IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 10 May 2024, accepted 10 June 2024, date of publication 13 June 2024, date of current version 21 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3414114

RESEARCH ARTICLE

Automatic Counting Method for Centipedes Based on Deep Learning

JIN YAO[1], WEITAO CHEN[2], TAO WANG[2], FU YANG[2], XIAOYAN SUN[1], CHONG YAO[1], AND LIANGQUAN JIA[2]
[1]Huzhou Central Hospital, Affiliated Central Hospital of Huzhou University, Huzhou, Zhejiang 313000, China
[2]School of Information Engineering, Huzhou University, Huzhou 313000, China

Corresponding authors: Chong Yao (ycu@hzhospital.com) and Liangquan Jia (02426@zjhu.edu.cn)

This work was supported in part by the Major Project for the Modernization of Traditional Chinese Medicine Jointly Established by the State Administration of Traditional Chinese Medicine and Zhejiang Province under Grant GZY-ZJ-KJ-24050, in part by the Research on Quality Grade Evaluation and Application of Traditional Chinese Medicine Pieces Based on ''Optimal Shape-Quality-Efficacy'' Correlation, and in part by Huzhou City Key Research and Development Project under Grant 2021ZD2003.

ABSTRACT The utilization of target detection algorithms for counting edible centipedes represents a novel endeavor in the field of traditional Chinese medicine materials using deep learning. However, the accuracy of current target detection algorithms is relatively low due to the complexity of the centipede background and the density of the detection targets, making them poorly suited for practical application scenarios. To address this, this study proposes a centipede target detection algorithm based on an improved Your Only Look Once V5 (YOLOv5) model, termed FD-YOLO. This algorithm enhances the original model by incorporating the CBAM attention mechanism and the BiFormer universal visual transformer to suppress irrelevant information and intensify focus on the desired detection targets, thereby improving the precision of the algorithm. Additionally, the FD-YOLO algorithm enhances the model's generalizability and robustness by improving the existing SPPF module. Experimental results demonstrate that compared to the original YOLOv5 prototype network, the improved YOLOv5 model has increased the AP@0.5 by 3.5%, reaching 97.2%, and raised the recall rate from 86.2% to 92.9%. Therefore, the enhanced YOLOv5 algorithm can effectively detect and count centipedes, making it more suitable for current practical application scenarios.

INDEX TERMS YOLOv5, computer vision, object detection, counting, centipede, medicinal materials.

I. INTRODUCTION

Centipedes have been used for medicinal purposes in China for thousands of years and can treat a variety of diseases [1]. They are particularly effective in treating rheumatism and arthritis. Furthermore, centipedes can promote blood circulation and enhance immunity, contributing positively to the body's resistance to illness [2]. In traditional Chinese medicine, centipedes also hold several other medicinal values, such as treating hemorrhoids and skin diseases. However, with the continuous increase in market demand, the scale of artificial centipede breeding has expanded, and the counting of commercial centipedes has become a pressing issue in the process of large-scale breeding. Traditional centipede counting involves manual sorting, bundling several centipedes together, counting by hand, and then packing for preservation. However, counting manually is inefficient, laborious, and too slow to satisfy the requirements of modern large-scale breeding scenarios. Therefore, improving detection efficiency and accuracy, thus enabling the automation of centipede counting, holds significant value and meaning for the development of the industry.

Object detection technology, as a major research direction in the fields of computer vision and digital image processing in recent years [3], has found broad applications across various industries, including traditional Chinese medicinal materials. Leveraging advanced computer vision, this technology significantly reduces the reliance on human labor in differentiating and identifying Chinese medicinal materials, thereby enhancing production efficiency and reducing costs,

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Olague.

which holds immense practical and economic significance. Given its substantial potential in practical applications, object detection has become a focal point of research in recent years, not only within the academic community but also in the pharmaceutical industry and other sectors. Concurrently, related algorithms continue to evolve rapidly.

Traditional object detection algorithms primarily rely on color differences within images to distinguish between objects and their backgrounds. Hussein et al. [4] conducted a study using three machine learning techniques: Linear Discriminant Analysis (LDA), Random Forest (RF), and Support Vector Machine (SVM). They explored the effectiveness of Synthetic Minority Over-sampling Technique (SMOTE) in enhancing classification performance for imbalanced datasets to address the issue of potential loss of feature details in automatic plant classification using machine learning techniques. Their results indicated varying performance of different techniques and data balancing methods across different plant families. Chen et al. [5] used network pharmacology and various machine learning techniques (including Support Vector Machine, Multivariate Linear Regression, Deep Learning, and Random Forest) to screen herbal candidates for Alzheimer's disease from the world's largest traditional Chinese medicine database. Their findings demonstrated that the models used achieved high R^2 values on both training and testing sets, indicating high predictive accuracy and suggesting potential in Alzheimer's disease treatment. Xu et al. [6] constructed a standardized traditional Chinese medicine dataset and developed a new machine learning model called Attention Pyramid Network (APN). This model leverages competitive attention and spatial collaborative attention mechanisms to adaptively process and model multi-scale features of traditional Chinese medicine images. Additionally, a new herbal identification framework was proposed. Experiments on the newly created dataset validated the effectiveness of this approach.

However, the accuracy and reliability of traditional object detection algorithms are often significantly limited in complex environments. For instance, in the task of detecting densely packed centipede images, traditional algorithms struggle to accurately differentiate individual centipedes due to occlusion between multiple centipedes and interference from complex background environments. The methods that segment based on color and shape perform poorly in practical centipede recognition scenarios.

With the advent of convolutional neural networks and the creation and evolution of residual structures, object detection algorithms have undergone a qualitative leap in a short period, offering more precise and robust detection methods. The new generation of object detection algorithms can better understand image content, achieving high precision in object recognition and localization even under conditions of changing lighting, occlusion, and background complexity. This progress has not only propelled the development of object detection technology but also opened new avenues for research and application in related fields.

