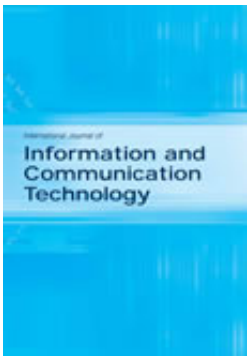


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Optimisation of intelligent recognition teaching system for Miao costume patterns integrating YOLOv5

Qiong Luo, Xubing Xu, Jan Zhou

DOI: 10.1504/IJICT.2025.10075238

Article History:

Received:	29 September 2025
Last revised:	29 October 2025
Accepted:	31 October 2025
Published online:	12 January 2026

Optimisation of intelligent recognition teaching system for Miao costume patterns integrating YOLOv5

Qiong Luo

College of Fashion and Art Design,
Donghua University,
Shanghai, 200000, China
and
School of Art and Design,
Anhui Institute of Information Technology,
Wuhu, 241000, China
Email: luoqiong@mail.dhu.edu.cn

Xubing Xu*

College of Fashion and Art Design,
Donghua University,
Shanghai, 200000, China
Email: 15305537665@163.com
*Corresponding author

Jan Zhou

School of Art and Design,
Anhui Institute of Information Technology,
Wuhu, 241000, China
Email: 13739298136@163.com

Abstract: This study builds a high-quality Miao pattern dataset, then applies label smoothing and mosaic data augmentation. To maximise multi-scale feature fusion, the spatial pyramid pooling fast (SPPF) module is utilised. Increasing the precision of bounding box regression and small target recognition, the focal loss and complete IoU Loss algorithms are combined. A web-based visual teaching platform is created with features for displaying cultural knowledge and inferring models. The research results indicate that, the enhanced YOLOv5 model outperforms comparable models like faster R-convolutional neural network and YOLOv4 with mean average precision@0.5 of 89.6% and mAP@0.5:0.95 of 61.5% on the test set. Compared to the original YOLOv5s, it has increased by 5.4% and 7.2% respectively. Meantime, the recall rate improvement in small pattern detection is greater than 6%, which is better than that of baseline models such as YOLOv4. The data confirm deep learning's potential in high-precision ethnic culture recognition and instruction.

Keywords: Miao costume patterns; YOLOv5; target detection; transformer attention mechanism; intelligent teaching system.

Reference to this paper should be made as follows: Luo, Q., Xu, X. and Zhou, J. (2025) 'Optimisation of intelligent recognition teaching system for Miao costume patterns integrating YOLOv5', *Int. J. Information and Communication Technology*, Vol. 26, No. 49, pp.57–74.

Biographical notes: Qiong Luo obtained her Bachelor's in Environmental Design from Anhui University of Engineering in 2009 and Master's in Design and Art from Anhui University of Engineering in 2012. She is currently pursuing a Doctoral in Design at the School of Fashion and Art Design, Donghua University in Shanghai. Since 2012, she has been teaching at the School of Art and Design, Anhui University of Information Technology. Her research interests include traditional clothing pattern design, digitalisation of intangible cultural heritage pattern design, and smart teaching to a reference.

Xubing Xu received his BS and MS in Aerospace Engineering from the University of Virginia, Charlottesville in 2001 and PhD in Mechanical Engineering from Drexel University, Philadelphia, PA in 2008. He graduated from Jiangnan University with a major in Fashion Design in July 1988. From 1996 to 2001, he studied abroad at Kyoto University of Arts and Crafts and the School of Japanese Culture and Fashion. He graduated with a Master's degree and became an Associate Professor at the School of Fashion and Art Design at Donghua University in September 2002. In September 2017, became a Professor and appointed as a Doctoral Supervisor in 2019.

Jan Zhou obtained his Bachelor's in Industrial Design (Engineering) from Anhui University of Engineering in 2008 and Master's in Design and Art from Jiangnan University in 2011. Since 2012, he has been teaching at the School of Art and Design, Anhui University of Information Technology. His research interests include interaction design and user research. He is a design expert at iFlytek Co., Ltd. and a Judge for the China Packaging Creative Design Competition. He is a national committee member of the Design Professional Committee of China Packaging Federation.

1 Introduction

Miao costume patterns are important carriers of the intangible cultural heritage (ICH) of the Chinese nation. Their unique geometric forms, natural images and cultural symbols contain profound historical memories and ethnic identity values (Tang et al., 2023; Jung and Choi, 2022). With the development of digital protection technology, image recognition methods based on deep learning have become an important path for intelligent research on cultural heritage (Sun et al., 2022; Alsuwaylimi et al., 2024). However, Miao patterns are characterised by complex structures, variable scales and small inter-class differences. Traditional detection models such as faster R-convolutional neural network (CNN) and the basic You Only Look Once (YOLO) series still have shortcomings in feature extraction, small target detection and multi-scale adaptability, resulting in limited recognition accuracy, which makes it difficult to support high-demand cultural education application scenarios (Olorunshola et al., 2023).

In the current cultural teaching practice, the teaching of Miao ethnic clothing patterns still heavily relies on manual identification and oral explanation, and has limitations like low teaching efficiency and insufficient standardisation of knowledge. Additionally, the

younger generation's understanding of traditional culture is becoming increasingly distant. This makes it difficult to effectively convey and quantify the deep cultural connotations such as totemic symbols and historical memories contained in patterns. Traditional teaching methods lack efficient visual interaction and technical assistance when faced with the high complexity of pattern shapes and rich cultural semantics. Therefore, intelligent recognition technology needs to be introduced to achieve digital preservation of cultural knowledge and immersive teaching experience.

This study suggests an enhanced and integrated approach for the intelligent identification and instruction of Miao costume patterns in order to address the above problems. The following three areas demonstrate its innovations:

1 To improve the YOLO version 5 (YOLOv5) model's capacity for collaborative perception of the global semantics and local aspects of patterns, incorporate the transformer attention mechanism and weighted k-means clustering anchor box design.

