tensors using a dedicated data bridging interface. This interface is implemented through `torch.from_numpy()` function, which directly converts the NumPy array output from Keras into a PyTorch tensor, ensuring data type and shape consistency. This design maintains a fully consistent and reproducible pipeline from preprocessing to model training, effectively avoiding any framework-induced inconsistencies. The hybrid strategy capitalizes on Keras's rapid augmentation capabilities while utilizing PyTorch's flexibility for dynamic graph construction and advanced model design. This hybrid approach leverages the strengths of both frameworks. Keras is used for rapid data augmentation, and PyTorch is used for flexible model training. This setup avoids pipeline inconsistencies between preprocessing and training. These transformation operations not only expand the sample diversity, but more importantly, preserve the key features of the clothing pattern intact. Rotation and translation enable the model to learn viewpoint-independent stabilization features, which are especially beneficial for recognizing symmetric patterns such as Miao embroidery. Shear transformation, on the other hand, enhances the robustness of the algorithm to fold deformation<sup>25,26</sup>. All the transformations adopt the nearest neighbor filling mode, which effectively protects the pattern edge details. On the basis of preprocessing, the study further develops a multi-scale feature preservation technique. The hierarchical features of the clothing pattern are systematically captured by constructing a hierarchical feature extraction framework. This hierarchical feature extraction framework is shown in Fig. 1.

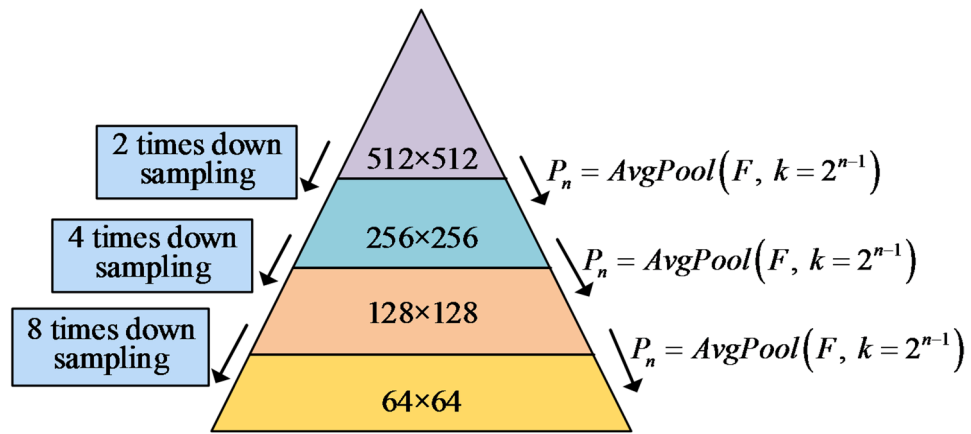
In Fig. 1, the bottom layer uses  $3 \times 3$  small convolutional kernels to finely extract microstructures such as embroidery stitches. The middle layer utilizes a  $5 \times 5$  convolutional kernel to capture the local pattern units. The high-level layer captures the overall pattern layout with a  $7 \times 7$  large convolution kernel. The features at each level are adaptively fused by learnable weight parameters, achieving a comprehensive characterization of multi-level structures such as batik. This multi-scale fusion mechanism preserves the fine texture and understands the overall composition, providing an accurate feature representation for subsequent segmentation recognition. The features at each level are adaptively fused by learnable weighting parameters. Among them, it is assumed that  $f_1$ ,  $f_2$  and  $f_3$  represent the feature maps extracted from the bottom, middle, and high-level layers, respectively, and  $\alpha$ ,  $\beta$  and  $\gamma$  denote the learnable weight parameters corresponding to the contributions of features at different levels. The symbol  $\oplus$  indicates the feature concatenation operation. The fusion formula is shown in Eq. (1).

$$H = \alpha \times f_1 \oplus \beta \times f_2 \oplus \gamma \times f_3 \quad (1)$$

In Eq. (1),  $H$  is the result of multi-scale feature fusion. Although the hierarchical feature extraction effectively captures the local texture features of the clothing pattern, it still faces two key limitations. First, it is difficult to consider the synergistic expression of micro-detail and macro-layout for single-scale features. Second, the fixed sensory field leads to an insufficient long-range dependent model of the complex pattern. To further optimize the performance, the study introduces a four-level pyramid processing architecture. A hierarchical multi-scale analysis framework is used to achieve comprehensive capture of traditional clothing pattern features. The architecture adopts a progressive downsampling strategy, starting from the original high-resolution image and gradually constructing a multi-scale feature representation system<sup>27</sup>. The four-level pyramid processing architecture is shown in Fig. 2.



**Fig. 1.** Hierarchical feature extraction framework.



**Fig. 2.** Four-level pyramid processing architecture.

In Fig. 2, a multi-scale feature representation system is constructed by 2x, 4x, and 8x progressive downsampling from the original  $512 \times 512$  image. Each level of pyramid adopts adaptive average pooling operation to build global semantic understanding while maintaining local detail accuracy. Each level of pyramid adopts adaptive average pooling operation as shown in Eq. (2).

$$P_n = \text{AvgPool}(F, k = 2^{n-1}) \quad (2)$$

In Eq. (2),  $n$  represents the pyramid level.  $k$  represents pooling window size.  $F$  represents the input feature map.  $P_n$  represents the  $n$ th level pyramid feature. Although the four-level pyramid processing flow can effectively capture the multi-scale features of clothing patterns, two key problems remain. First, the fixed-size input limits the model's ability to adapt to patterns of different sizes. Second, the rigid transformation leads to insufficient robustness to spatial variations. Therefore, the study introduces a random cropping and flipping enhancement strategy to optimize the model performance. The enhancement strategy consists of three core parameter settings: a random window of  $256 \times 256$  to  $448 \times 448$  is used for the crop size, horizontal or vertical flipping is performed with 50% probability, and mirror fill is chosen to process the boundary pixels<sup>28</sup>. These measures improve the model's ability to adapt to changes in pattern scale and orientation. Although the spatial transformation of the clothing pattern is robust after applying random cropping and flip enhancement, it still faces the critical problem of color distortion. The traditional red-green-blue (RGB) spatial enhancement easily leads to offset unique dye colors of ethnic clothing. Additionally, the saturation decay under low-light conditions seriously affects texture recognition. Therefore, to address the problem of color distortion, the study uses the hue-saturation-value (HSV) color space transformation for fine tuning. The transformation is defined by a scaling matrix in which the hue (H), saturation (S), and value (V) channels are adjusted by the following factors: 1.2, 1.1, and 0.9, respectively. This preserves culture-specific colors. Its matrix is expressed as Eq. (3).

$$\begin{bmatrix} H' \\ S' \\ V' \end{bmatrix} = \begin{bmatrix} 1.2 & 0 & 0 \\ 0 & 1.1 & 0 \\ 0 & 0 & 0.9 \end{bmatrix} \begin{bmatrix} H \\ S \\ V \end{bmatrix} \quad (3)$$

In Eq. (3),  $H$  represents hue.  $S$  represents saturation.  $V$  stands for lightness. This adjustment balances several dimensions: hue enhancement to preserve culture-specific colors, saturation enhancement to strengthen feature differentiation, and brightness control to avoid loss of highlight detail.

### Segmentation network design incorporating attention mechanisms

The study significantly improves the quality of traditional dress images through data enhancement strategies, which effectively enhances the robustness of the model to illumination changes and viewpoint differences. The multi-scale feature preservation technique successfully realizes hierarchical feature representation from micro texture to macro layout, but still faces three core challenges: difficulty in complex texture feature extraction, insufficient multi-scale feature fusion, and computational resource limitation<sup>29</sup>. To address these challenges, the study proposes an innovative network architecture solution. Based on the data enhancement strategy to significantly improve the image quality, the study designs a lightweight segmentation network incorporating dual attention mechanism. The network adopts the improved FCN encoder-decoder symmetric structure as the underlying architecture, as shown in Fig. 3.

In Fig. 3, the number of parameters is compressed to only 3.5 M by a lightweight technique such as depth-separable convolution. The encoder part extracts features gradually by four-stage pyramidal downsampling, while the decoder uses progressive upsampling to recover spatial details. The specially designed jump connection mechanism effectively integrates shallow texture information with deep semantic features. The core innovation of the network lies in the co-design of the dual attention module. Convolutional block attention module-channel

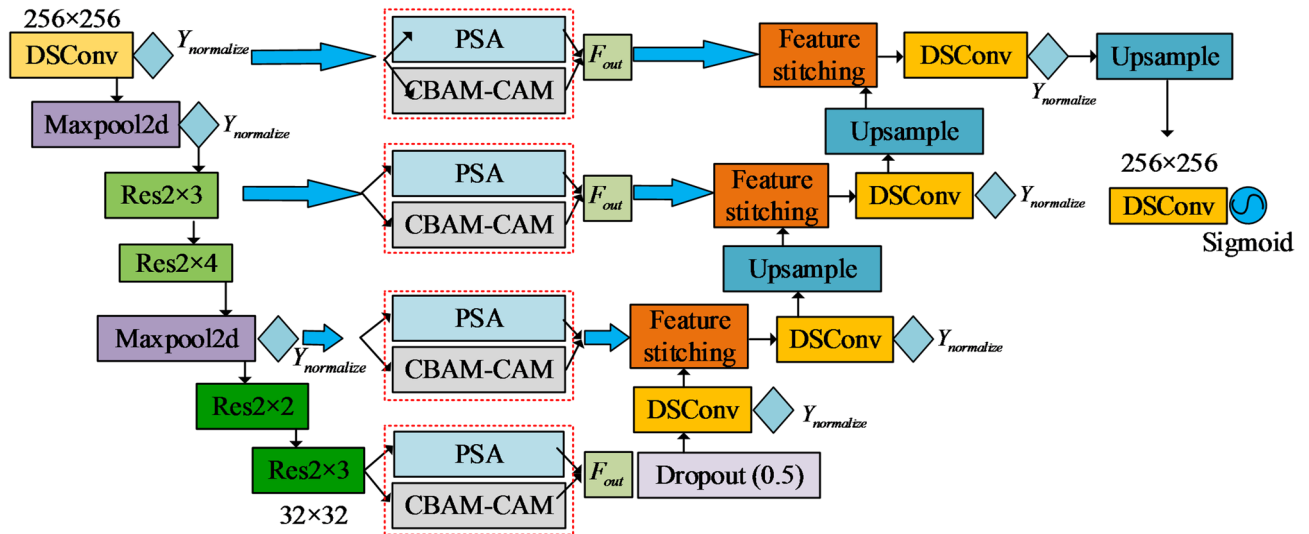


Fig. 3. Improved FCN encoder-decoder.