Two-stage detection algorithms divide the detection process into two phases: initially generating region proposals and then classifying and refining these through convolutional neural networks. Since their inception, such algorithms have been extensively applied in the field of traditional Chinese medicine, with Fast R-CNN [7], Faster R-CNN [8], and Mask-RCNN [9] being among the most representative. Gang et al. [10] proposed a new lightweight convolutional neural network (CNN), named CCNet, designed to meet the computational constraints of edge mobile devices. The model compresses modern CNN architectures through a "bottleneck" structure and utilizes a downsampling rate of 3 to obtain multi-scale features. It also employs stacked GCIR modules and MDCA attention mechanisms to enhance model performance. Additionally, the model was validated on a newly created fine-grained traditional Chinese medicine (TCM) identification dataset called "TCM-100," achieving high classification accuracy. The experimental results showed that the model is not only suitable for edge devices, but its design philosophy and architecture also provide valuable references for classification or identification tasks in other fields. Zheng et al [11] introduced the Dual-Teacher Supervision Decay (DTSD) method, which refines soft output labels and adjusts decay parameters while using a dynamic combination loss strategy for knowledge distillation in teacher models. Combined with an exponential warm-up learning rate decay strategy, the DTSD method achieved a high accuracy of 98.60% across 10 herbal image classification tasks, while maintaining extremely low detection time and a small model size. The experimental results demonstrate that this improved model not only provides an effective method for herbal image recognition, but also offers a practical solution for lightweight models in mobile applications. Li et al. [12] presented a new fractional-order convolutional neural network (CFO-CNN) approach for traditional Chinese medicine (TCM) image recognition to improve classification accuracy and efficiency. The method uses the Caputo fractional-order gradient descent algorithm to update model parameters, optimizing gradient descent during training and avoiding the issue of traditional integer-order gradient descent falling into local optima. Furthermore, a fractional-order backpropagation function helps the model more effectively find the global optimum and realize deep feature recognition. Experimental results indicate that this approach can effectively perform deep classification and recognition of TCM, encompassing variety, type, and grade.

One-stage detection algorithms, such as YOLO [13], RetinaNet [14], and SSD [15], represent an end-to-end detection method that bypasses the candidate region generation phase, directly producing the class probabilities and location coordinates of objects. As they allow for direct final detection results through a single evaluation, these algorithms feature significantly faster detection speeds. Su et al. [16] utilized the YOLOv8 model to train on a Chinese Herbal Slice (CHS) dataset, collected and annotated from the internet, achieving a mean precision of 95.8% through Mosaic data augmentation. Compared to traditional image classification methods, this approach not only improved the identification rate of traditional Chinese herbal slices but also helped promote the awareness and acceptance of traditional Chinese medicine culture. Shiqi [17] tackled the challenges of small differences between current samples and imbalanced types by using the K-means++ clustering algorithm to recluster anchor boxes and optimizing YOLO-V4's loss function based on the Focal Loss function. Experimental results demonstrated that these optimizations effectively improved the detection and identification of small packages of traditional Chinese medicine samples, allowing the improved YOLO-V4 algorithm to outperform the original version. Tian et al [18] proposed an intelligent identification method based on YOLO v5 for honeysuckle pollen grain microscopic images by combining deep learning with microscopic image analysis of traditional Chinese medicine. Furthermore, their research validated the model's scalability in microscopic feature identification under different magnifications. This technology not only enhances the quality of traditional Chinese medicine but also aids in achieving quality standardization, demonstrating promising applications.

However, most of the models for centipede counting and detection available in the market primarily adopt traditional, unoptimized original architectures. This lack of innovation and customized improvement leads to various performance deficiencies, especially when these models are applied to specific centipede datasets. Specifically, these standard models often exhibit lower recognition accuracy, which, in practical applications, could result in inaccurate estimations of centipede numbers, thereby affecting the outcomes of actual production and biological research.

In light of the issues above, it is necessary to carry out in-depth improvements and optimizations on existing centipede counting and detection models. This study proposes a brand new centipede counting detection model, FD-YOLO, which is based on a series of optimizations and improvements made on YOLOv5. These modifications ensure that the model achieves high detection accuracy for centipedes. The main contributions of this paper are as follows:

(1) To address the issues of dense centipede populations and significant environmental disturbances in centipede images, this study has incorporated the BiFormer universal visual transformer into the original model. This enhancement strengthens the feature extraction capability for obscured objects and suppresses the interference from complex environments, thereby improving the model's detection ability.

(2) To enhance the model's generalizability and robustness, this study has employed the SPPFCSP module to improve the existing SPPF module, thereby increasing the model's flexibility and applicability. This enhancement boosts the model's feature extraction and representation capabilities.

(3) In order to further improve the model's recognition accuracy, the CBAM attention module has been integrated into the model. The CBAM attention module captures target spatial and channel information more effectively without increasing the number of model parameters, suppresses irrelevant information, and extracts valuable information, thereby enhancing the model's detection precision.

(4) To ensure that the model achieves optimal performance in centipede recognition, our team has created a new centipede target detection dataset. The dataset was captured in multiple sessions according to real-world application scenarios to ensure its extensive applicability.

II. MATERIALS AND METHODS

A. EXPERIMENTAL DATA

In order to thoroughly simulate the performance of the centipede detection model in actual usage scenarios, this research initiated a close collaboration with Huzhou Central Hospital. On September 24, 2023 (overcast) and September 25, 2023 (sunny), our team conducted systematic centipede data collection in the professional laboratory of Huzhou Central Hospital located in Wuxing District, Huzhou City, Zhejiang Province, China. This data collection aimed to mimic the real-world occurrences of centipedes under different environmental conditions, covering outdoor scenes during overcast and sunny conditions as well as indoor situations with direct and backlighting, to ensure a wide and realistic coverage of the dataset scenarios.

Specifically, our team captured photos of centipedes under the aforementioned four different lighting and environmental conditions, taking 80 pictures for each setting, thus obtaining a total of 320 high-quality images. The distance of the photos was maintained between approximately 500 to 1500 millimeters to ensure the clarity and detail of the centipedes. Notably, considering the significant size differences among individual centipedes, where larger ones could obscure smaller individuals, this phenomenon could negatively impact the model's detection accuracy in real scenarios. To overcome this challenge and build a more robust centipede counting model, our team employed a multi-angle photography approach to increase the diversity and dimensions of the data. This method not only helps reduce the issue of inter-individual occlusion but also improves the model's recognition capabilities for centipedes of different angles and sizes, thereby enhancing the model's generalizability and robustness. Ultimately, a diverse and high-quality dataset of centipede images was successfully collected, providing a solid foundation for subsequent model training and validation. The overall results of the photography are illustrated in Figure 1.

B. IMAGE ANNOTATION AND DATASET CONSTRUCTION

To ensure the accuracy and broad applicability of the dataset annotation, this study utilized Labelimg for data marking [19]. We saved the annotation labels in txt format to facilitate easier subsequent processing. To minimize the impact of dataset annotation errors on the final experimental results, the research team carefully screened the images before labeling, excluding those blurred or indistinguishable due to accidental factors. Simultaneously, a professional dataset annotation team was meticulously selected to ensure their extensive experience and precise labeling skills. To guarantee consistency in annotation, the team established strict dataset labeling standards to ensure each bounding box accurately reflects the targets in the images. Through diligent efforts, a total of 316 images were successfully selected, resulting in 31,600 annotation boxes.

In the final stage of data preparation, to further increase the randomness of the training, validation, and test sets, the entire dataset was randomly shuffled, disrupting the original sequence of images. Ultimately, the annotated images were allocated into the training, validation, and test sets in an 8:1:1 ratio to ensure better generalization performance during model training and evaluation. Through this series of stringent data processing steps, a high-quality and diverse data foundation was provided for the final experiments, laying a solid foundation for the reliability and scientific integrity of the research.