2 Construct a detection framework that integrates the spatial pyramid pooling fast (SPPF) multi-scale fusion (MSF) module and the focal loss-complete IoU loss (CIoU) combined loss function to improve the robustness to complex backgrounds and small-sized patterns.

3 Develop a teaching system including a web front-end and model inference service to realise the integrated functions of pattern recognition, cultural interpretation and interactive learning.

This study systematically expounds the model optimisation strategy, system construction method and experimental verification results, aiming to provide a promotable technical solution for the digital inheritance of ICH.

2 Related work

Deep learning-based object detection technology has shown wide-ranging application promise in clothing-related domains in recent years, offering a crucial theoretical point of reference and methodological underpinning for the study. In the field of general clothing recognition, Chang and Zhang used YOLOv5 to realise the automatic recognition of clothing styles, verifying the efficiency of single-stage detectors in clothing image classification tasks (Chang and Zhang, 2022). By altering the network structure and loss function of YOLOv5, Li et al. (2024) and Zhou et al. (2023) enhanced the detection accuracy and robustness of fabric flaws, respectively. Their optimisation strategies for small targets provided important inspiration for this study (Li et al., 2024; Zhou et al., 2023). In the field of clothing part recognition, Cao and Tuo (2024) combined YOLOv5 with ResNet50 to achieve accurate recognition of women's clothing sleeves, proving the effectiveness of multi-model fusion in complex clothing structure analysis. From the perspective of auxiliary technology, Rocha et al. (2023) developed a clothing defect recognition system for visually impaired people using object detection technology, demonstrating the social value of computer vision technology. In industrial detection scenarios, Nguyen et al. (2024) conducted heuristic optimisation for the detection of construction site protective equipment based on YOLOv5. Yusro et al. (2023) compared YOLOv5 and faster R-CNN's performance in overlapping object recognition. Both

provided technical references for solving problems such as dense patterns and occlusions in this study (Nguyen et al., 2024; Yusro et al., 2023). In the agricultural field, Moya et al. (2024) applied the YOLOv5 image analysis method in crop detection, further expanding the application boundary of this model under complex backgrounds.

In summary, although existing research has provided many improvement ideas for general clothing detection, there are still significant limitations in dealing with Miao clothing patterns with high cultural specificity:

1 The lack of high-quality, finely annotated specialised datasets leads to insufficient recognition accuracy of the model in complex backgrounds and small sample situations.

2 Improvement strategies often focus on the model itself and fail to fully consider the global semantic and cultural contextual features of patterns.

3 Most research results remain at the algorithmic level, lacking support from visual teaching platforms that are deeply integrated with educational practice.

Based on the above limitations, this study proposes a model for constructing a dedicated dataset, integrating attention mechanisms and multi-scale optimisation. It is applied to a web teaching platform to fill the gap between algorithm innovation and cultural education practice.

3 Intelligent recognition model of Miao costume patterns based on YOLOv5

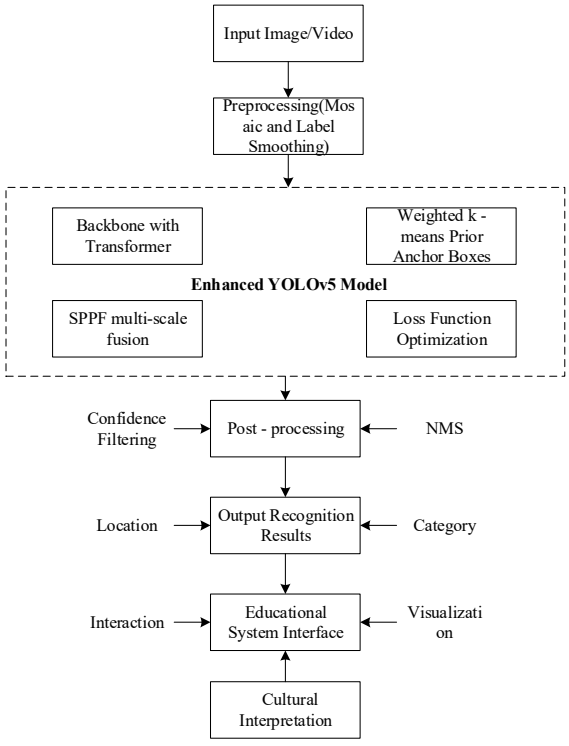
3.1 Overall framework of intelligent recognition model for Miao costume patterns

The proposed intelligent recognition and teaching system for Miao costume patterns takes the YOLOv5 object detection model as the core and combines a variety of deep learning optimisation strategies to construct an end-to-end pattern recognition and teaching assistance platform. The entire system is divided into four modules: the data input and pre-processing module, the pattern detection model module, the post-processing and result analysis module, and the teaching system integration module. Its overall architecture is shown in Figure 1, where non-maximum suppression (NMS) stands for NMS.

Figure 1 shows the end-to-end processing flow of the model from data input to result output. The data input and pre-processing module is responsible for receiving image or video streams and performing size scaling, normalisation, and other operations to provide standardised input for the model. The pattern detection model module serves as the core, and its optimisation includes: in backbone, the C3TR module enhances the model's ability to model the global context of patterns by introducing transformer self-attention mechanism; the SPPF module improves the efficiency of feature extraction for patterns of different scales by performing multi-scale pooling and fusion on feature maps; the prior boxes generated by weighted k-means provide more matching initial prediction references for detection layers of different scales in the detection head. Subsequently, the post-processing and analysis module mainly uses NMS to remove redundant detection boxes and obtain the final pattern category and position information. Finally, all results

are transmitted to the teaching system integration module for web visualisation display and cultural knowledge association.