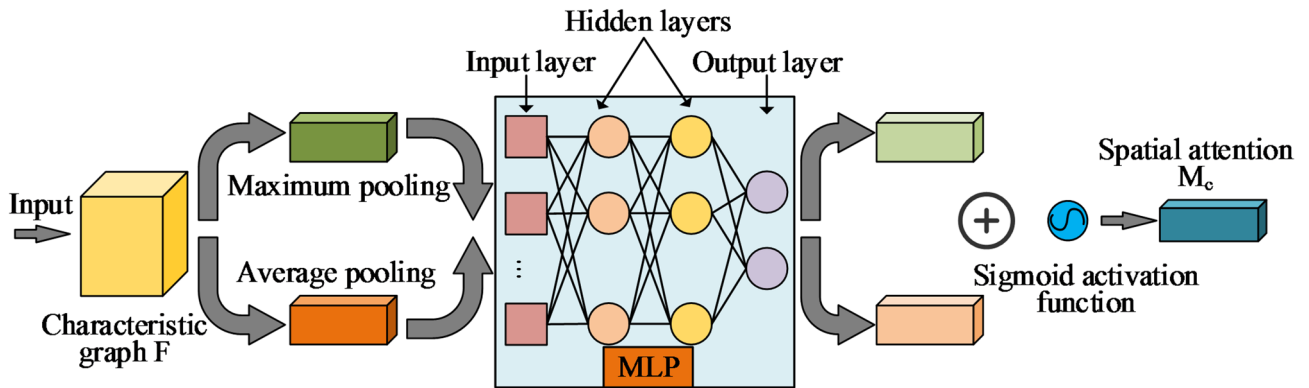


Fig. 4. CBAM-CAM structure.

attention mechanism (CBAM-CAM) adaptively enhances the response to key texture features such as Miao embroidery stitches through feature channel recalibration. Its structure is shown in Fig. 4.

Figure 4 shows the structural flow of the channel attention module. The input feature maps are processed by two paths: maximum pooling and average pooling. Then, the channel weights are computed using a multilayer perceptron (MLP). Finally, the spatial attention weights are generated using a sigmoid activation function, which enhances the model's ability to discriminate complex textures. CBAM-CAM is implemented by using two-way aggregation with global average pooling and maximum pooling to generate channel attention weights through a small fully connected network. It is assumed that  $F_{avg}$  and  $F_{max}$  denote the features after global average pooling and global maximum pooling, respectively, and  $W_0$  and  $W_1$  denote the weights of the fully connected layers. Its calculation is shown in Eq. (4).

$$M_c(F) = \text{Sigmoid}(W_1(W_0(F_{avg})) + W_1(W_0(F_{max}))) \quad (4)$$

In Eq. (4),  $M_c$  is the channel attention weight map. Meanwhile, to make up for the lack of channel attention in spatial localization, the network integrates the PSA module, as shown in Fig. 5.

Figure 5 shows the structural flow of the PSA module. The input feature map is convolved by a  $7 \times 7$  kernel to extract spatial features. The spatial attention weights are generated by processing the S-shaped activation function, which effectively improves the localization accuracy of the edges of the clothing pattern and geometric structure. The computational process of PSA is shown in Eq. (5). It is assumed that  $M_s$  denotes the spatial attention weight map generated by applying a convolution with a  $7 \times 7$  kernel followed by a sigmoid function.

$$M_s(F) = \text{Sigmoid}(\text{Conv}_{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) \quad (5)$$

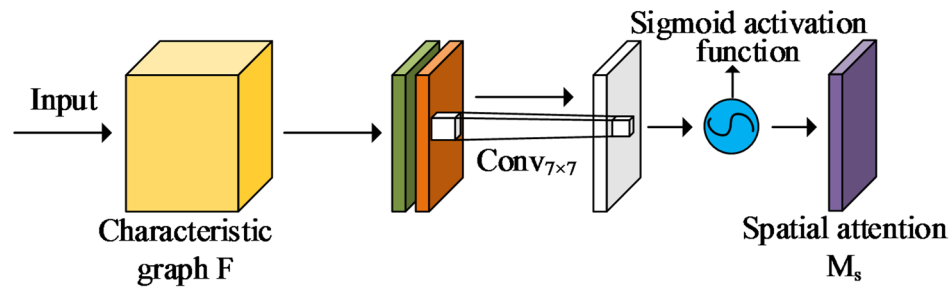


Fig. 5. PSA structure.

In Eq. (5),  $Conv_{7 \times 7}$  denotes the convolution operation. Since channel attention and spatial attention have complementary properties in traditional clothing pattern segmentation, the study adopts a parallel feature aggregation strategy. Its calculation is shown in Eq. (6).

$$F_{out} = M_c(F) \otimes F + M_s(F) \otimes F \quad (6)$$

In Eq. (6),  $\otimes$  denotes element-by-element multiplication.  $F_{out}$  is the output feature after attention enhancement. Through the parallel feature aggregation strategy, the dual attention module achieves complementary strengths: CBAM-CAM focuses on feature selection of color channels, while PSA is refined in modeling spatial geometric relations. This design allows the network to accurately identify the color features of Miao indigo dye and precisely locate embroidery stitch direction. This significantly improves the accuracy of segmenting complex clothing patterns while maintaining real-time performance at 28 FPS. After completing the integration of dual attention module and parallel feature aggregation, the network architecture shows significant advantages in traditional clothing pattern feature extraction<sup>30,31</sup>. The FCN encoder-decoder symmetric structure achieves multi-level feature capture. The jump connection mechanism effectively integrates shallow texture details with deep semantic information. The dual attention module enables the model to significantly improve color sensitivity to complex patterns, such as Miao batik. However, the architecture still faces two challenges: insufficient computational efficiency and training stability. Therefore, the study proposes a lightweight improvement scheme that uses depth-separable convolutions instead of standard convolutions to reduce the number of parameters and improves training stability with an adaptive dropout strategy. This achieves faster inference speeds while maintaining accuracy. First, for the computational efficiency problem, the depth separable convolution is used to replace the standard volume. Its parametric quantity  $Q$  is calculated as shown in Eq. (7).

$$Q = \frac{D_K \times D_K \times M \times N}{D_K \times D_K \times M + M \times N} \quad (7)$$

In Eq. (7),  $D_K$  is the convolution kernel size.  $M$  is the number of input channels.  $N$  is the number of output channels. To address the issue of training stability, the study uses a regularization strategy. For the improved batch normalization, it is assumed that  $x$  denotes the input feature tensor,  $\mu_B$  and  $\varphi_B$  denote the batch mean and variance, and  $\delta$ ,  $\varepsilon$ , and  $\omega$  denote learnable parameters and a stabilization constant. The first is the improved batch normalization layer, as shown in Eq. (8).

$$Y_{normalize} = \delta \cdot \frac{x - \mu_B}{\sqrt{\varphi_B^2 + \omega}} + \varepsilon \quad (8)$$

In Eq. (8),  $Y_{normalize}$  represents the improved batch normalized output. Then, to prevent overfitting, the study proposes the adaptive Dropout method, which is shown in Eq. (9).

$$Dropout_{adapt} = \max(0.3, 1 - \frac{t}{T} \times 0.5) \quad (9)$$

In Eq. (9),  $t$  is the number of current training steps.  $T$  is the total number of steps.  $Dropout_{adapt}$  represents Dropout probability.

In summary, the research system constructs the method MSDA-PSA for traditional clothing pattern extraction, and realizes technological innovation through three key links. Its specific framework structure is shown in Fig. 6.