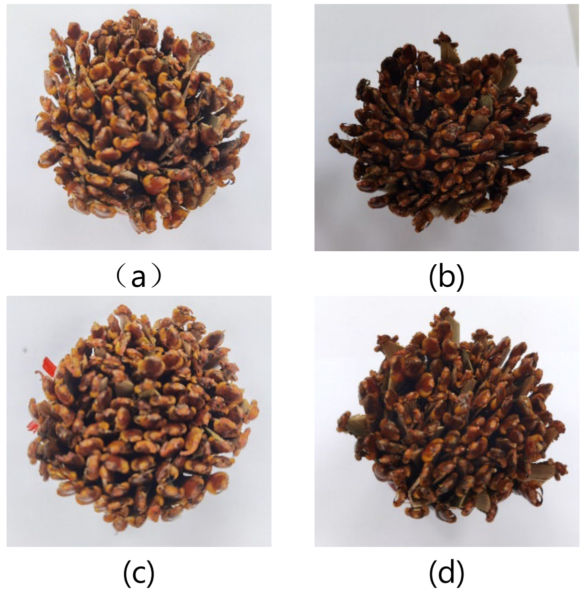

FIGURE 1. Centipede Dataset. (a) Centipede image under sunny daylight (b) Centipede image under overcast daylight (c) Centipede image under direct artificial lighting (c) Centipede image under backlight artificial lighting.

C. DESIGN FOR FD-YOLO

"You Look Only Once" (YOLO) is an iconic algorithm in the field of single-stage object detection [13], simplifying the detection task into a regression problem and directly predicting bounding boxes and categories through a single neural network [20]. The performance of the network varies with its depth and width. YOLOv5 is divided into five different scales of models: s, m, l, x, n, with the computational resource requirements increasing as the model size increases. To date, many scholars have evolved it into multiple versions adapted to different applications, each demonstrating excellent performance in their specific domains. However, different datasets exhibit various characteristics within these diverse models. Among them, YOLOv5 is widely recognized for its outstanding generalization capabilities and efficient performance. Despite nearly four years since its release, YOLOv5 continues to demonstrate strong competitiveness. Given this, our research has chosen YOLOv5 as the basis for improvement.

Considering the practical application needs and laboratory equipment conditions, our team has selected YOLOv5s as the primary framework for our study, which consists of four main parts: image input, backbone network (for feature extraction), neck network (for feature fusion), and the output layer (for object detection).

Due to the low recognition accuracy and inaccuracies in the original model, which hinder precise classification of traditional Chinese medicinal materials in practical applications, this study proposes a new and efficient centipede counting model, FD-YOLO, as illustrated in Figure 2. This method combines YOLOv5 with the BiFormer universal vision transformer, taking full advantage of both algorithms to maintain high precision while achieving rapid recognition and counting. Furthermore, based on the original improvements, this study has added the CBAM attention mechanism in the backbone part, which can suppress ineffective information and extract useful information, thereby enhancing the detection precision. Finally, this study has improved the existing SPPF module, thereby increasing the model's flexibility and applicability and enhancing the model's feature extraction and representation capabilities.

1) 2.3.1 CONVLUTIONAL BLOCK ATTENTION(CBAM)

When detecting and counting the number of centipedes, the dense population of centipedes and the complex background environment can introduce certain disturbances, affecting the final accuracy of centipede number recognition. To address this challenge, incorporating attention mechanisms into the algorithm has become a conventional and effective solution. It enhances the model's ability to selectively focus on relevant features while filtering out irrelevant information. The most commonly used attention mechanisms include SE [21], CA [22], ECA [23], and CBAM [24], among others. The Convolutional Block Attention Module (CBAM), one of the most representative attention mechanisms currently, combines the Channel Attention Module (CAM) and the Spatial Attention Module (SAM), forming a simple and effective end-to-end universal module. According to our team's experiments, this algorithm can effectively resolve the disturbances caused by complex environments in centipede images, thereby improving the detection accuracy of the algorithm. The structure of CBAM is illustrated in Figure 3.

When designing the channel attention module, the input feature maps are first processed through global maximum pooling and global average pooling. The aim is to extract two sets of one-dimensional feature vectors from the original feature mappings, each possessing different statistical

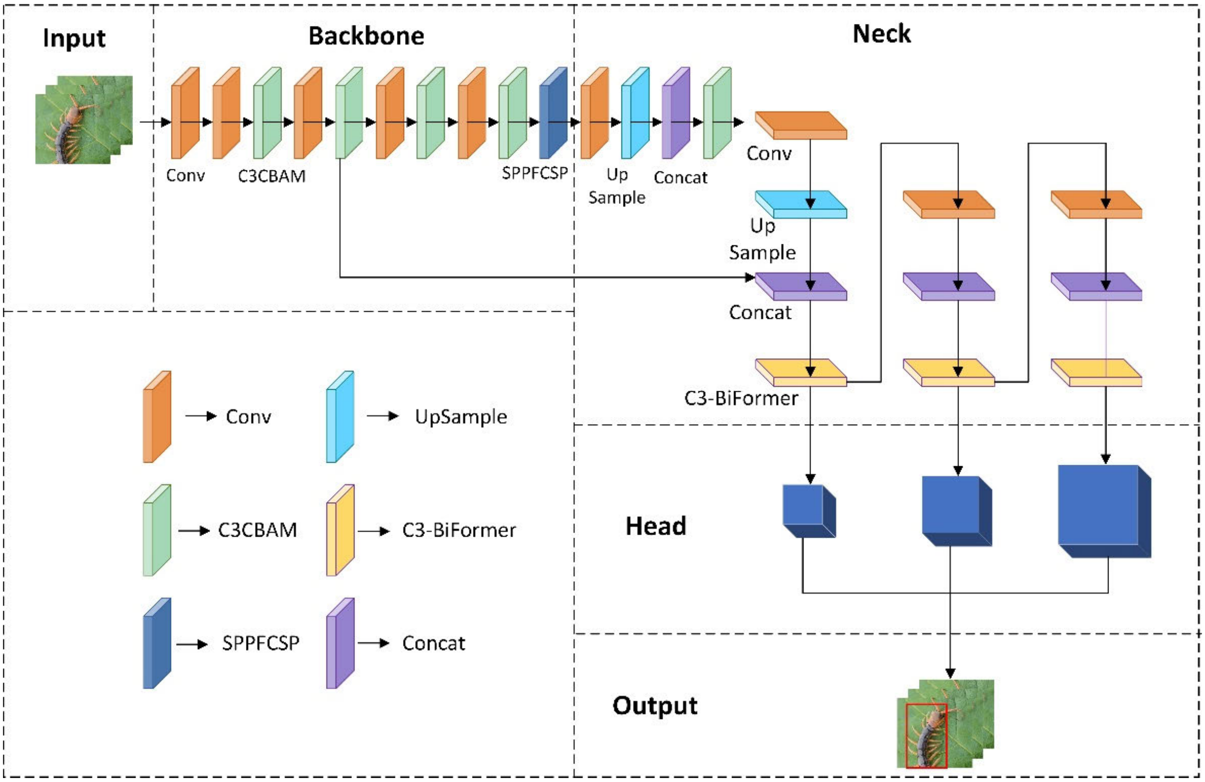

FIGURE 2. Schematic Diagram of the FD-YOLO Model.