Figure 1 Model framework of intelligent recognition of Miao costume patterns (see online version for colours)



3.2 The construction and enhancement of Miao costume pattern dataset

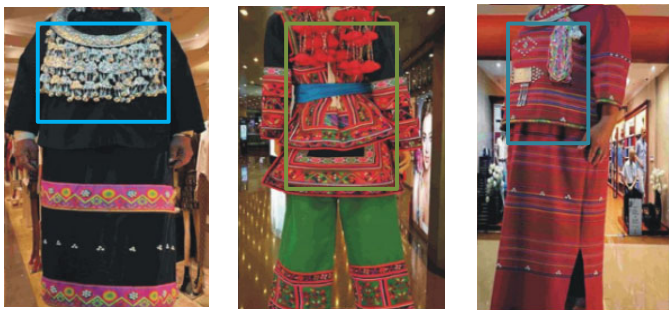
3.2.1 Data acquisition and labelling

For specific targets like Miao costume patterns, which feature rich cultural details and diversity, this study first constructs a high-quality and large-scale dataset. This dataset takes the National Culture Instruction-Following Dataset (NCIFD) as the base dataset. The NCIFD dataset contains 151,159 data entries, including images in seven fields: architecture, costumes, craftsmanship, food, etiquette, language, and customs. Among them, this study selects a large number of images of traditional Chinese ethnic costumes (13,900 images in total) from the costume section (https://github.com/letsgoLakers/NCIFD).

After obtaining the original images, this study uses the LabelImg open-source tool for manual annotation. Then, according to the characteristics of Miao costume patterns, category labels including ‘geometric patterns’ and ‘animal patterns’ are set. Annotators draw bounding boxes according to the positions and categories of patterns in the images and selected corresponding category labels. The annotation results are saved in a text format supported by YOLOv5. Each text file contains the category indices of all target

boxes in the corresponding image, as well as the normalised coordinates of the centre points, width and height information. Figure 2 shows some examples of annotated images.

Figure 2 Examples of some images after labelling (see online version for colours)



Following the completion of all annotations, the dataset is split at random into an 8:2 training and validation set. While the validation set assesses model performance, modifies hyperparameters, and avoids overfitting during training, the training set learned the model's parameters. Segmentation of the Miao costume pattern dataset (after screening) is displayed in Table 1.

Table 1 Classification of Miao costume pattern dataset (after screening)

Classification name	*Number of training set images*	*Number of verification set images*	*Total*
Geometric patterns	2,664	666	3,330
Animal patterns	2,736	684	3,420
Plant patterns	2,600	650	3,250
Embroidery patterns	1,600	400	2,000
Silver ornaments	1,520	380	1,900
Total	11,120	2,780	13,900

In addition, this study optimises the annotation results through multi person collaborative annotation and cross review: three annotation personnel with image recognition backgrounds participated to develop detailed annotation specifications, and clarify the definition standards and boundary box drawing principles for various patterns. During the annotation process, each image is initially annotated by one annotator and then independently reviewed by another annotator. Controversial annotation samples are discussed and agreed upon. Finally, 10% of the annotated results are randomly selected for sampling inspection by domain experts to ensure the accuracy and consistency of the annotated results.

3.2.2 Mosaic data enhancement and label smoothing strategy

Following the creation of a high-quality dataset of Miao costume patterns, this study uses two training procedures to enhance the model's resilience and generalisation skills and avoid overfitting to a limited number of samples during the training phase: enhancement

of mosaic data and label smoothing (Mao et al., 2025; Zhao et al., 2023). During the model training process, data augmentation and label smoothing strategies work together from both the data and label levels to enhance the robustness of the model. Data augmentation enhances the adaptability of models to complex real-world situations by creating complex composite images with multiple scales and backgrounds, forcing them to learn how to locate and recognise targets in different contextual environments. Meanwhile, label smoothing softens hard labels, reduces the model's overconfidence in training samples, encourages it to learn more generalised feature boundaries, and effectively prevents overfitting. The combination of these two strategies aims to systematically improve the generalisation performance of the model on unknown data.

Mosaic data augmentation is an efficient data augmentation technology widely used in YOLOv4 and YOLOv5 (Mathew and Mahesh, 2022). The fundamental idea behind it is to randomly scale, crop, and arrange four distinct training photos into a new composite image, combining the matching bounding boxes in the process (Wang et al., 2022). At the label processing level, the label smoothing regularisation technology is introduced (Zhou et al., 2023). Among them, label smoothing improves generalisation ability by 'softening' this hard label distribution.

$$q'(k) = (1-\epsilon) * \delta_{k,y} + \epsilon / K \tag{1}$$

In equation (1), q'(k) represents the smoothed label distribution. $\delta_{k,y}$ is an indicator function, which equals 1 when category k is the true category y, and 0 otherwise. ϵ is a small smoothing hyperparameter (usually set to 0.1) used to control the degree of smoothing. K denotes the total number of categories in the dataset.

In the specific implementation, the random scaling ratio range of each input image is set to 0.5 to 1.5 times to ensure the diversity of training samples in scale. Subsequently, a designated area (640 × 640) is randomly cropped from the scaled image, and the four cropped sub-images are randomly laid out and pieced together to form a new composite image. Moreover, the target box coordinates corresponding to the image are proportionally offset and adjusted based on the relative position of each subgraph in the synthesised image, thereby generating multi-target and multi-scale training samples with precise annotations for the model.

3.3 Optimisation of pattern detection model based on YOLOv5

3.3.1 Backbone network with transformer attention mechanism

The YOLOv5 object detection model's backbone network is primarily in charge of retrieving multi-level feature information from the input photos. When analysing Miao costume patterns with intricate structures and long-distance dependencies, classic CNN backbone networks have considerable limitations in their ability to model the global environment, despite their ability to capture local characteristics efficiently (Sadiq et al., 2022; Dewi et al., 2022). This study incorporates a Transformer module based on the SAM into the YOLOv5 backbone network.

The SAM at the heart of the transformer design is capable of determining the dependency relationship between any two elements in a sequence and calculating attention weights using the three vectors of query, key, and value (Moustapha et al., 2023). The following is computation procedure, as shown in equation (2).

$$\text{Att}(t, s, e) = \text{soft}\left(\frac{ts^T}{\sqrt{d_s}}\right)e \tag{2}$$

t, s and e are matrices obtained by linear transformation of input sequence. d_s is the dimension of key vector. $\sqrt{d_s}$ plays a scaling role to prevent the gradient of softmax function from disappearing due to excessive dot product.