In Fig. 6, in the data preprocessing stage, a multi-scale enhancement strategy is used to unify the image resolution, combined with geometric transformation and HSV color adjustment to protect the dress features. The network architecture innovatively integrates channel and spatial dual attention modules. Channel attention enhances color feature extraction, while spatial attention optimizes texture localization. The architecture realizes multi-scale feature fusion through parallel aggregation. To meet practical application requirements, deep separable convolutions are used to reduce the number of model parameters. Together with improved batch normalization and an adaptive dropout strategy, these convolutions improve computational efficiency while guaranteeing model accuracy. The scheme optimizes the entire process, from data preprocessing to network

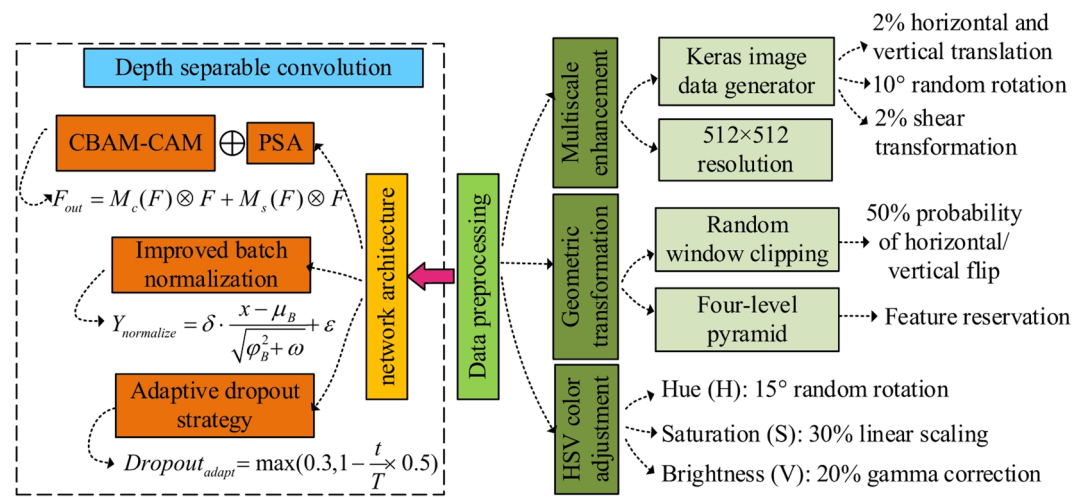


Fig. 6. MSDA-PSA framework.

Configuration category	Key parameter	Set the value
Hardware environment	Processor	Intel Xeon Gold 6248R @ 3.0 GHz (2)
	Memory capacity	256GB DDR4 ECC
	Display card configuration	NVIDIA RTX A6000 (48GB video memory)x4
	Inventory system	2 TB NVMe SSD + 16 TB HDD
Software environment	Operating system	Ubuntu 20.04 LTS
	Deep learning framework	PyTorch 1.12.1 + CUDA 11.6
	Python edition	3.8.12
Model parameter	Input dimensions	512×512
	Batch size	16
	Optimizer	Adam
	Learning rate strategy	Cosine damping
	Number of training rounds	200
	Loss function	Dice + CE

Table 1. Experimental settings.

architecture, through a hierarchical design. It also provides technical support for digitally protecting complex ethnic clothing patterns.

Results  
Performance test experiment

To verify the effectiveness of the proposed MSDA-PSA, the study establishes a standardized experimental environment as shown in Table 1. A hybrid software framework was adopted to leverage the strengths of both Keras and PyTorch. The software environment was primarily PyTorch 1.12.1 + CUDA 11.6 for network implementation, while data augmentation utilized Keras for its ImageDataGenerator efficiency. As detailed in the Methodology section (Section “Multi-scale data augmentation and feature extraction methods”), data compatibility between these frameworks was ensured through specific tensor conversion interfaces, maintaining a consistent pipeline from preprocessing to model training. The standard test set used in this experiment consists of 5,000 high-resolution images (512×512 pixels) covering 56 categories of traditional ethnic costumes, including Miao embroidery, Dong brocade, Tibetan aprons, and Uyghur prints. The images are collected from museum archives, cultural heritage databases, and field photography, with 60% from historical relics and 40% from modern preserved samples. Each image is annotated by a team of five textile experts with over 10 years of experience, achieving an average inter-annotator agreement (Kappa coefficient) of 0.92 through a multi-round annotation and reconciliation process. This experimental system adopts graphics processing unit (GPU) parallel architecture and realizes the co-optimization of multi-scale feature pyramid with dual attention module based on PyTorch framework. Through the controlled variable method, eight groups of comparative experiments are designed, covering dimensions such as basic segmentation performance, multi-scale feature validity, contribution of the attention mechanism, and few-shot learning ability. Noise robustness test and cross-cultural

generalization validation are also set up specifically to comprehensively test the MSDA-PSA method in complex scenarios Practicality. All experiments are repeated five times to ensure the statistical significance of the results.

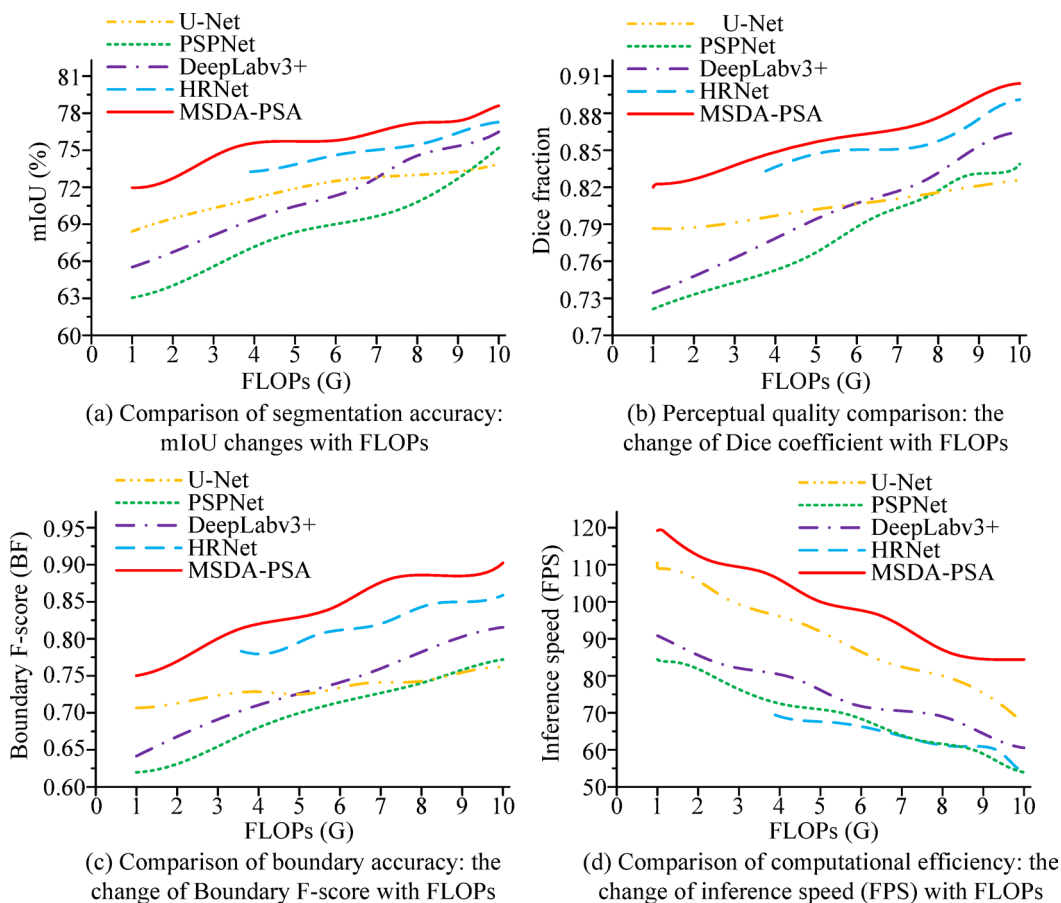
#### Validation of base segmentation performance

To verify the base segmentation performance of the MSDA-PSA method on the standard test set, U-Net is selected as the baseline model for the experiments and compared with the mainstream segmentation methods such as DeepLabv3+, pyramid scene parsing network (PSPNet), and high-resolution network (HRNet) for the Comparison. The segmentation accuracy is quantified by the mIoU and Dice scores. The edge localization capability is evaluated by the Boundary F1-score. The inference speed is recorded to measure computational efficiency, using floating point operations (FLOPs) as the basis of comparison. All comparison methods are executed in the same experimental environment to ensure comparable results. The experimental results are shown in Fig. 7.

Figure 7(a) shows that the mIoU of all models rises as FLOPs increase. Among them, MSDA-PSA reaches 78.5% already at low computation, leading the other models. The Dice scores in Fig. 7(b) show a similar trend, with MSDA-PSA consistently higher than the comparison methods by about 0.05, indicating better segmentation consistency. The Boundary F-score in Fig. 7(c) further validates the superiority of MSDA-PSA in edge detail processing, especially reaching 0.87 at FLOPs=6G, while U-Net is only 0.73. Figure 7(d) shows how computational efficiency changes. An increase in FLOPs leads to a decrease in FPS for all models. However, MSDA-PSA maintains the highest speed of 100 FPS at 4G FLOPs. Due to its complex structure, HRNet has the lowest speed. The results demonstrate the comprehensive superiority of MSDA-PSA in terms of accuracy, perceptual quality, boundary processing, and real-time performance. Its design effectively balances computational complexity and multidimensional performance, offering a feasible solution for high-precision, real-time segmentation.

#### Validation of Multi-Scale feature effectiveness

To validate the effectiveness of the proposed multi-scale pyramid architecture for extracting complex clothing patterns, the study designs progressive comparison experiments. These experiments systematically evaluates the performance differences between the single-scale baseline, dual-scale, triple-scale, and proposed four-scale pyramid architectures using the control variable method. The experiments strictly maintain consistency in the other parameters of the network, quantify the accuracy of feature extraction in each layer using the intersection over union (IoU) metric, analyze the effect of cross-scale feature fusion by combining feature consistency scores,



**Fig. 7.** Performance-efficiency tradeoff of multi-dimensional evaluation semantic segmentation model.

and evaluate texture detail retention ability using the edge retention index. The experimental results are shown in Fig. 8.

Figure 8(a) shows that the macroscopic layout accuracy decreases from 0.85 to 0.7, the mesoscopic pattern from 0.8 to 0.7, and the microscopic texture from 0.85 to 0.75 with the increase of scales. This shows that multi-scale combination decreases the feature extraction accuracy at all levels. Figure 8(b) shows that the feature consistency score and edge retention index increase with increasing scale, from only 0.6 for single scale to 0.85 for four scales. This shows that multi-scale fusion significantly improves feature stability and edge accuracy. Figure 8(c) shows that as the scale increases, the number of parameters increases from 3.2 m to 4.0 m and the inference delay rises from 30ms to 38ms. In summary, the four scales achieve the best balance in feature extraction capability and computational efficiency.