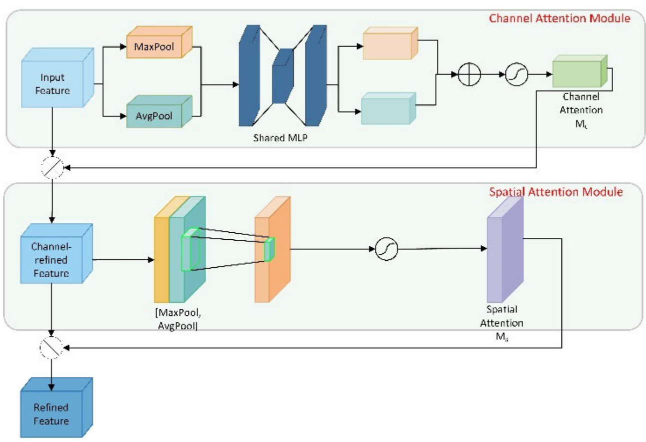

FIGURE 3. Schematic Diagram of the CBAM Attention Mechanism.

properties. This step is intended to capture global contextual information, thereby providing a richer input for subsequent feature analysis. Following this, these two sets of feature vectors are fed into the same multi-layer perceptron (MLP) [25] simultaneously. The purpose of this multi-layer perceptron is to delve deeper into and transform the information within the feature vectors, refining more detailed and useful feature representations through nonlinear activation functions. After processing by the multi-layer perceptron, the resulting output feature vectors are combined through an element-wise addition operation. Finally, a Sigmoid activation function is applied to this combined vector to generate the final channel attention weights. These weights represent the importance of each channel, guiding the model to focus more on those feature channels that are more informative. The corresponding formulas are illustrated below.

$$\begin{aligned} M_c(F) &= \sigma\left(MLP\left(AvgPool\left(F\right)\right) + MLP\left(MaxPool\left(F\right)\right)\right) \\ &= \sigma\left(W_1\left(W_0\left(F^c_{avg}\right)\right) + W_1\left(W_0\left(F^c_{max}\right)\right)\right) \end{aligned} \quad (1)$$

where c represents the channel attention module, σ represents the sigmoid function. $AvgPool$ and $MaxPool$ represent average pooling and maximum pooling, respectively. F^c_{avg} and F^c_{max} represent the average pooling features and maximum pooling features, respectively. W_0 and W_1 are the weights of the MLP.

To achieve more detailed and precise attention feature capture in the spatial dimension, this study conducted a series of meticulously designed operations on the feature maps optimized by the preliminary channel attention module. Specifically, a key step was first performed on these optimized feature maps: global average pooling and global maximum pooling based on their width and height. This process generates two feature maps with different statistical properties, capturing the global average and extreme information of the features, respectively. Next, to construct a more comprehensive and enriched feature representation, these two pooled feature maps were concatenated. This concatenation strategy aims to integrate different feature dimensions, thereby providing a more comprehensive feature base for subsequent analysis. Then, the combined feature map is further processed through a specifically configured convolutional layer for deeper feature extraction. This step refines and enhances the model's ability to understand and capture spatial information [26]. Finally, by applying the Sigmoid activation function to the features processed by the convolutional layer, the final spatial attention weights are successfully generated. These weights significantly guide the model to focus its attention on the most important spatial regions within the image. This refined spatial attention mechanism significantly improves the model's ability to focus during the feature extraction and recognition process, ensuring that the model can more effectively identify and understand the key information and regions in the image. The computation process is as follows:

$$\begin{aligned} M_s(F) &= \sigma\left(f^{7\times 7}\left([AvgPool(F); MaxPool(F)]\right)\right) \\ &= \sigma\left(f^{7\times 7}\left(\left[F^s_{avg}; F^s_{max}\right]\right)\right) \end{aligned} \tag{2}$$

where s represents the spatial attention module. *AvgPool* and *MaxPool* represent average pooling and maximum pooling, respectively. $f^{7\times 7}$ denotes a convolution operation with a filter size of 7×7.

2) BIFORMER

Our team's research found that factors affecting the model's detection accuracy are not only due to disturbances caused by complex environments but may also stem from issues related to the centipedes themselves, leading to a decrease in model accuracy. The dense nature of centipede images can cause the model to struggle in effectively distinguishing individual centipedes, potentially resulting in multiple centipedes being classified as a single entity and ultimately leading to counting discrepancies. To address this, our study suggests adding a BiFormer structure to the existing model.

The visual Transformer model BiFormer represents an innovative universal visual transformer design [27]. As shown in Figure 4, the BiFormer structure adopts a four-layer pyramid shape, reflecting the latest ideas in visual transformer design. This model particularly introduces a unique dual-pathway attention mechanism, enabling the BiFormer to dynamically focus on a few key tokens, thereby not only enhancing model performance but also significantly improving computational efficiency. This aspect is particularly prominent in scenarios requiring dense predictions, such as image classification and object detection, where BiFormer demonstrates superior performance in these applications.

In the BiFormer architecture, a multi-level pyramid layer strategy is employed to extract features at different hierarchical levels. This dual-pathway strategy allows the high-level global features to complement and integrate with the low-level local features, thereby promoting a smoother flow of information and a richer expression of features. Specifically, the model uses overlapping patch embedding

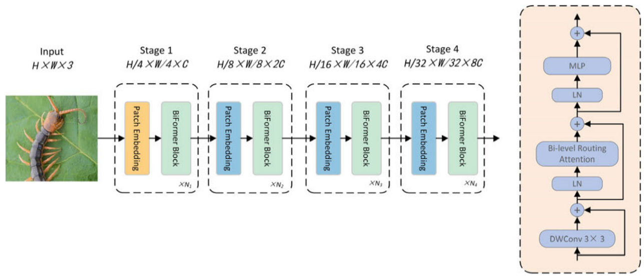

FIGURE 4. Left: The overall architecture of BiFormer Right: Detailed information of the BiFormer Block.

techniques in the initial phase to enhance its feature-capturing ability, and in the subsequent stages, it employs patch merging techniques to optimize spatial resolution and increase the number of channels. This is followed by several BiFormer blocks for deep feature transformation, further enhancing the model's performance.

Numerous studies have validated that the BiFormer model demonstrates exceptional performance in recognizing centipedes, thanks to its dual-layer routing attention mechanism. This enables the model to simultaneously capture the subtle differences and global connections of the targets, effectively distinguishing between the target objects and the background.