By repeatedly using the aforementioned attention process, multi-head (MH) attention enables the model to focus on data from various representation subspaces, as shown in equations (3)–(4).

$$\text{MH}(t, s, e) = \text{Concat}\left(\text{h}_1, ..., \text{h}_h\right)W^O \tag{3}$$

$$\text{where head}_i = \text{Att}\left(tW_i^t, sW_i^s, eW_i^e\right) \tag{4}$$

W_i^t, W_i^s, and W_i^e are the projection weight matrices corresponding to the i^{th} head. W^O is the output projection matrix, and h is the number of attention heads.

Following the C3 module in the centre of the backbone, this study presents a simpler transformer encoder block based on the attention mechanism mentioned before. This block includes a feed-forward neural network and a MH self-attention layer. To guarantee training stability, layer normalisation and residual connections are performed around each layer. This improved module is referred to as the C3TR module in this study, which replaces one of the C3 modules in the original YOLOv5s model.

The C3TR module maintains the same input and output feature map dimensions as the original C3 module in its design, ensuring that it can replace and embed into the backbone network of YOLOv5s. Specifically, after serialising the input feature map, the module first captures the global contextual relationship through a MH self-attention layer (with four heads), and then uses a feedforward neural network to perform nonlinear transformation on it. In addition, this study verifies the effectiveness of the C3TR module in enhancing the model's ability to perceive global pattern structures by comparing the mean average precision (mAP) improvement results on the validation set before and after introducing the module.

3.3.2 Weighted k-means clustering for prior box generation

In object detection models, the size and ratio of prior anchor boxes directly affect the efficiency of the model in regressing target positions. YOLOv5 generally uses general anchor boxes obtained by clustering based on the common objects in context (COCO) dataset, but there is a significant difference between these anchor boxes and the actual size distribution of Miao costume patterns. Building on this, this study employs a weighted k-means clustering algorithm to automatically derive more appropriate prior anchor box sizes from the established dataset (Liu et al., 2023; Xu et al., 2023).

The traditional k-means clustering uses Euclidean distance as the measurement standard, but this will lead to larger errors for large-sized bounding boxes than for small ones. However, the evaluation standard for object detection is intersection over union (IoU) (Kim et al., 2023). Therefore, this study adopts 1 – IoU as the distance metric for clustering.

$$d(box, centroid) = 1 - \text{IoU}(box, centroid) \tag{5}$$

In equation (5), *box* represents the bounding box with real annotations in the dataset, and *centroid* represents the cluster canter (i.e., a candidate anchor box). IoU represents the ratio of the area where two boxes overlap (intersection area) to the total area they cover together (union area). Finding k cluster centres is the aim of this computation, which aims to minimise the sum of the distances between each bounding box and its corresponding cluster centre.

The standard k-means algorithm assumes that each sample contributes equally to the cluster centre. The amount of pattern instances in the dataset used for this study, however, differs depending on the category. Direct clustering will lead to categories with a larger number of instances dominating the clustering results. To address this issue, this study introduces weighted k-means. This algorithm assigns a weight to each sample, and during the clustering process, samples with higher weights will have a greater impact on the position of the cluster centre. Therefore, by associating a weight ω_i with each bounding box, the objective function becomes minimising the sum of weighted distances, as shown in equation (6).

$$D = \sum_{i=1}^{N} \omega_i \cdot d\left(box_i, centroid\right) \tag{6}$$

N is the total number of bounding boxes in the dataset, and ω_i is the weight of the i^{th} bounding box. The setting of weights mainly involves assigning higher weights to small-sized patterns. This is to encourage cluster centres to better match these small targets, thereby alleviating the common challenge of small target detection in object detection tasks. The weight ω_i of each bounding box is inversely proportional to its area and can be calculated using equation (7):

$$\omega_i = 1 \Big/ \sqrt{\text{Area}\left(box_i\right)} \tag{7}$$

This strategy aims to ensure that smaller target boxes have greater weight influence during the clustering process, thereby guiding the generated anchor boxes to more densely cover the large number of small-sized patterns present in the dataset. Subsequent studies will validate the effectiveness of this strategy by comparing the model performance under different weighting schemes in ablation experiments.

Based on the above clustering process, this study obtains nine prior anchor boxes, which are assigned to the three detection layers of YOLOv5 according to their scales. These three detection layers are respectively responsible for detecting small-sized, medium-sized, and large-sized pattern targets.

3.3.3 Multi-scale feature fusion optimisation

Miao costume patterns exhibit significant scale variations in images, where large layout patterns and fine tiny embroideries often coexist. To ensure the model can effectively detect targets of these different scales simultaneously, this study optimises the multi-scale feature fusion module. YOLOv5 itself adopts the structure of feature pyramid network (FPN) and path aggregation network (PAN). However, based on this foundation, this study replaces the original spatial pyramid pooling (SPP) module with a more efficient SPPF module.

SPP module performs parallel pooling operations using multiple maximum pooling kernels of different sizes, then concatenates the results to fuse features under different

receptive fields. However, multiple large-kernel pooling layers incur high computational costs. The core idea of the SPPF module is to simulate the effect of large-kernel pooling by connecting multiple small-sized pooling kernels in series. Specifically, the SPPF module uses multiple 5×5 maximum pooling layers connected in series: two consecutive 5×5 pooling layers are equivalent to one 9×9 pooling layer in terms of receptive field, and three consecutive 5×5 pooling layers are equivalent to one 13×13 pooling layer. This design can reduce the amount of computation. Its processing process is as follows:

$$Y = \mathrm{Concat}\left(X, P_5(X), P_5\left(P_5(X)\right), P_5\left(P_5\left(P_5(X)\right)\right)\right) \tag{8}$$

In equation (8), X denotes the input feature map. $P_5(\cdot)$ represents the maximum pooling operation with a kernel size of 5×5, and Concat stands for the concatenation operation along the channel dimension. The final output Y integrates the original input and features after one, two, and three rounds of pooling, thereby capturing multi-scale contextual information.