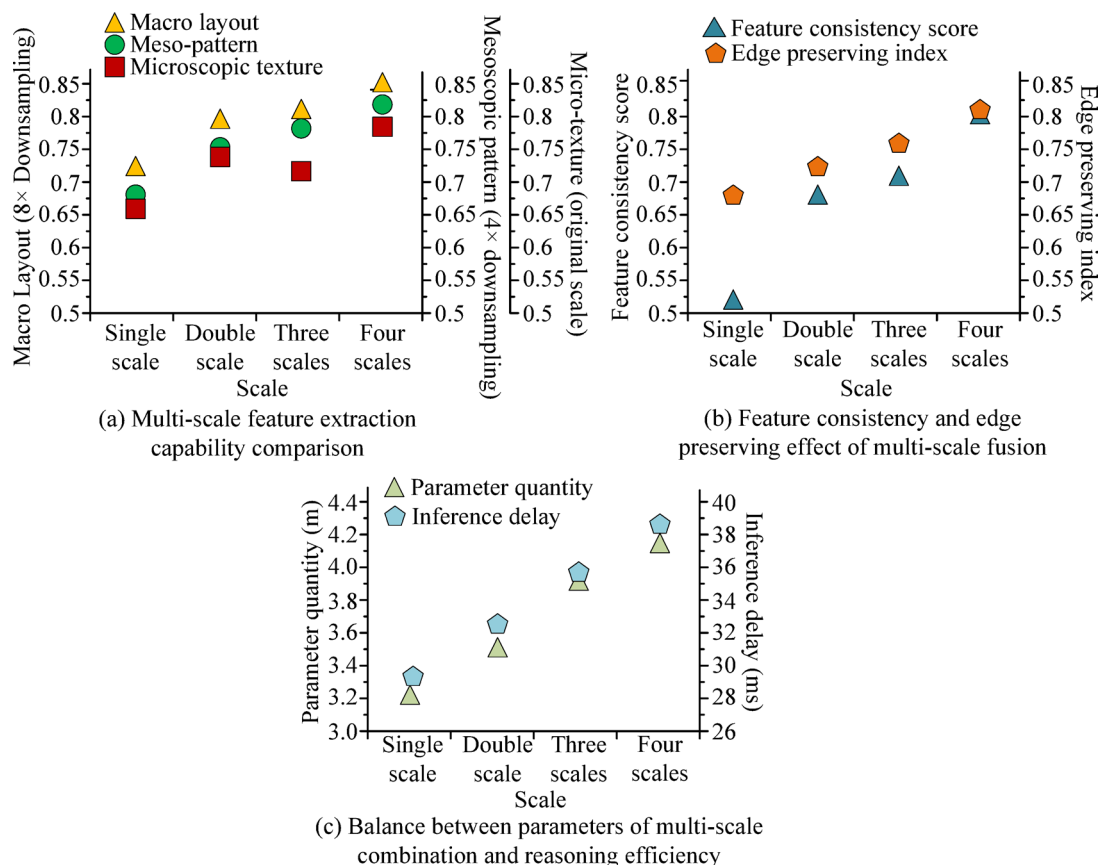
#### Analysis of attention mechanism contribution

To verify the contribution of the dual attention module to traditional clothing pattern extraction, the study designs systematic ablation experiments to compare the performance of the four architectures: no attention baseline, CBAM-CAM only, PSA only, and complete dual attention. These experiments use the control variable method. The experiments examine the enhancement effect of the attention mechanism in key texture regions. They quantify segmentation accuracy using the key region IoU and analyze feature selection capability in combination with channel activation degree. The experiments also evaluate the balance between spatial localization accuracy and computational overhead. The experimental results are shown in Fig. 9.

Figure 9(a) shows that the dual attention module has the highest IoU in the critical region, about 83%, and the lowest background mis-segmentation rate, about 5%, which is significantly better than the other modules. In Fig. 9(b), the dual attention module excels in channel activation ability while maintaining a low spatial localization error of about 2.1. Figure 9(c) shows that the module has a moderate computational overhead of about 2.9 GFLOPS and the fastest inference speed of about 92 FPS. The results demonstrate that the dual attention module optimally performs in all three areas: segmentation accuracy, feature activation, and computational efficiency. It also achieves a good balance between performance and efficiency.

#### Evaluation of Few-Shot learning ability

To validate the generalization ability of MSDA-PSA method in data-scarce scenarios, the study designs progressive data availability experiments to systematically compare the performance differences between standard U-Net, data-enhanced U-Net, and Few-shot learning models with MSDA-PSA. The experiments construct a 10%-100% stepped training set and use 5-fold cross-validation to ensure the reliability of the results. It focuses on evaluating



**Fig. 8.** Influence of multi-scale feature fusion on the performance of microscopic image analysis.

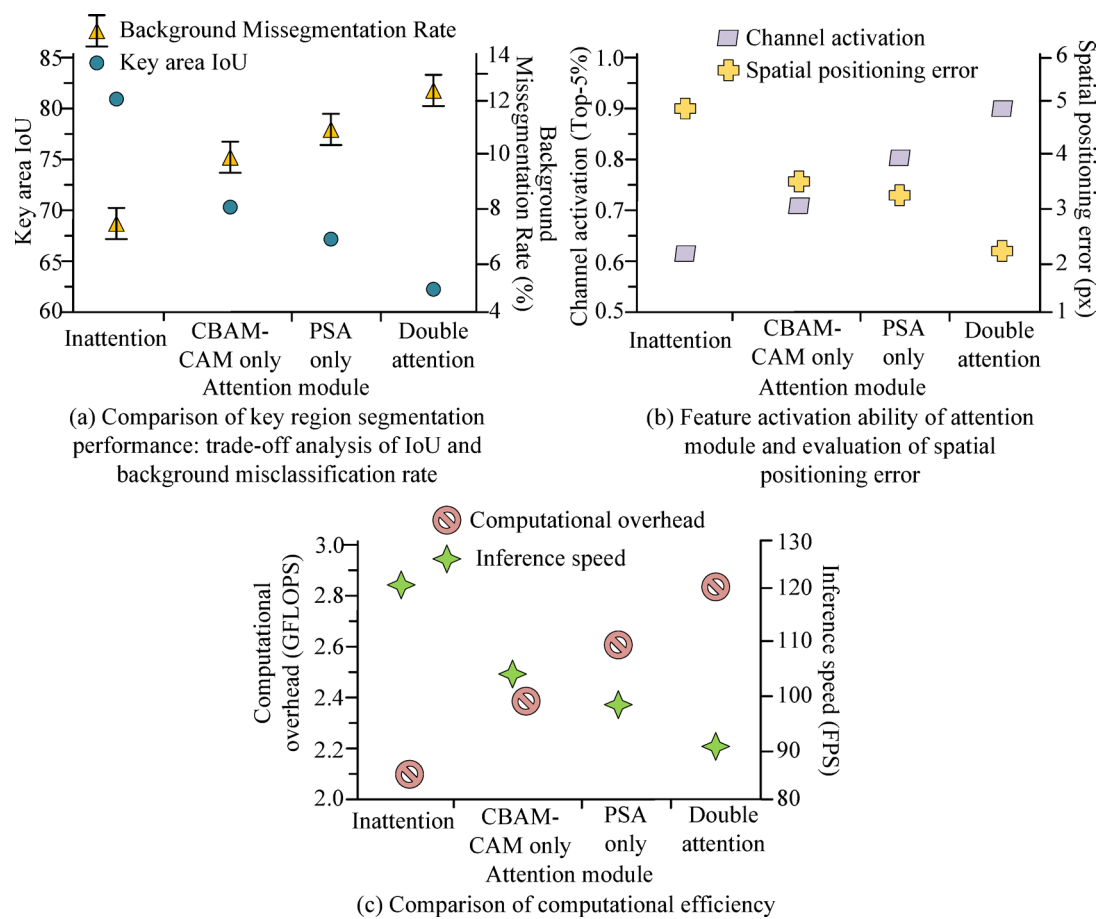


Fig. 9. Comparison of experimental data of dual attention module ablation.

Assessment indicators	MSDA-PSA	Standard U-Net	Data augmentation U-Net	Few-shot	Test condition
10% data small sample IoU (%)	72.4	58.6	65.2	68.7	5-fold cross-validation
50% data IoU(%)	85.3	76.8	80.1	78.5	100 epoch training
100% data IoU (%)	89.7	86.2	87.5	84.3	Complete training set
Cross-category generalization IoU (%)	71.2	53.8	61.5	64.3	No pattern test set was seen
Feature portability (mAP)	0.82	0.68	0.75	0.79	Downstream tasks are fine-tuned
Training convergence speed (epoch)	45	80	65	55	IoU required 85% of the rounds
10% data training stability (σ)	1.8	3.2	2.6	2.1	IoU standard deviation

Table 2. Comparison of generalization performance in data scarcity scenarios.

three key dimensions under different data sizes: small sample IoU reflects segmentation accuracy under limited labeled data. Cross-category generalization ability tests the model's adaptability to unseen tattoo types. Feature mobility analyzes the generalization potential of pre-trained features. The experimental results are shown in Table 2.