3) SPPFCSP

In YOLOv5, the Spatial Pyramid Pooling (SPP) module enables the rapid processing of feature maps of different sizes [28]. The enhanced version, SPPF, increases the receptive field by processing feature maps at multiple scales in parallel, capturing information of objects at various scales. This method improves the model's ability to detect targets of various sizes while maintaining a faster processing speed. However, our experiments have shown that this module's performance on the centipede dataset is less than satisfactory, as excessive pooling still poses the risk of overfitting. To enhance the algorithm's capability in recognizing centipede targets, this study has added the CSP module to the original model, proposing the SPPFCSP module.

SPPFCSP, or Spatial Pyramid Pooling Fast Cross-Stage Partial network, represents an innovative improvement based on the SPPNet (Spatial Pyramid Pooling Network) [29]. It aims to efficiently process input images of various sizes, enhancing the model's semantic capture capability for small-scale features and ensuring the effective integration of local and global features. SPPFCSP integrates multi-scale Maxpool layers, significantly reducing feature redundancy in the convolutional neural network, lowering computational load, speeding up processing, and enhancing detection accuracy. Furthermore, this structure not only combines spatial pyramid pooling and fully connected layers to enhance feature representation but also introduces the CSP (Cross-Stage Partial) design, effectively reducing the model's parameters and computational complexity [30]. SPPFCSP is committed to improving the efficiency of feature extraction and

TABLE 1. Baseline model comparison.

Algorithms	Parameters	mAP	FPS
Faster RCNN	32	89.7	32
VGG16	138	84.3	18
Resnet-50	25	88.1	24
Resnet-101	44	89.9	18
Alexnet	60	91.2	43
SSD	13.8	83.9	59
Mobilenetv3-small	1.53	86.7	51
Mobilenetv3-large	4.22	90.2	26
Yolov5-s	7.01	93.7	89

information fusion, aiming to maintain or even enhance model performance while optimizing resource consumption.

As shown in the architecture in Figure 5, the first branch begins with the CBS (Cross-Stage Bottleneck Suppression) module, which is responsible for preliminary feature extraction. Additionally, this branch integrates three Maxpool modules of different sizes, specifically for 5 × 5, 9 × 9, and 13 × 13 feature sizes, aimed at capturing feature information of various scales. These three scales of feature maps are then integrated together and connected with the end feature map of size 1 × 1. In the other branch of the framework, we constructed a network based on the CSPNet architecture, characterized by the CSP module splitting the feature layer into two parts: one inherits the most recent pooling path; the other is realized by combining the CBS with the Maxpool layer after feature extraction. This design effectively cuts the computation in half, thereby speeding up the model's operation and improving the accuracy of detection.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. MODEL SELECTION

To ensure the practicality of the subsequent models in real-world applications, this study aimed to select a target detection model that combines efficient image processing capabilities, high-precision detection abilities, and minimal hardware requirements. To this end, comparative experiments were conducted on the nine most commonly used fundamental models available on the market. These models were evaluated based on the number of parameters, the mean average precision, and operational speed. The final comparative results are presented in Table 1. It is noteworthy that all experiments were performed using an RTX 3060 device.

As indicated by the results in Table 1, in terms of model size, the parameter count of YOLOv5-s is second only to MobileNetV3-Small and MobileNetV3-Large. In the aspect of mean average precision, YOLOv5-s outperforms all other models. Furthermore, in real-time detection, YOLOv5-s significantly surpasses the other models, achieving 89 FPS. Therefore, after comprehensive consideration of the model's parameter count, mean average precision, and operational speed, this study has ultimately selected YOLOv5-s as the foundational model for this research.

TABLE 2. Main configuration of experimental equipment.

Items	values
Operating system	Window 11
CPU	I7-11700
GPU	RTX3060
Memory	12GB
Deep learning framework	Pytorch 1.11
Cuda version	11.7

B. EXPERIMENTAL ENVIRONMENT PARAMETERS AND TRAINING PARAMETERS

Experimental Environment: All experiments in this study are based on the Pytorch deep learning framework, with Python as the programming language. The main configurations of the engineering machine used in this experiment are shown in Table 2.

Training Parameter Settings: Before model training, hyperparameters are set in advance, as appropriate hyperparameters can enhance model performance. The YOLOv5 algorithm includes 28 hyperparameters, comprising settings for the learning rate, weights for different loss functions, weight decay coefficients, and various data augmentation parameters. More significant hyperparameters are shown in Table 3. In the model configuration, 'lr0' and 'lrf' represent the initial value of the learning rate and its final ratio, respectively; 'momentum' denotes the inertia value used in stochastic gradient descent; box controls the weight of the bounding box loss; 'cls' and 'cls_pw' respectively signify the weight of the classification loss and the weight of the positive samples in the binary cross-entropy loss for classification; 'obj' and 'obj_pw' represent the weight of the confidence loss and the weight of the positive samples in the binary cross-entropy loss for confidence, respectively; 'iou_t' refers to the IoU threshold used for matching predicted boxes with true boxes; 'anchor_t' denotes the threshold for anchor selection; 'hsv_h', 'hsv_s', and 'hsv_v' are parameters for enhancing hue, saturation, and brightness in the HSV color space of images; 'translate', 'scale', 'fliplr', and 'mosaic' are parameters used for data augmentation. It is noteworthy that the values of these hyperparameters are the default settings for the YOLOv5 model and are also considered by many scholars to be the most suitable hyperparameter values for this model.

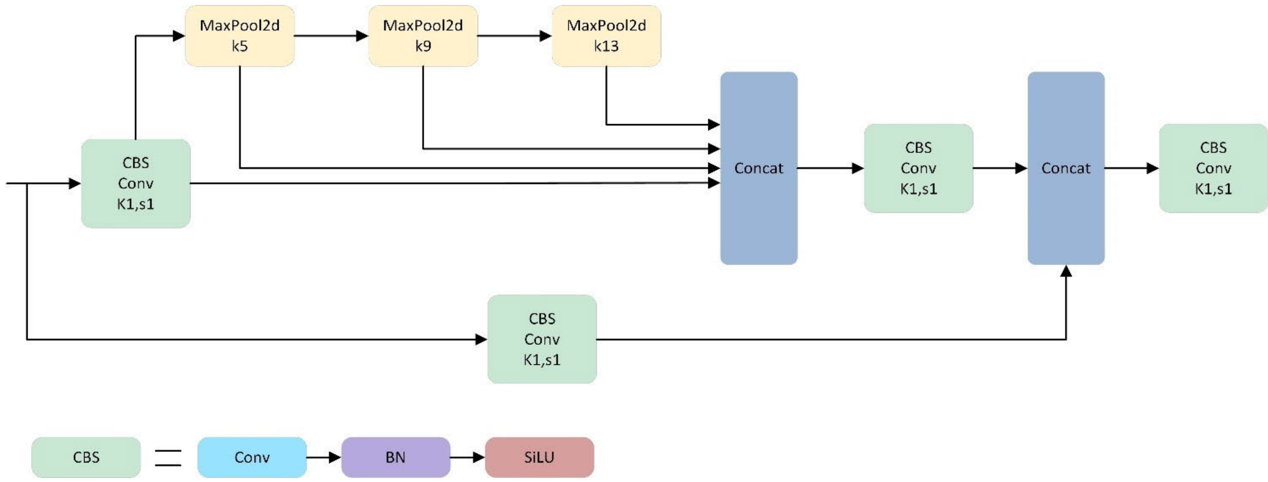

FIGURE 5. Structure Diagram of SPPFCSP.