While shallow-layer features provide accurate localisation but insufficient semantic information, deep-layer features in the FPN+PAN structure have rich semantic information but limited resolution. Multi-scale feature fusion aims to effectively combine feature maps from different layers. A simplified feature fusion operation can be expressed as:

$$F_o = Conv\left(\left[F_h \| F_l\right]\right) \tag{9}$$

In equation (9), F_h refers to the deep-layer feature map from a higher network level. F_l denotes the shallow-layer feature map from a lower network level. Upsample represents the upsampling operation, which is used to adjust the resolution of F_h to match that of F_l. Concat performs concatenation of the two feature maps along the channel dimension. Finally, a convolution layer is applied for fusion and dimension reduction, outputting the fused feature map F_o.

3.3.4 Loss function optimisation

The original YOLOv5 loss function consists of three parts: classification loss, object confidence loss, and bounding box regression loss. To enhance the model's performance in the Miao costume pattern identification task, this study implements targeted adjustments to the loss function to optimise the detection effect of small targets under complex backgrounds. The primary optimisations include using focal loss to improve the classification branch and CIoU Loss to replace the original IoU loss for bounding box regression.

Through the addition of a modulation factor to traditional cross-entropy loss, focal loss dynamically lessens the impact of simple samples on the total loss, freeing the model up to concentrate more on the learning of hard data. The following is its equation:

$$FL\left(y_t\right) = -m_t\left(1 - y_t\right)^{\tau} \log\left(y_t\right) \tag{10}$$

In equation (10), y_t is the probability that the model predicts the target belongs to the true class. m_t is a balancing factor used to balance the weights of positive and negative samples. τ is an adjustable parameter that controls the rate at which the weights of simple samples are reduced. When $\tau = 0$, focal loss degenerates into the standard cross-entropy

loss. As τ increases, the effect of the modulation factor $(1 - y_t)^\tau$ is enhanced, the loss of samples correctly classified by the model (with large y_t) is significantly reduced, while the loss of hard-to-classify samples (with small y_t) remains relatively unchanged.

CIoU Loss, as a regression loss, comprehensively considers three geometric factors: overlapping area, centre point distance, and aspect ratio. The following is its definition:

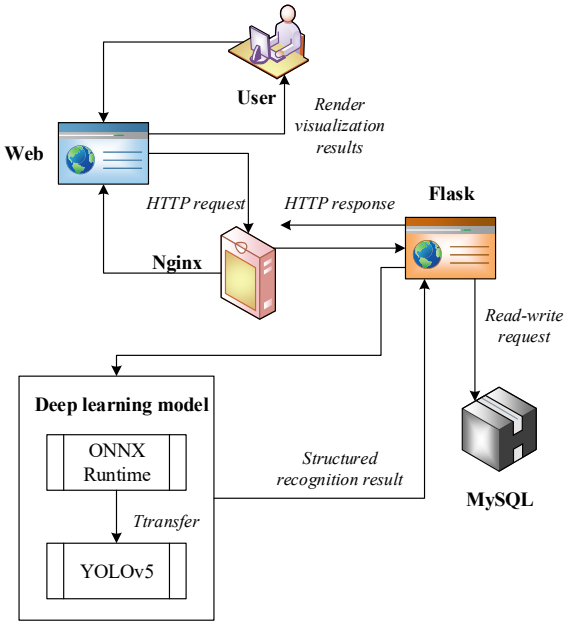
$$\mathcal{L}_{CIoU} = 1 - IoU + \frac{\rho^2\left(d, d^{rt}\right)}{D^2} + w\varphi \tag{11}$$

In equation (11), ρ calculates the Euclidean distance between the centre point d of the predicted box and the centre point d^{rt} of the ground truth box. D is the diagonal distance of the smallest enclosing region that contains both the predicted box and the ground truth box. φ is a parameter used to measure the consistency of the aspect ratio, and w is a weight coefficient for balancing φ.

3.4 *Teaching system integration based on web visualisation and interactive design*

To transform the optimised YOLOv5 Miao costume pattern recognition model into a practically usable teaching tool, this study designs and implements a web-based visual interactive system. The goal of this system is to provide users with an intuitive, easy-to-use, and feature-rich platform, enabling them to experience pattern recognition technology and gain in-depth insights into the underlying ethnic cultural connotations. The system's traditional front-end and back-end separation architecture guarantees outstanding user experience, scalability, and maintainability. In Figure 3, the system's overall technical architecture is displayed.

Figure 3 Teaching system architecture based on web visualisation and interactive design (see online version for colours)



The system uses Vue.js to build the front-end interface, which includes four modules: an image upload area, a real-time video detection area, a result visualisation display area, and a pattern cultural knowledge base. It supports drag-and-drop/click upload or calling the camera for real-time detection. Results are annotated with bounding boxes indicating categories and confidence levels, with associated display of cultural backgrounds of patterns. The back-end provides a representational state transfer application programming interface (RESTful API) based on Flask, which invokes the YOLOv5 model [stored in open neural network exchange (ONNX) format and accelerated by the ONNX Runtime inference engine] for prediction. User data is stored in a My Structured Query Language (MySQL) database. The workflow is as follows: the front-end sends requests via the hyper text transfer protocol (HTTP), the back-end pre-processes the images before model inference, and the results are returned to the front-end in JavaScript Object Notation (JSON) format for rendering.

4 Experimental results and analysis

4.1 Experimental environment setting

A workstation with an NVIDIA graphics processing unit (GPU) is used for the study in order to thoroughly assess how well the optimised YOLOv5 model performs in the Miao costume pattern detection challenge. The specific hardware configuration and software environment are shown in Table 2.

Table 2 Configuration of experimental hardware and software environment

Component type	*Configuration details*
Central processing unit (CPU)	Intel Core i9-10900K @ 3.70 GHz
GPU	NVIDIA GeForce RTX 3080 Ti (12 GB GDDR6X)
Memory	64 GB DDR4 3200MHz
Operating system	Ubuntu 20.04.4 LTS
Deep learning framework	PyTorch 1.9.0 + CUDA 11.1
Programming language	Python 3.8
Acceleration library	cuDNN 8.0.5

Hyperparameter setting used in model training is beneficial to reproduce the experimental results. Based on this, this study follows the conventional training strategy of YOLOv5 model, and makes targeted adjustments to some key parameters. The specific configuration is shown in Table 3.