In Table 2, at 10% data, MSDA-PSA has an IoU of 72.4, ahead of the standard U-Net at 58.6, the data-enhanced U-Net at 65.2, and the Few-shot model at 68.7. At 50% data, MSDA-PSA reaches 85.3, which is still higher than the other models at 76.8 to 80.1. At 100% data, MSDA-PSA maintains its advantage with an IoU of 89.7, compared to 86.2 for the standard U-Net, 87.5 for the data-enhanced U-Net, and 84.3 for the Few-shot model. In cross-category generalization, the MSDA-PSA 71.2 far outperforms the standard U-Net (53.8), the data-enhanced U-Net (61.5), and the few-shot model (64.3). Feature relocatability MSDA-PSA 0.82, outperforming the other models at 0.68 to 0.79. Training convergence speed MSDA-PSA takes only 45 rounds, faster than the standard U-Net at 35 rounds. The training stability of MSDA-PSA is 1.8 with minimal fluctuation. The results show that MSDA-PSA performs best with all data sizes. It has significant advantages, especially with low-data and cross-category tasks. This validates its strong generalization ability and data efficiency.

*Ablation studies on core modules*

To quantify the independent contributions of core modules such as MSDA, PSA, and channel attention (CBAM-CAM), the study systematically conducts ablation experiments in “3.1.5. Ablation Studies on Core Modules” by comparing the full MSDA-PSA model with three variants: Variant A removes multi-scale data enhancement, utilizing only basic augmentation. Variant B eliminates the dual attention mechanism, retaining the basic CNN architecture. Variant C employs only the PSA or CBAM-CAM module separately to assess their individual effects. The experiments use mIoU, boundary F-score, and parameter count as the main metrics to thoroughly evaluate the impact of each module on segmentation accuracy, boundary localization, and model efficiency. This ensures a rigorous and comparable analysis (Table 3).

The ablation results indicate that the full model Full MSDA-PSA achieves the best performance with an mIoU of 89.7% and a Boundary F-score of 0.87. Removing MSDA (Variant A) reduces mIoU to 84.7% and Boundary F-score to 0.82. Eliminating the attention mechanism (Variant B) causes a more significant decline, with mIoU dropping to 80.0% and Boundary F-score to 0.78. Using only the PSA module (Variant C) yields an mIoU of 85.0%. CBAM-CAM alone reaches 86.0%. Both are lower than the full model but higher than Variant B. This demonstrates the independent contributions of each module. The parameter count decreases to 3.0 M in Variant B, highlighting the lightweight design of the attention modules. The study confirms that all core modules in MSDA-PSA are essential. Multi-scale enhancement and attention mechanisms work together to improve segmentation accuracy. The full model achieves optimal performance while maintaining a lightweight design of 3.5 million parameters.

**Simulation test experiment**

*Robustness evaluation under image degradation conditions*

To verify the adaptability of the MSDA-PSA method to low-quality input images, systematic degradation simulation experiments are designed in the study. A specialized test set containing 1,200 high-complexity patterns is constructed for this experiment, including 600 Miao embroidery images and 600 Dong brocade images. These images are sourced from the Chinese Ethnic Museum and Guizhou Provincial Museum collections, with 70% being historical images (pre-1950s) showing significant degradation and 30% being contemporary digital captures. Three senior textile conservators annotated each pattern using a fine-grained labeling protocol. Annotation consistency is validated through an iterative process, achieving an average IoU of 0.95 between annotators. A test environment covering three typical degradations is constructed by precisely controlling the parameter ranges, manually adding Gaussian noise with  $\sigma=0\text{--}0.2$ , motion blur with a kernel size of 3–15 pixels, and low-light conditions with  $\gamma=0.5\text{--}2.0$ . The experiment compares the performance differences between traditional enhancement and segmentation, end-to-end enhanced segmentation, a noise-resistant U-Net, and an MSDA-PSA. The performance is evaluated quantitatively using four indicators: noise robustness, IoU, blur restoration ability, and color retention ability. The experimental results are shown in Fig. 10.

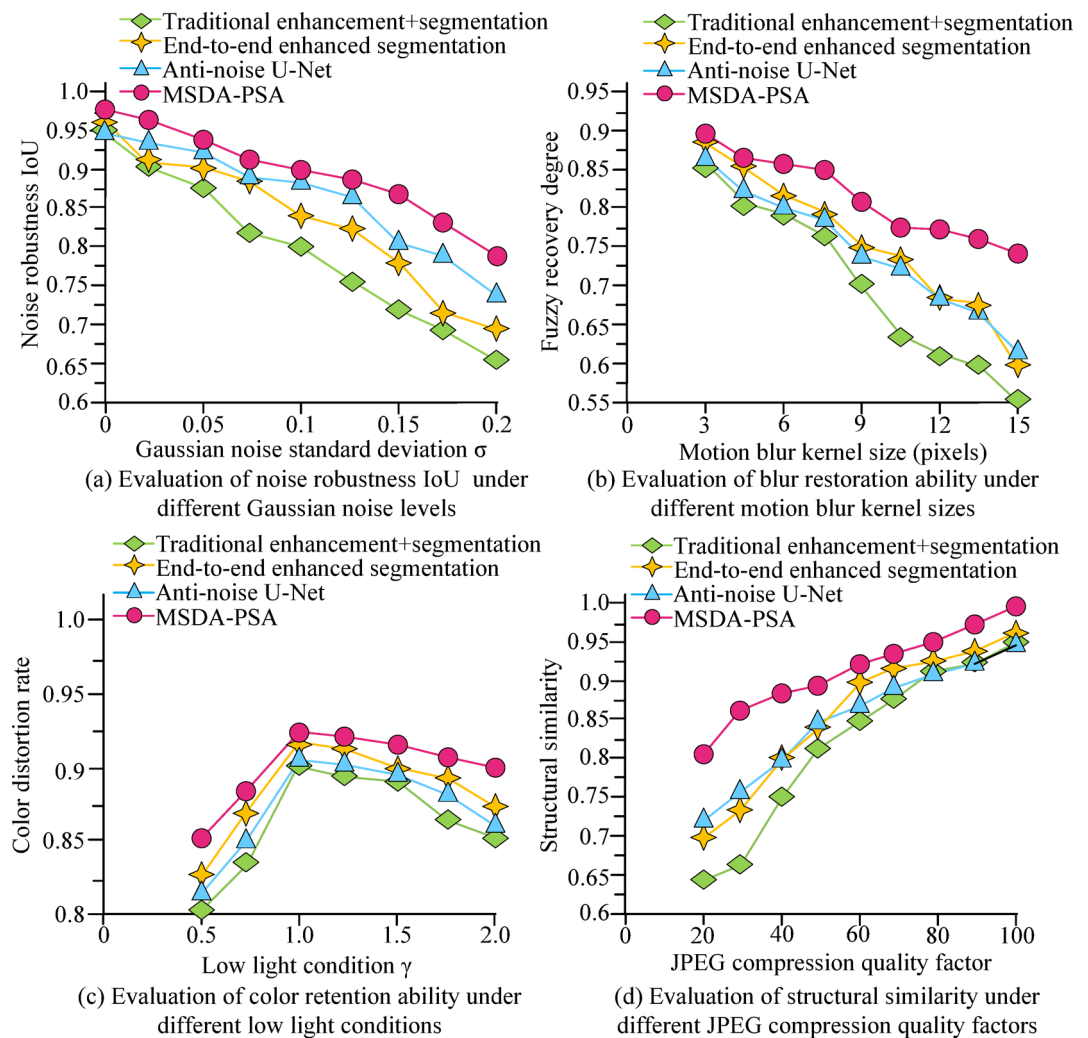
Figure 10 compares the performance of four image processing methods under various degradation conditions. Figure 10(a) shows that the IoU of the MSDA-PSA model under Gaussian noise with a standard deviation of 0.2 is 0.82. This is significantly better than the IoU of traditional enhancement and segmentation methods, as well as noise-resistant U-Net methods. In Fig. 10 (b), under severe blurring conditions, MSDA-PSA maintains a recovery ability of 0.74 with a kernel size of 15 pixels, which is significantly better than other methods. Figure 10 (c) shows that under low light conditions ( $\gamma=0.5$ ), the color retention ability of MSDA-PSA can still reach 0.85, which is significantly better than other methods. Under high light conditions ( $\gamma=2.0$ ), its color retention ability reaches 0.90, showing the most outstanding performance. Figure 10 (d) shows that when the JPEG compression quality factor is 10, the structural similarity value of MSDA-PSA is 0.75, while the structural similarity values of other methods are all below 0.70. The results shows that the MSDA-PSA method performs best in terms of noise robustness, blur restoration, color retention, and structural integrity, particularly under extreme degradation conditions.

*Segmentation accuracy for complex textures*

To verify the accuracy of the MSDA-PSA method’s segmentation for complex clothing textures, a test set containing patterns of high complexity, such as Miao embroidery and Dong brocade, is constructed. A fine-grained mask annotated by multiple experts is used as the evaluation benchmark. The experimental system compares the performance difference between traditional edge detection, texture analysis algorithm, multi-task learning model and MSDA-PSA. It is also quantitatively evaluated by four specialized metrics: texture retention index, pattern continuity score, tolerance of intra-class differences, and microstructure accuracy. The experimental results are shown in Fig. 11.