TABLE 3. Values of various parameters for yolov5_s.

Hyperparameters	Value
lr0	0.01
lrf	0.01
momentum	0.937
box	005
cls	0.5
cls_pw	1.0
obj	1.0
obj_pw	1.0
iou_t	0.20
anchor_t	3
hsv_h	0.015
hsv_s	0.7
hsv_v	0.4
translate	0.1
scale	0.5
fliplr	0.5
mosaic	1.0

This experiment uses the improved FD-YOLO as the training model for this centipede dataset. Standardize the image size to 640 × 640 pixels, and divide the dataset into training, testing, and validation sets in a ratio of 8:1:1. A total of 254 images were obtained for the training set, 31 images for the testing set, and 31 images for the validation set. During training, we set the epoch, image size, and mosaic to the conventionally recognized optimal values for the traditional YOLO algorithm, which are 300, 640 × 640, and 1, respectively. Considering the performance of the experimental equipment, we set the batch size to 16. mAP0.5 refers to the mean Average Precision at an IOU threshold set to 0.5 for all classes, while mAP0.95 refers to the mean Average Precision for the IOU threshold increasing in steps of 0.05 from 0.5 to 0.95 and the weighted average.

C. MODEL EVALUATION METRICS

To comprehensively evaluate the efficacy of the centipede detection model, this study adopts three core evaluation metrics: Precision, Recall, and Average Precision (AP), aiming to accurately measure the model's detection performance on the centipede dataset. In the field of object detection, the AP value plays a crucial role as it reflects the average precision at various recall levels, that is, the area under the precision-recall curve. Therefore, a higher AP value indicates that the object detection model performs better on the specified dataset.

$$P = \frac{TP}{FP + TP} \tag{3}$$

$$R = \frac{TP}{FN + TP} \tag{4}$$

$$AP = \sum \int_0^1 P(R)dR \tag{5}$$

where P and R represent Precision and Recall, respectively. TP (True Positive) represents the number of centipedes correctly identified by the network model. FP (False Positive) represents the number of centipedes incorrectly identified by the network model. FN (False Negative) represents the actual number of centipedes that the network model failed to detect.

D. COMPARISON OF ABLATION STUDY

In this study, both the network structure model and the loss function of the YOLOv5 detection algorithm were improved. To verify whether all improvements in YOLO could enhance performance, an ablation study was conducted on the dataset constructed for this research. To ensure the correctness of the ablation experiments, the settings of the model hyperparameters and the running environment were kept consistent. The experimental results are shown in Table 4.

The results indicate that modifications in different parts of the model have a positive impact. From the table 4, we can see that compared to the original Yolov5-s model, replacing the SPPF module with the SPPFCSP module, although

TABLE 4. Comparison of ablation study.

Algorithms	BiFormer	CBAM	SPPFCSP	Parameters	FLOPs(G)	Recall	AP
Yolov5s				7012822	15.8	86.2%	93.7%
Proposed Methods(1)	√			8078080	57.2	91.7%	95.6%
Proposed Methods(2)		√		6782620	15.2	91.1%	95.3%
Proposed Methods(3)			√	13438166	21.1	90.3%	95.1%
Proposed Methods(4)	√	√		7838364	56.5	92.3%	96.3%
Proposed Methods(5)	√		√	14506230	62.4	93.2%	96.5%
Proposed Methods(6)		√	√	13170912	20.1	93.0%	96.9%
Proposed Methods	√	√	√	14226656	61.1	92.9%	**97.2%**

the computational load has increased, correspondingly, the improved model has seen increases in Recall and AP in centipede detection, improving to 90.3% and 95.1% respectively. The introduction of the BiFormer visual transformer in 2023 has further enhanced the model's performance in centipede number detection, with Recall and AP increasing from the original 90.3% and 95.1% to the current 93.0% and 96.9%. Moreover, adding the CBAM attention mechanism to the original model has increased Precision, Recall, and AP while slightly reducing the model's number of parameters, from the original 7.0M to now 6.7M. The experimental results demonstrate that the various improvements made to the Yolov5 in this study have played their expected roles compared to the original model and have a clear advantage in the detection and counting of the centipede dataset.

E. COMPARISON OF DIFFERENT MODELS

To further examine the practicality of the FD-YOLO algorithm, this study conducted a comparative analysis with currently popular single-stage object detection algorithms. This includes the latest algorithms such as YOLOv7 and YOLOv8, as well as improved versions of YOLOv5-s combined with MobileNet and ShuffleNet, among others. To ensure the accuracy and reliability of the experimental results, we utilized a unified dataset for training and testing these algorithms, including the same training set, validation set, and test set. Detailed experimental results are presented in Table 5.

The experimental results show that the improved model proposed in this paper, compared to other improved models such as YOLOv5-ShuffleNet2, YOLOv5-MobileNet3, and YOLOv5-MobileNet3-CARAFE, exhibits significant improvements in recall and precision rates. Specifically, the recall rates have increased by 9.5%, 5.9%, and 1.1%, respectively. In terms of precision, they have increased by 5.7%, 2.7%, and 1.5%, respectively. Compared to current mainstream base models like YOLOv7 and YOLOv8-s, FD-YOLO still maintains a significant advantage, with precision improvements of 0.9% and 0.4%, respectively. Figure 6 visually represents the performance of eight models on the centipede dataset, where FD-YOLO shows a clear advantage

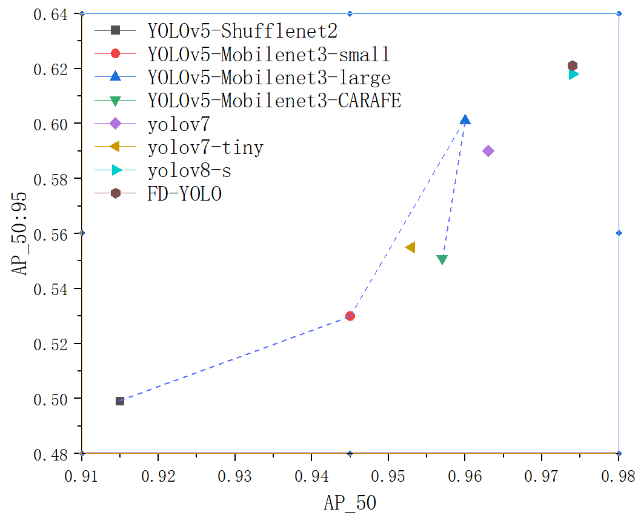

FIGURE 6. Comparison of Different Models.

in terms of AP_50 and AP_50:95. Therefore, the improved model FD-YOLO has a distinct advantage in centipede target detection, and its higher accuracy can also perform well in practical applications.