The above hyperparameters are selected based on the conventional training strategy of YOLOv5, and targeted adjustments and verifications have been made according to the characteristics of the task in this study. The key parameters such as initial learning rate and weight decay are referenced from the pre-training experience of the model on the COCO dataset. Based on this, local optimisation is carried out through grid search and pre-experiments. The batch size is mainly limited by the GPU memory capacity and is set to 16 to achieve a balance between training efficiency and stability. The learning rate scheduler adopts cosine annealing strategy to achieve smooth convergence at the end of

training. All hyperparameter settings are validated for their effectiveness in this task through controlled variable experiments, ensuring optimal model performance.

Table 3 Model training hyperparameter setting

Hyperparameter	*Setting value*
Input image size	640 × 640
Batch size	16
Training cycle	300
Optimiser	Stochastic gradient descent (SGD)
Initial learning rate	0.01
Momentum	0.937
Weight attenuation	0.0005
Learning rate scheduler	Cosine annealing

This study employs industry-standard criteria for object detection to gauge the model's performance based on assessment metrics. The key evaluation measures are frames per second (FPS) and mAP. mAP is calculated with various IoU thresholds: mAP@0.5 is defined as the average precision when the IoU threshold is 0.5, and it accounts for detection accuracy; mAP@0.5:0.95 stands for the average mAP value in the IoU threshold interval of 0.5 to 0.95, which is used to evaluate the model's performance under different positional accuracy specifications. FPS is a crucial metric for assessing the algorithm's viability and is used to assess the model's inference speed on certain hardware.

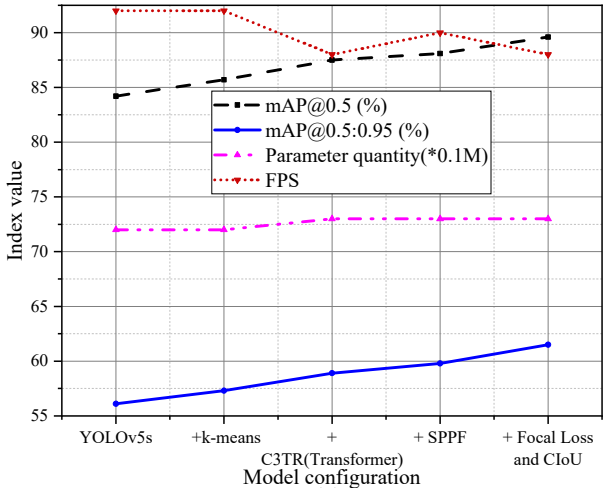
4.2 Performance verification of intelligent recognition model for Miao costume patterns

4.2.1 Ablation experiment

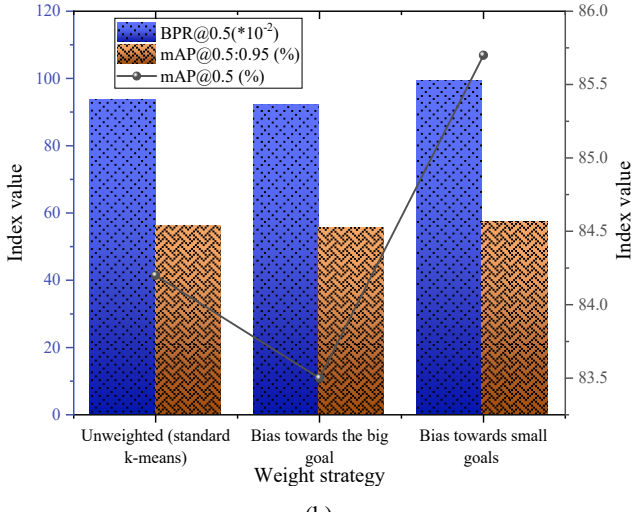
Ablation experiments are carried out to confirm the precise contribution of each optimisation method to the Miao costume pattern recognition model's performance. The experiments use the standard YOLOv5s model as the baseline. Under the same training dataset and experimental environment, weighted k-means anchor boxes; transformer attention mechanism (C3TR), SPPF module, and the combination of focal loss and CIoU loss function are introduced sequentially. Finally, a complete optimised model is obtained, and the changes in various performance indicators are recorded. The results are shown in Figure 4.

Figure 4 shows that based on the YOLOv5s baseline model (mAP@0.5 84.2%), after sequentially introducing weighted k-means anchor boxes, C3TR transformer, SPPF, and the combination of focal loss and CIoU, the fully optimised model achieves an mAP@0.5 of 89.6%. The parameter count is stably maintained at 7.3M, and the FPS is kept between 88 and 92. Among these, the weighted k-means strategy optimised for small targets (BPR@0.5 = 0.992) achieves a 3.2% gain in small target detection compared with standard k-means. However, the strategy biased toward large targets leads to a 0.9% decrease in performance. This confirms the proposed weight assignment strategy's efficacy.

Figure 4 Ablation experimental results, (a) ablation experiment results of different module combinations (on the test set), (b) k-means clustering with different weighting strategies and model performance comparison (see online version for colours)



(a)



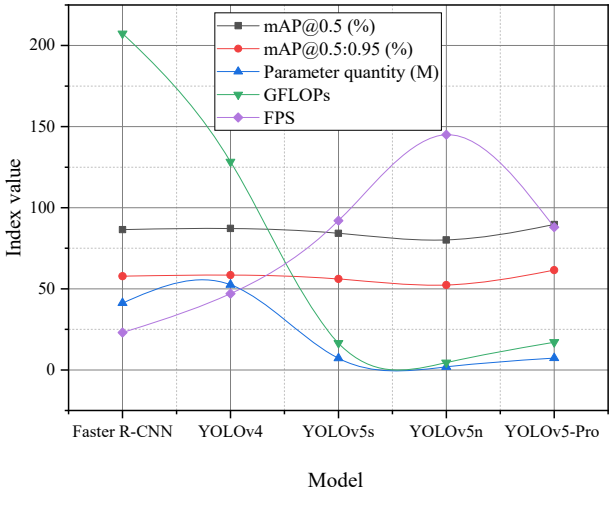
(b)