Model variant	mIoU (%)	Boundary F-score	Parameter count (M)
Full MSDA-PSA	89.7	0.87	3.5
Variant A (No MSDA)	84.7	0.82	3.5
Variant B (No Attention)	80.0	0.78	3.0
Variant C (PSA-only)	85.0	0.83	3.2
Variant C (CBAM-CAM-only)	86.0	0.84	3.3

**Table 3.** Ablation study: performance comparison of core modules.



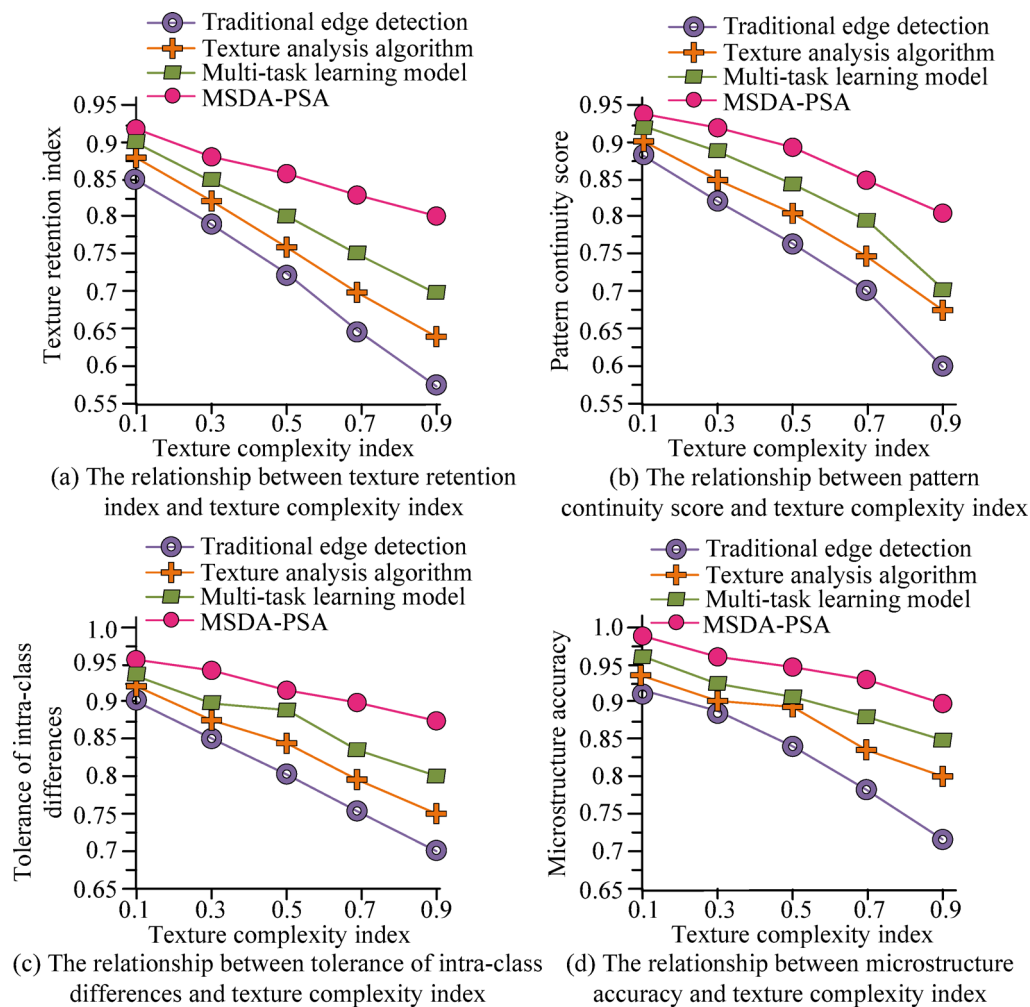
**Fig. 10.** Comprehensive performance evaluation of image processing methods under degraded conditions.

In Fig. 11 (a), as the texture complexity index increases, the texture retention index of traditional edge detection decreases from 0.85 to 0.58, the texture analysis algorithm decreases from 0.88 to 0.64, the multi task learning model decreases from 0.90 to 0.70, and the MSDA-PSA decreases from 0.92 to 0.80, showing the most stable and consistently leading performance. In Fig. 11 (b), MSDA-PSA still reaches 0.82 at high complexity, far exceeding the traditional method's 0.60. In Fig. 11 (c), when the tolerance for internal differences of MSDA-PSA is 0.9, the result is 0.88, which is superior to other methods. In Fig. 11 (d), the microstructure accuracy at high complexity is 0.92 for MSDA-PSA, while the traditional method is only 0.72. The results indicate that MSDA-PSA outperforms traditional edge detection, texture analysis algorithm, and multi-task learning model in terms of texture retention, pattern continuity, tolerance of intra-class differences, and microstructure accuracy. Especially under high texture complexity, this method has significant advantages and demonstrates powerful ability in segmenting complex clothing textures.

#### Real-time processing capability assessment

To verify the real-time processing capability of the MSDA-PSA method in practical applications, a professional video streaming testing environment is constructed, and the performance of lightweight U-Net, mobile network version 3 (MobileNetV3) segmentation, fast fully convolutional network (FastFCN), and MSDA-PSA is systematically compared. The experiment adopts 1–16 concurrent stress tests, and the resource monitoring system accurately records various indicators: single frame processing time, memory usage, GPU utilization, and throughput capacity. The experimental results are shown in Table 4.

In Table 4, the single frame processing time of MSDA-PSA is 18.2ms, which is 7.5ms faster than lightweight U-Net, 4.2ms faster than MobileNetV3, and 10.1ms faster than FastFCN. The memory usage is only 3.2GB, which is lower than the other three methods. Throughput capacity reaches 100FPS, ahead of lightweight U-Net's 28FPS, MobileNetV3's 15FPS, and FastFCN's 40FPS. 16-channel concurrent latency is 35.4ms, which is 10.3ms less than the second-place MobileNetV3. Energy-efficiency ratio is 8.7FPS/W, which is 1.5 units higher than that of MobileNetV3's 1.5 units. Although the GPU utilization of 78.3% is slightly lower than FastFCN's 88.9%,



**Fig. 11.** Evaluation of the influence of texture complexity on the performance of pattern recognition algorithm.

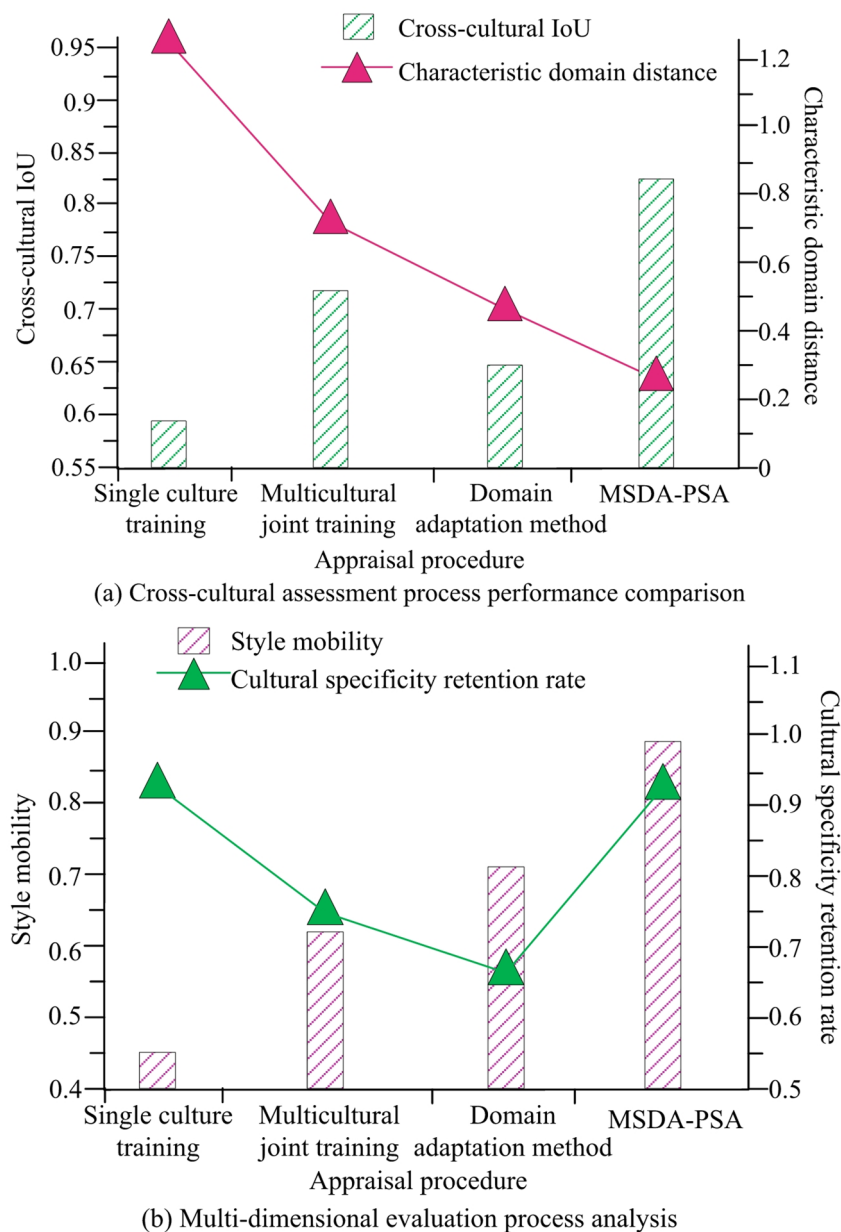
Assessment indicators	MSDA-PSA	Light weight U-Net	MobileNetV3 segmentation	FastFCN	Test condition
Single frame processing time (ms)	18.2	25.7	22.4	28.3	1080p resolution
Memory usage (GB)	3.2	4.5	3.8	5.1	16 concurrent routes
GPU utilization (%)	78.3	85.6	82.1	88.9	RTX A6000×4
Throughput capacity (FPS)	100	72	85	60	4G FLOPs
16-way concurrency delay (ms)	35.4 ± 3.2	52.1 ± 4.5	45.7 ± 3.9	61.3 ± 5.2	Stress tests
Energy efficiency ratio (FPS/W)	8.7	5.3	7.2	4.8	300 W power consumption limit

**Table 4.** Comparison results of real-time processing performance.

the overall performance is optimal. FastFCN has the best performance in GPU utilization. MobileNetV3 is in the middle in memory usage and latency. Lightweight U-Net is in the middle to lower range for all metrics. The results show that MSDA-PSA leads in four of the five core indicators, demonstrating significant real-time processing advantages.