IV. DISCUSSION

In this study, aimed at meeting the demand for efficient and accurate counting of centipedes in real-life applications, we made a series of innovative improvements based on the YOLOv5 architecture. We introduced multiple efficient modules and proposed the FD-YOLO model, thereby achieving significant improvements in detection accuracy. To ensure the practical applicability of the model, we meticulously simulated real usage scenarios, captured, and organized images of centipedes under various environmental conditions, constructing a comprehensive and practical dataset for model training. In comparative experiments, FD-YOLO showed significant improvements in accuracy compared to the standard YOLOv5s model and even demonstrated advantages in detection accuracy and speed when compared with the latest YOLO series model, YOLOv8. The identification results of

TABLE 5. Comparison results of different models.

Algorithms	Image Size	Parameters	Precision	Recall	AP_50	AP_50:95
YOLOv5-Shufflenet2	640×640	3177626	0.898	0.834	91.5%	49.9%
YOLOv5-Mobilenet3-small	640×640	1386812	0.935	0.870	94.5%	53%
YOLOv5-Mobilenet3-large	640×640	5081352	0.941	0.924	96.0%	60.1%
YOLOv5-Mobilenet3-CARAFE	640×640	5075472	0.936	0.918	95.7%	55.1%
Yolov7	640×640	9320380	0.944	0.907	96.3%	59%
Yolov7-tiny	640×640	6014988	0.948	0.895	95.3%	55.5%
Yolov8-s	640×640	11125971	0.951	0.936	96.8%	61.8%
FD-YOLO	640×640	14226656	0.960	0.929	**97.2%**	**62.1%**

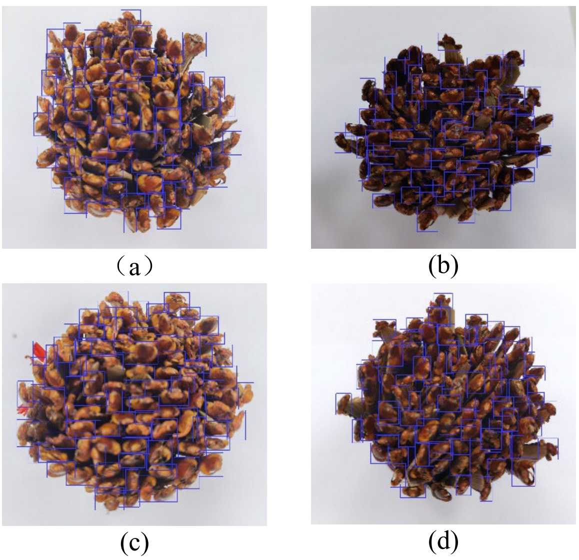

FIGURE 7. Examples of Detection Results Based on the FD-YOLO Model: (a) Detection results of images taken in a sunny environment. (b) Detection results of images taken in an overcast environment. (c) Detection results of images under direct artificial lighting. (d) Detection results of images under backlight artificial lighting.

our method on the test set are shown in Figure 7. These results strongly prove the effectiveness of our improved algorithm, addressing the issues most models face in accurately identifying and counting centipedes, making it effectively applicable to actual production and living environments.

This study aims to deploy deep learning models on mobile devices to facilitate the real-time detection and counting of centipedes. By employing advanced deep learning techniques, the system can identify and locate centipedes in various environments, significantly aiding management personnel in quickly and accurately obtaining information on centipede populations, thus substantially reducing their workload. To achieve this, the study utilized the NCNN (NCNN Convolutional Neural Network) framework for deploying the model on mobile platforms. NCNN is a high-performance neural network inference framework optimized for mobile devices, enabling the practical implementation of a centipede detection application. This application, after being integrated with the actual workflows of herbal medicine managers, has been proven to provide effective detection results and an excellent user experience.

During our in-depth research process, we encountered a series of challenges, which seem to be common issues for current object detection models. Particularly in performing centipede target detection tasks, the model needs to deal with a series of difficulties, including visual obstructions caused by camera shake, overexposure, and the high-density aggregation of centipedes. These factors combined may lead to missed or false detections by the target detection models. To overcome these issues, we recommend adopting more advanced data augmentation techniques, such as random occlusion and noise addition. These measures can help the model more accurately recognize subtle features, thereby significantly enhancing the model's generalization performance.

This study not only advances the application of object detection technology in the more niche field of medicine but also plays an important role in promoting the development and popularization of traditional Chinese medicinal materials. Furthermore, our work provides new ideas and methods for other similar studies. Looking forward, we will continue to explore further optimization and streamlining of the model, striving to achieve its application in a broader range of fields to meet more real-world demands and bring more innovation and value to traditional Chinese medicine materials and related areas.

V. CONCLUSION

In response to the development trends of the traditional Chinese medicine materials market and to meet the specific needs for centipede detection and counting in actual production scenarios, this study introduces a new strategy for centipede detection based on an improved YOLOv5 algorithm. This method not only inherits the efficiency of the original YOLOv5 algorithm but also significantly enhances the model's ability to detect centipedes and count them accurately through the integration of two innovative technologies: the CBAM attention mechanism and the BiFormer universal visual transformer. Additionally, to further strengthen the model's generalization and robustness against centipedes under various backgrounds, we have replaced the standard SPPF module in YOLOv5 with the more advanced SPPFCSP module. In practice, to verify the effectiveness and practicality of the proposed algorithm, our research team adopted a comprehensive experimental design, systematically capturing and accurately annotating centipedes under different lighting and environmental conditions. This approach ensured the diversity and authenticity of the model training and testing data. Experimental results show that, compared to the traditional YOLOv5 algorithm, our FD-YOLO algorithm has significantly improved in terms of accuracy and recall rates, reaching 97.2% and 92.9% respectively. By enhancing the YOLOv5 algorithm, the precision of this algorithm in detecting edible centipedes has been improved. This not only provides crucial technical support for the accurate counting of centipedes and advancements in the field of traditional Chinese medicine materials but also offers significant referential value for the future intelligent development of traditional Chinese medicine materials, considering the universality of the improvements made to the algorithm.

Given these results, we plan to further optimize the FD-YOLO algorithm in future research, especially regarding its performance under complex backgrounds and variable lighting conditions. Our goal is to enhance the adaptability and robustness of the algorithm, enabling it to achieve efficient and accurate centipede detection and counting in a wider range of application scenarios, thereby bringing more innovation and value to the field of traditional Chinese medicine.

ACKNOWLEDGMENT

(Jin Yao and Weitao Chen contributed equally to this work.)