4.2.2 Comparison experiment

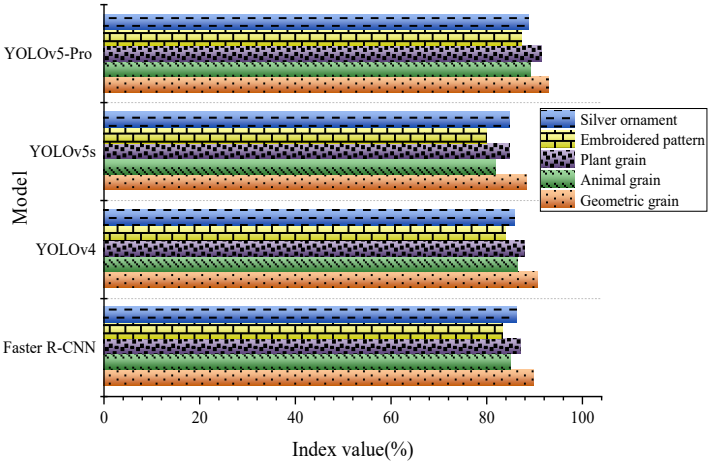
Using the same Miao costume pattern test set, this study compares the performance of the suggested optimised model (henceforth referred to as YOLOv5-Pro) with existing popular object identification algorithms in order to fully assess its overall effectiveness. YOLOv4 and the original YOLOv5s from the YOLO series of single-stage detectors, a lightweight network called YOLOv5n, and faster R-CNN, a traditional representation of two-stage detectors, are among the models included in the comparison. All comparable models are trained using the same training set and employ the same data augmentation

methods. To guarantee the fairness of the comparison, their performance is ultimately assessed using the same test set. Figure 5 presents the findings of the thorough performance comparison.

Figure 5 Comparison experiment, (a) the performance comparison of different target detection models on Miao clothing pattern test set, (b) comparison of AP@0.5 of different models on various patterns (%)(see online version for colours)



(a)



(b)

Figure 5 shows that YOLOv5-Pro achieves better overall performance than other models on the Miao costume pattern test set. Its mAP@0.5 reaches 89.6% (higher than 87.2% of YOLOv4 and 86.5% of Faster R-CNN), and its mAP@0.5:0.95 is 61.5%. In the detection of various patterns, YOLOv5-Pro also performs better in terms of AP@0.5, with outstanding performance in geometric patterns (93.2%) and plant patterns (91.8%). Compared with YOLOv5s, these two indicators increase by 4.7% and 6.8% respectively,

which verifies its strong generalisation ability for complex patterns. Despite having a slightly lower frame rate than YOLOv5n, the model's great accuracy and efficiency make it appropriate for practical use.

4.3 Display and analysis of application effect of teaching system

This study carries out a user experience test to confirm the efficacy of the intelligent recognition training system for Miao costume patterns integrated with YOLOv5 in practical application settings. The evaluation results are shown in Figure 6.

Figure 6 Evaluation results of system application effect (see online version for colours)

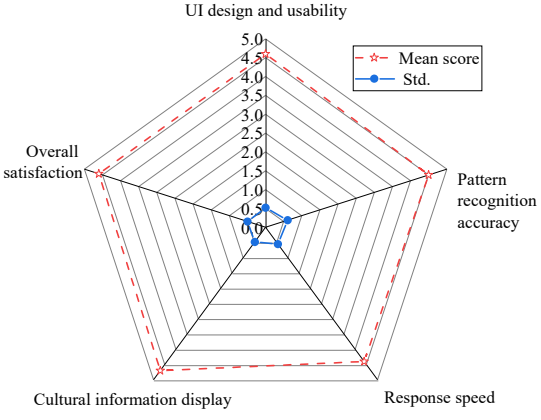


Figure 6 shows that the intelligent recognition teaching system for Miao costume patterns integrated with YOLOv5 has received high evaluations among 30 non-professional users. The average score of each dimension is higher than 4.4. Among these dimensions, 'cultural information display' has the highest score (4.7), while 'response speed' is slightly lower (4.4), and the overall satisfaction score is 4.6. The low standard deviation (0.48–0.61) indicates good consistency in the evaluation scores, which verifies the practical application effectiveness of the system in terms of interface friendliness (4.6), pattern recognition accuracy (4.5), and cultural communication value.

5 Conclusions

This study aims to address the technical challenges in the automated recognition of Miao costume patterns and the teaching of cultural inheritance. By integrating deep learning with the needs of ethnic cultural protection, an efficient intelligent recognition teaching system is constructed. With the YOLOv5 model as its core, the system incorporates multiple technologies, including the transformer attention mechanism, weighted k-means clustering for anchor box optimisation, SPPF MSF, and the combination of focal loss and CIoU loss function. These technologies enhance the model's ability to extract features and accurately detect complex patterns. Meanwhile, an interactive teaching platform with real-time detection, image recognition, and cultural knowledge display features is created using web technology. According to experimental results, the optimised model

outperforms mainstream detection algorithms overall, achieving mAP@0.5 of 89.6% and mAP@0.5:0.95 of 61.5% on the Miao costume pattern dataset. The system also receives high user satisfaction evaluations in practical applications. However, this study remains limited. For example, the precision with which edge segmentation is performed on small accessories has room for improvement, and the model's ability to withstand harsh lighting conditions requires additional enhancement. Future research will concentrate on multi-modal data fusion, lightweight deployment on mobile terminals, and optimising small target detection algorithms. This will increase the system's usability and appeal even more while offering more dependable technical assistance for the preservation and teaching of ethnic cultures online.