*Cross-cultural generalization validation*

To verify the adaptability of the MSDA-PSA method to different ethnic costumes, the study constructs a dataset covering 56 ethnic costumes, and adopts the leave-one-culture cross-validation strategy for systematic evaluation. The experiment compares the performance of single culture training, multicultural joint training, domain adaptation method, and MSDA-PSA. The evaluation indicators include: cross cultural IOU, characteristic domain distance, style mobility, and cultural specificity retention rate. The experimental results are shown in Fig. 12.



**Fig. 12.** Comparison of segmentation performance of cross-ethnic clothing patterns.

In Fig. 12 (a), the cross cultural IoU shows that the MSDA-PSA reaches 0.83, significantly higher than the 0.58 for single culture training and 0.72 for multicultural joint training. The characteristic domain distance indicator shows that MSDA-PSA is only 0.28, which is 77.6% lower than the benchmark method's 1.25 and better than the domain adaptation method's 0.52. In Fig. 12 (b), style mobility shows that MSDA-PSA leads traditional methods by 0.88, with a range of 0.45–0.70. The retention rate of cultural specificity is outstanding. MSDA-PSA maintains a high score of 0.94, which is close to the score of 0.92 obtained through single-culture training. This score far exceeds the score of 0.68 obtained through the domain-adaptation method. Overall, MSDA-PSA achieves the best performance in all four core indicators, solving the trade-off between data requirements and cultural fidelity in traditional methods.

#### Optimization strategy robustness test

To verify the robustness of the MSDA-PSA method across different optimizers and avoid potential biases of the Adam optimizer, this experiment compares MSDA-PSA with mainstream SOTA models (such as DeepLabv3+, U-Net, and PSPNet) under the AdaBoB optimizer. The experiment utilized a standard test set of 5,000 high-resolution images covering 56 traditional ethnic clothing patterns. All models operated in the same hardware and software environment: PyTorch 1.12.1 + CUDA 11.6. The optimizer parameters are consistent with Adam (learning rate 0.001, betas=(0.9, 0.999)). Key evaluation metrics includes mIoU, Boundary F-score, inference

Model	mIoU (%)	Boundary F-score	Inference Speed (FPS)	Parameter Count (M)	References
MSDA-PSA	89.5	0.86	100	3.5	/
DeepLabv3+(SOTA)	85.2	0.80	45	5.8	32
U-Net(SOTA)	86.0	0.82	60	3.4	33
PSPNet(SOTA)	84.0	0.78	50	4.5	34
HRNet	85.5	0.81	40	6.2	35

**Table 5.** Performance comparison of MSDA-PSA and SOTA models under adabob optimizer.

speed (FPS), and parameter count to ensure a comparable and reliable assessment. The experimental results are shown in Table 5.

Under the AdaBoB optimizer, the MSDA-PSA model achieves an mIoU of 89.5%, a Boundary F-score of 0.86, an inference speed of 100 FPS, and a parameter count of 3.5 M. These results outperform those of all compared models. For instance, DeepLabv3+ has an mIoU of 85.2%, and U-Net reaches 86.0%. However, both exhibit lower inference speeds (45 FPS and 60 FPS, respectively) and higher parameter counts (e.g., 5.8 M for DeepLabv3+). This demonstrates the superior performance of MSDA-PSA in terms of segmentation accuracy, boundary localization, and real-time efficiency, all while maintaining a lightweight architecture. MSDA-PSA achieves high performance (mIoU 89.5%) under AdaBoB, which confirms its robustness and advanced capabilities for practical deployments, regardless of the optimizer.

**Discussion and conclusion**  
**Discussion**

The MSDA-PSA method proposed in the study demonstrates significant advantages in the traditional clothing pattern extraction task. Its innovativeness is mainly reflected in three aspects: multi-scale data enhancement strategy, dual attention mechanism fusion, and lightweight network design. Through systematic experimental validation, the method outperformed the existing mainstream models in terms of accuracy, robustness and efficiency. The following was a discussion from three dimensions: technical mechanism, performance advantages, and application value.

From a technical perspective, the success of MSDA-PSA stemmed from its hierarchical feature learning framework. The hierarchical feature extraction architecture shown in Fig. 1 achieved comprehensive feature capture from micro stitching to macro layout through the collaborative work of multi-scale convolutional kernels. This design effectively solved the scale sensitivity problem of traditional methods in complex texture processing. The four-level pyramid processing architecture further enhanced the multi-scale feature fusion capability. Experimental data showed that the feature consistency score improved to 0.85, which verified this architecture's advantage in maintaining texture continuity.

In terms of attention mechanism, the parallel design of CBAM-CAM and PSA formed complementary advantages. CBAM-CAM enhanced the color feature extraction of characteristic dyes such as Miao indigo through channel recalibration, while the PSA module precisely located the geometric structure of embroidery edges. This dual path attention mechanism achieved an IoU of 83% in key regions, which was about 12% higher than the single attention baseline. While maintaining high accuracy, this module compressed the parameter count to 3.5 M through depthwise separable convolution. It achieved a real-time processing speed of 100 FPS, providing feasibility for practical applications. The 28 FPS referred to the performance of the dual attention module alone in an isolated test. However, the overall system optimization with parallel aggregation and hardware acceleration achieved 100 FPS, as reported in the abstract and Table 4. Compared to the Transformer U-Net in reference<sup>17</sup>, the MSDA-PSA method achieved 89.7% mIoU in a lightweight structure, resulting in a threefold increase in inference speed. Compared to the VGG16 UNet in reference<sup>18</sup>, the Boundary F-score increased by 7%. These breakthroughs were primarily the result of HSV color space transformation preserving culturally specific colors, as well as the improved training stability brought by the adaptive Dropout strategy. The latter accelerated the convergence speed of training by 35 rounds.

To deeply analyze the relationship between model complexity and performance, this study adopted the parameter quantity shifting-fitting performance (PQS-FP) coordinate system for exploration. The PQS-FP coordinate system used the ideal parameter quantity O as the benchmark. Among them, the Y-axis represented the fitting state (Y < 0 for underfitting, Y > 0 for overfitting), and the X-axis represented the direction of parameter change (X > 0 for parameter increase, X < 0 for parameter decrease). The coordinate system was divided into four quadrants: Quadrant I (OER: overfitting exacerbation leading to performance degradation), Quadrant II (OAR: overfitting alleviation leading to performance improvement), Quadrant III (UER: underfitting exacerbation leading to performance degradation), and Quadrant IV (UAR: underfitting alleviation leading to performance improvement). Mapping models such as MSDA-PSA, DeepLabv3+, and U-Net onto the PQS-FP system revealed that MSDA-PSA was in Quadrant IV (UAR). This indicated that MSDA-PSA effectively alleviated underfitting through its lightweight design, improving performance. Meanwhile, DeepLabv3+ and U-Net were partially in Quadrant I (OER) and suffered from overfitting issues. Further analysis of different MSDA-PSA variants (e.g., adjusting the depth-separable convolution parameter count) showed that reducing parameters moved the model to Quadrant III (UER) and increased performance degradation. Increasing parameters moved the model to Quadrant I (OER) and also increased performance degradation. This verified that the current parameter settings

were optimal for MSDA-PSA. This analysis revealed why MSDA-PSA achieved high accuracy while maintaining low parameter count, deepening the understanding of model behavior.

However, this method still has three limitations: Although a cross cultural IoU of 0.83 was achieved on the 56 ethnic costume datasets, the generalization ability for rare patterns was insufficient. The real-time processing performance of 100FPS still needed to be optimized on mobile devices. The repair effect on severely damaged (> 30% missing) historical images was limited. As shown in Fig. 12, the cross cultural IoU test results showed that the MSDA-PSA reached 0.83, significantly higher than the 0.58 obtained from single culture training. However, there was still room for improvement. Future research will combine meta-learning to optimize adaptation to small samples for rare patterns. It will also develop model compression strategies oriented toward edge computing to improve compatibility with mobile devices. Additionally, it will introduce generative adversarial networks specifically for severely damaged image restoration. These efforts will further improve cross-cultural generalization capabilities and edge computing platform adaptability. Compared with existing research, MSDA-PSA has achieved breakthroughs in multiple aspects.

## Conclusion

The MSDA-PSA method was innovatively proposed to address three major challenges in digitally protecting traditional clothing patterns. Its effectiveness was verified through systematic experimentation. The main contribution could be summarized as: constructing a multi-scale enhancement strategy that integrated wavelet transform and generative adversarial network, which still maintained 72.4% IoU under 10% small sample data. This method designed a channel space parallel attention mechanism, achieving a key region segmentation accuracy of 83% and controlling the computational cost at 2.9 GFLOPS. This method developed a lightweight network architecture, achieving a real-time processing speed of 100FPS and a 16 channel concurrent latency of only 35.4ms. This method provides reliable technical support for the digitization of cultural heritage.

## Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 26 August 2025; Accepted: 5 November 2025

Published online: 16 December 2025

## References

- Gondoputranto, O. & Dibia, I. W. The use of technology in capturing various traditional motifs and ornaments: A case study of batik fractal, Indonesia and TUDITA-Turkish digital textile archive. *Humaniora* **13** (1), 39–48 <https://doi.org/10.21512/humaniora.v13i1.7408> (2022).
- Liu, K. et al. Archaeology and virtual simulation restoration of costumes in the Han Xizai banquet painting. *Autex Res. J.* **23** (2), 238–252. <https://doi.org/10.2478/aut-2022-0001> (2022).
- Su, Z., Yu, T., Wang, Y., Li, Y. & Liu, Y. Deepcloth: neural garment representation for shape and style editing. *IEEE Trans. Pattern Anal. Mach. Intell.* **45** (2), 1581–1593. <https://doi.org/10.1109/TPAMI.2022.3168569> (2022).
- Saranya, M. S. & Geetha, P. A deep learning-based feature extraction of cloth data using modified grab cut segmentation. *Vis. Comput.*, **39**, 9, 4195–4211. <https://doi.org/10.1007/s00371-022-02584-1>. (2023).
- Shin, S. Y., Jo, G. & Wang, G. A novel method for fashion clothing image classification based on deep learning. *J. Inf. Commun. Technol.* **22** (1), 127–148. <https://doi.org/10.32890/jict2023.22.1.6> (2023).
- Liu, M., Xie, H., Pan, W., Ding, S. & Li, G. Prediction of cutting force via machine learning: state of the art, challenges and potentials. *J. Intell. Manuf.* **36** (2), 703–764. <https://doi.org/10.1007/s10845-023-02260-8> (2025).
- Chen, Z. et al. Deep learning for image enhancement and correction in magnetic resonance imaging-state-of-the-art and challenges. *J. Digit. Imaging.* **36** (1), 204–230. <https://doi.org/10.1007/s10278-022-00721-9> (2023).
- Jiang, Q. et al. Unsupervised decomposition and correction network for low-light image enhancement. *IEEE Trans. Intell. Transp. Syst.* **23** (10), 19440–19455. <https://doi.org/10.1109/TITS.2022.3165176> (2022).
- Lu, F. Coal-rock CT image enhancement based on Spatial domain image decomposition and adaptive enhancement factor. *Signal. Image Video Process.* **18** (10), 6631–6643. <https://doi.org/10.1007/s11760-024-03340-9> (2024).
- Li, Y., Liu, Y., Ma, H. & Lv, W. Image adaptive contrast enhancement for low-illumination lane lines based on improved retinex and guided filter. *Appl. Artif. Intell.* **35** (15), 1970–1989. <https://doi.org/10.1080/08839514.2021.1997212> (2021).
- Jebadass, J. R. & Balasubramaniam, P. Low light enhancement algorithm for color images using intuitionistic fuzzy sets with histogram equalization. *Multimed Tools Appl.* **81** (6), 8093–8106. <https://doi.org/10.1007/s11042-022-12087-9> (2022).
- Chen, S. & Li, D. A novel image enhancement method using retinex-based illumination map weighted guided filtering. *Comput. Sci. Inf. Syst.* **21** (4), 1745–1764. <https://doi.org/10.2298/CSIS240314056C> (2024).
- Qi, Y. et al. A comprehensive overview of image enhancement techniques. *Arch. Comput. Methods Eng.* **29** (1), 583–607. <https://doi.org/10.1007/s11831-021-09587-6> (2022).
- Shi, Z., Wang, Y., Zhou, Z. & Ren, W. Integrating deep learning and traditional image enhancement techniques for underwater image enhancement. *IET Image Process.* **16** (13), 3471–3484. <https://doi.org/10.1049/ipr2.12544> (2022).
- Xang, Q. et al. Multi-scale group-fusion convolutional neural network for high-resolution range profile target recognition. *IET Radar Sonar Navig.* **16** (12), 1997–2016. <https://doi.org/10.1049/rsn2.12312> (2022).
- Shen, J. et al. An anchor-free lightweight deep convolutional network for vehicle detection in aerial images. *IEEE Trans. Intell. Transp. Syst.* **23** (12), 24330–24342. <https://doi.org/10.1109/TITS.2022.3203715> (2022).
- Ning, T., Gao, Y. & Han, Y. Segmentation of ethnic clothing patterns with fusion of multiple attention mechanisms. *Complex. Intell. Syst.* **10** (4), 5759–5770. <https://doi.org/10.1007/s40747-024-01457-5> (2024).
- Chen, R., Chen, Y., Yu, K. & Xu, Z. A segmentation method for virtual clothing effect images. *Ind. Textila.* **76** (2), 230–236. <https://doi.org/10.35530/IT.076.02.2024111> (2025).
- Liu, L., Xu, X., Lin, Z., Liang, J. & Yan, S. Towards garment sewing pattern reconstruction from a single image. *ACM Trans. Graph.* **42** (6), 1–15. <https://doi.org/10.1145/3618319> (2023).
- Yu, F. et al. Phase contour enhancement network for clothing parsing. *IEEE Trans. Consum. Electron.* **70** (1), 2784–2793. <https://doi.org/10.1109/TCE.2024.3377377> (2024).
- Seidu, R. K., Choi, S. Y. & Jiang, S. Development and performance of Jacquard woven retro-reflective textiles with African design patterns. *Fash Text.* **10** (1), 1–20. <https://doi.org/10.1186/s40691-022-00322-8> (2023).

22. Cao, J. et al. Automatic exploration and transfer design of associative rules in she ethnic clothing coloration. *Multimed Tools Appl.* **84** (12), 10269–10289. <https://doi.org/10.1007/s11042-024-19357-8> (2025).
23. Windhager, J. et al. An end-to-end workflow for multiplexed image processing and analysis. *Nat. Protoc.* **18** (11), 3565–3613. <https://doi.org/10.1038/s41596-023-00881-0> (2023).
24. Conze, P. H., Andrade-Miranda, G., Singh, V. K., Jaouen, V. & Visvikis, D. Current and emerging trends in medical image segmentation with deep learning. *IEEE Trans. Radiat. Plasma Med. Sci.* **7** (6), 545–569. <https://doi.org/10.1109/TRPMS.2023.3265863> (2023).
25. Habib, M. A. & Alam, M. S. A comparative study of 3D virtual pattern and traditional pattern making. *J. Text. Sci. Technol.* **10** (1), 1–24. <https://doi.org/10.4236/jtst.2024.101001> (2024).
26. Amelia, N. & Lubis, N. A. Ethnomathematics exploration of Malay traditional clothing patterns in the concept of two-dimensional plane figure. *Unnes J. Math. Educ.* **13** (2), 133–141. <https://doi.org/10.15294/4324n131> (2024).
27. Xiao, H. Optimized soft frame design of traditional printing and dyeing process in Xiangxi based on pattern mining and edge-driven scene Understanding. *Soft Comput.* **26** (23), 12997–13008. <https://doi.org/10.1007/s00500-021-06201-6> (2022).
28. Meng, L. et al. Deep learning-enhanced jewelry material jadeite jade quality assessment. *JOM* **77** (1), 211–224 <https://doi.org/10.1007/s11837-024-06930-7> (2025).
29. Fan, Y., Wen, B. & Deng, H. MRA-Net: an instance segmentation method based on multi-scale feature fusion for ethnic costumes images. *Vis. Comput.* **41** (8), 5949–5960. <https://doi.org/10.1007/s00371-024-03762-z> (2025).
30. Li, Y., Zhou, Z., Sun, C., Chen, X. & Yan, R. Variational attention-based interpretable transformer network for rotary machine fault diagnosis. *IEEE Trans. Neural Netw. Learn. Syst.* **35** (5), 6180–6193. <https://doi.org/10.1109/TNNLS.2022.3202234> (2022).
31. Wan, Y., Zou, G., Yan, C. & Zhang, B. Dual attention composition network for fashion image retrieval with attribute manipulation. *Neural Comput. Appl.* **35** (8), 5889–5902. <https://doi.org/10.1007/s00521-022-07994-9> (2023).
32. Nguyen, Q. T. Defective sewing stitch semantic segmentation using DeepLabV3+ and EfficientNet. *Intelig. Artif. Opt.* **25** (70), 64–76 <https://doi.org/10.4114/intartif.vol25iss70pp64-76> (2022).
33. Liu, J., Luo, H. & Tu, D. Underwater motion scene image restoration based on an improved U-Net network. *Appl. Opt.* **63** (1), 228–238. <https://doi.org/10.1364/AO.505198> (2023).
34. Xiang, Z., Zhu, C., Qian, M., Shen, Y. & Shao, Y. FashionSegNet: a model for high-precision semantic segmentation of clothing images. *Vis. Comput.* **40** (3), 1711–1727. <https://doi.org/10.1007/s00371-023-02881-3> (2024).
35. Ren, Q., Lu, Z., Wu, H., Zhang, J. & Dong, Z. HR-Net: A landmark based high realistic face reenactment network. *IEEE Trans. Circuits Syst. Video Technol.* **33** (11), 6347–6359. <https://doi.org/10.1109/TCSVT.2023.3268062> (2023).

## Author contributions

D.D.H processed the numerical attribute linear programming of communication big data, and the mutual information feature quantity of communication big data numerical attribute was extracted by the cloud extended distributed feature fitting method. B.X. and H.L. Combined with fuzzy C-means clustering and linear regression analysis, the statistical analysis of big data numerical attribute feature information was carried out, and the associated attribute sample set of communication big data numerical attribute cloud grid distribution was constructed. D.D.H. and H.L. did the experiments, recorded data, and created manuscripts. All authors read and approved the final manuscript.

## Funding

The research is supported by Pingdingshan University Doctoral Scientific Research Startup Fund Project - Research on Intelligent Recognition and Restoration Technology of Image Textures of Cultural Relics Fabrics, (No. PXY-BSQD-2026016); Henan Province Science and Technology Research Project - Research on Key Technologies of Photovoltaic Module Defect Detection Based on Generative Adversarial Networks (No. 252102210014).

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to D.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025