REFERENCES

[1] E. P. Cherniack and E. Cherniack, "Bugs as drugs, part two: Worms, leeches, scorpions, snails, ticks, centipedes, and spiders," *Altern Med. Rev.*, vol. 16, no. 1, pp. 8–50, 2011.

[2] V. Haddad, "Tropical dermatology: Venomous arthropods and human skin: Part II. Diplopoda, Chilopoda, and Arachnida," *J. Amer. Acad. Dermatol.*, vol. 67, no. 3, p. 347, 2012.

[3] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proc. IEEE*, vol. 111, no. 3, pp. 257–276, Mar. 2023.

[4] B. R. Hussein, "Automated classification of tropical plant species data based on machine learning techniques and leaf trait measurements," in *Proc. Comput. Sci. Technol. 6th (ICCST)*, Kota Kinabalu, Malaysia. Cham, Switzerland: Springer, Aug. 2019, pp. 29–30.

[5] H.-Y. Chen, "Deep learning and random forest approach for finding the optimal traditional Chinese medicine formula for treatment of Alzheimer's disease," *J. Chem. Inf. Model.*, vol. 59, no. 4, pp. 1605–1623, 2019.

[6] Y. Xu, G. Wen, Y. Hu, M. Luo, D. Dai, Y. Zhuang, and W. Hall, "Multiple attentional pyramid networks for Chinese herbal recognition," *Pattern Recognit.*, vol. 110, Feb. 2021, Art. no. 107558.

[7] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[8] S. Ren, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 1137–1149.

[9] A. O. Vuola, S. U. Akram, and J. Kannala, "Mask-RCNN and U-Net ensembled for nuclei segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 208–212.

[10] H. Gang, S. Guanglei, W. Xiaofeng, and J. Jinlin, "CCNNet: A novel lightweight convolutional neural network and its application in traditional Chinese medicine recognition," *J. Big Data*, vol. 10, no. 1, p. 114, Jul. 2023.

[11] L. Zheng, W. Long, J. Yi, L. Liu, and K. Xu, "Enhanced knowledge distillation for advanced recognition of Chinese herbal medicine," *Sensors*, vol. 24, no. 5, p. 1559, Feb. 2024.

[12] T. Li, "Deep recognition of Chinese herbal medicines based on a caputo fractional order convolutional neural network," in *Proc. Int. Workshop Internet Things Big Data Healthcare*. Cham, Switzerland: Springer, 2023, pp. 41–51.

[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[14] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.

[15] W. Liu, "SSD: Single shot multibox detector," in *Proc. 14th Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands. Cham, Switzerland: Springer, Oct. 2016, pp. 21–37.

[16] Y. Su, B. Cheng, and Y. Cai, "Detection and recognition of traditional Chinese medicine slice based on YOLOv8," in *Proc. IEEE 6th Int. Conf. Electron. Inf. Commun. Technol. (ICEICT)*, Jul. 2023, pp. 214–217.

[17] L. Shiqi, G. Jia, W. Jianyu, and M. Xuan, "Recognition algorithm for small packages of traditional Chinese medicine based on improved YOLOV4," in *Proc. Int. Conf. Cyber-Phys. Social Intell. (ICCSI)*, Dec. 2021, pp. 1–4.

[18] G. Tian, X. Li, Y. Wu, A. Liu, Y. Zhang, Y. Ma, W. Guo, X. Sun, B. Fu, and D. Li, "Recognition effect of models based on different microscope objectives," in *Proc. 3rd Int. Symp. Artif. Intell. Med. Sci.*, Oct. 2022, pp. 133–141.

[19] A. Yakovlev and O. Lisovychenko, "An approach for image annotation automatization for artificial intelligence models learning," *Adapt. Syst. Autom. Control*, vol. 1, no. 36, pp. 32–40, 2020.

[20] Z. Gui, J. Chen, Y. Li, Z. Chen, C. Wu, and C. Dong, "A lightweight tea bud detection model based on YOLOv5," *Comput. Electron. Agricult.*, vol. 205, Feb. 2023, Art. no. 107636.

[21] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[22] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13708–13717.

[23] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539.

[24] S. Woo, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[25] L. Du, Z. Lu, and D. Li, "Broodstock breeding behaviour recognition based on Resnet50-LSTM with CBAM attention mechanism," *Comput. Electron. Agricult.*, vol. 202, Nov. 2022, Art. no. 107404.

[26] M. Canayaz, "C+EffxNet: A novel hybrid approach for COVID-19 diagnosis on CT images based on CBAM and EfficientNet," *Chaos, Solitons Fractals*, vol. 151, Oct. 2021, Art. no. 111310.

[27] L. Zhu, X. Wang, Z. Ke, W. Zhang, and R. Lau, "BiFormer: Vision transformer with bi-level routing attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 10323–10333.

[28] X. Hu, D. Kong, X. Liu, J. Zhang, and D. Zhang, "FM-STDNet: High-speed detector for fast-moving small targets based on deep first-order network architecture," *Electronics*, vol. 12, no. 8, p. 1829, Apr. 2023.

[29] P. Purkait, C. Zhao, and C. Zach, "SPP-Net: Deep absolute pose regression with synthetic views," 2017, *arXiv:1712.03452*.

[30] D. Wu, S. Jiang, E. Zhao, Y. Liu, H. Zhu, W. Wang, and R. Wang, "Detection of camellia Oleifera fruit in complex scenes by using YOLOv7 and data augmentation," *Appl. Sci.*, vol. 12, no. 22, p. 11318, Nov. 2022.

JIN YAO was born in January 1983. She received the degree from Zhejiang University of Traditional Chinese Medicine.

She is currently with Huzhou Central Hospital, specializing in traditional Chinese medicine.

WEITAO CHEN was born in May 1999. He is currently pursuing the master's degree with the School of Information Engineering, Huzhou University, Huzhou, China.

His main research interests include object detection, image processing, and image segmentation.

TAO WANG was born in March 1998. He is currently pursuing the master's degree with the School of Information Engineering, Huzhou University, Huzhou, China.

His main research interests include artificial intelligence and image segmentation.

FU YANG was born in November 1999. He is currently pursuing the master's degree with the School of Information Engineering, Huzhou University, Huzhou, China.

His main research interests include deep learning, hyperspectral imaging, and carbon flux processing.

XIAOYAN SUN was born in April 1969. She received the degree from Zhejiang University of Traditional Chinese Medicine.

She was with Huzhou Central Hospital. She has engaged in traditional Chinese medicine work for more than 30 years.

CHONG YAO was born in December 1984.

He is currently the Deputy Director of Huzhou Central Hospital. His current research interests include the quality evaluation of Chinese herbal pieces and the development of traditional Chinese medicine preparations.

LIANGQUAN JIA was born in October 1983.

He is currently an AI Engineer with the School of Information Engineering, Huzhou University. He is also a Postdoctoral Researcher with Zhejiang University. His current research interests include object detection, semantic segmentation, and image processing.