Acknowledgements

This work is supported by the Anhui Provincial Philosophy and Social Sciences Planning Project: Research on the Media Communication and Activation of Moral Images in Huizhou Carving Based on AIGC (AHSKQ2024D139) and the Anhui University of Information Science and Technology Intangible Cultural Heritage Leadership Team Project: Digital Intangible Cultural Heritage Team (25kytdlj001).

Declarations

No conflict of interest exists in the submission of this manuscript.

References

Alsuwaylimi, A.A., Alanazi, R., Alanazi, S.M., Alenezi, S.M., Saidani, T. and Ghodhbani, R. (2024) 'Improved and efficient object detection algorithm based on yolov5', *Engineering, Technology & Applied Science Research*, Vol. 14, No. 3, pp.14380–14386.

Cao, H.Y. and Tuo, J.Y. (2024) 'A method for identifying women's sleeves based on improved YOLOv5 and ResNet50', *Advanced Textile Technology*, Vol. 32, No. 1, p.45.

Chang, Y.H. and Zhang, Y.Y. (2022) 'Deep learning for clothing style recognition using YOLOv5', *Micromachines*, Vol. 13, No. 10, p.1678.

Dewi, C., Chen, R.C., Zhuang, Y.C. and Christanto, H.J. (2022) 'Yolov5 series algorithm for road marking sign identification', *Big Data and Cognitive Computing*, Vol. 6, No. 4, p.149.

Jung, H.K. and Choi, G.S. (2022) 'Improved yolov5: efficient object detection using drone images under various conditions', *Applied Sciences*, Vol. 12, No. 14, p.7255.

Kim, K., Kim, K. and Jeong, S. (2023) 'Application of YOLO v5 and v8 for recognition of safety risk factors at construction sites', *Sustainability*, Vol. 15, No. 20, p.15179.

Li, F., Xiao, K., Hu, Z. and Zhang, G. (2024) 'Fabric defect detection algorithm based on improved YOLOv5', *The Visual Computer*, Vol. 40, No. 4, pp.2309–2324.

Liu, B., Wang, H., Wang, Y., Zhou, C. and Cai, L. (2023) 'Lane line type recognition based on improved YOLOv5', *Applied Sciences*, Vol. 13, No. 18, p.10537.

Mao, M., Ma, J., Lee, A. and Hong, M. (2025) 'Enhancing fabric detection and classification using YOLOv5 models', *Engineering Proceedings*, Vol. 89, No. 1, p.33.

Mathew, M.P. and Mahesh, T.Y. (2022) 'Leaf-based disease detection in bell pepper plant using YOLO v5', *Signal, Image and Video Processing*, Vol. 16, No. 3, pp.841–847.

Moustapha, M., Tasyurek, M. and Ozturk, C. (2023) 'A novel yolov5 deep learning model for handwriting detection and recognition', *International Journal on Artificial Intelligence Tools*, Vol. 32, No. 4, p.2350016.

Moya, V., Quito, A., Pilco, A., Vásconez, J.P. and Vargas, C. (2024) 'Crop detection and maturity classification using a yolov5-based image analysis', *Emerging Science Journal*, Vol. 8, No. 2, pp.496–512.

Nguyen, N.T., Tran, Q., Dao, C.H., Nguyen, D.A. and Tran, D.H. (2024) 'Automatic detection of personal protective equipment in construction sites using metaheuristic optimized YOLOv5', *Arabian Journal for Science and Engineering*, Vol. 49, No. 10, pp.13519–13537.

Olorunshola, O.E., Irhebhude, M.E. and Evwiekpaefe, A.E. (2023) 'A comparative study of YOLOv5 and YOLOv7 object detection algorithms', *Journal of Computing and Social Informatics*, Vol. 2, No. 1, pp.1–12.

Rocha, D., Pinto, L., Machado, J., Soares, F. and Carvalho, V. (2023) 'Using object detection technology to identify defects in clothing for blind people', *Sensors*, Vol. 23, No. 9, p.4381.

Sadiq, M., Masood, S. and Pal, O. (2022) 'FD-YOLOv5: a fuzzy image enhancement based robust object detection model for safety helmet detection', *International Journal of Fuzzy Systems*, Vol. 24, No. 5, pp.2600–2616.

Sun, K., Zhang, Y.J., Tong, S.Y., Tang, M.D. and Wang, C.B. (2022) 'Study on rice grain mildewed region recognition based on microscopic computer vision and YOLO-v5 model', *Foods*, Vol. 11, No. 24, p.4031.

Tang, X., Wang, C., Su, J. and Taylor, C. (2023) 'An elevator button recognition method combining yolov5 and OCR', *Computers, Materials and Continua*, Vol. 75, No. 1, pp.117–131.

Wang, N., Qian, T., Yang, J., Li, L., Zhang, Y. and Zheng, X. (2022) 'An enhanced YOLOv5 model for greenhouse cucumber fruit recognition based on color space features', *Agriculture*, Vol. 12, No. 10, p.1556.

Xu, H., Zheng, W., Liu, F., Li, P. and Wang R. (2023) 'Unmanned aerial vehicle perspective small target recognition algorithm based on improved yolov5', *Remote Sensing*, Vol. 15, No. 14, p.3583.

Yusro, M.M., Ali, R. and Hitam, M.S. (2023) 'Comparison of faster R-CNN and yolov5 for overlapping objects recognition', *Baghdad Science Journal*, Vol. 20, No. 3, p.15.

Zhao, Q., Wei, H. and Zhai, X. (2023) 'Improving tire specification character recognition in the yolov5 network', *Applied Sciences*, Vol. 13, No. 12, p.7310.

Zhou, H., Ou, J., Meng, P., Tong, J., Ye, H. and Li, Z. (2023) 'Research on kiwi fruit flower recognition for efficient pollination based on an improved YOLOv5 algorithm', *Horticulturae*, Vol. 9, No. 3, p.400.

Zhou, S., Zhao, J., Shi, Y.S., Wang, Y.F. and Mei, S.Q. (2023) 'Research on improving YOLOv5s algorithm for fabric defect detection', *International Journal of Clothing Science and Technology*, Vol. 35, No. 1, pp.88–106.