CA-YOLOv5: Detection model for healthy and diseased silkworms in mixed conditions based on improved YOLOv5

Hongkang Shi[1,2], Wenfu Xiao[2], Shiping Zhu[1,3,4*], Linbo Li[1,2], Jianfei Zhang[2]

(1. *College of Engineering and Technology, Southwest University, Chongqing 400700, China*;
2. *Sericultural Research Institute, Sichuan Academy of Agricultural Sciences, Nanchong 637000, Sichuan, China*;
3. *Yibin Academy of Southwest University, Yibin 644000, Sichuan, China*;
4. *State Key Laboratory of Resource Insects, Southwest University, Chongqing 400700, China*)

Abstract: The accurate identification and localization of diseased silkworms is an important task in the research of disease precision control technology and equipment development in the sericulture industry. However, the existing deep learning-based methods for this task are mainly based on image classification, which fails to provide the location information of diseased silkworms. To this end, this study proposed an object detection-based method for identifying and locating healthy and diseased silkworms. Images of mixed healthy and diseased silkworms were collected using a mobile phone, and the category and location of each silkworm were labeled using LabelImg as a labeling tool to construct an image dataset for object detection. Based on the one-step detection model YOLOv5s, the ConvNeXt-Attention-YOLOv5 (CA-YOLOv5) model was designed in which the large kernel with depth-wise separable convolution (7×7 dw-conv) of ConvNeXt was adopted to expand receptive fields and the channel attention mechanism ECANet was added to enhance the capability of feature extraction. Experiments showed that the mean average precision (mAP) values of CA-YOLOv5 for healthy and diseased silkworms reached 96.46%, which is 1.35% better than that achieved via YOLOv5s. At the same time, the overall performance of CA-YOLOv5 was significantly better than state-of-the-art one-step models, such as Single Shot MultiBox Detector (SSD), CenterNet, and EfficientDet, and even improved YOLOv5 using image attention mechanism and a lightweight backbone, like SENet-YOLOv5 and MobileNet-YOLOv5. The results of this study can provide an important basis for the accurate positioning of diseased silkworms in precision disease control technology and equipment development.

Keywords: diseased silkworm detection, YOLOv5, mixed conditions, image attention mechanism, object detection

DOI: 10.25165/j.ijabe.20231606.7854

Citation: Shi H K, Xiao W F, Zhu S P, Li L B, Zhang J F. CA-YOLOv5: Detection model for healthy and diseased silkworms in mixed conditions based on improved YOLOv5. Int J Agric & Biol Eng, 2023; 16(6): 236–245.

1 Introduction

The silkworm (*Bombyx mori*) is an insect with high economic value, mainly used for silk production and widely reared in China, India, and Southeast Asia. Due to their small size and the being density of rearing boxes, silkworms are highly susceptible to disease infestation, which directly leads to mortality or non-cocooning. Silkworm diseases are generally difficult to treat effectively with drugs and are highly contagious, spreading rapidly in a short period of time and causing widespread infection[1,2]. Therefore, disease prevention and control is an essential task when raising silkworms. The most common method of disease prevention comprises strict disinfection and sterilization measures. At the same time, timely identification and discarding of diseased silkworms is also critical to cut off the spread of the pathogen and prevent further infection. In traditional small-scale sericulture, diseased silkworms can be screened out by manual identification. However, with the development of modern agriculture, the mode of sericulture is gradually shifting to large-scale mechanized and intelligent farming[3], in which manual identification and screening cannot satisfy the demand. Thus, there is an urgent need for an efficient and accurate method for locating diseased silkworms to support the research and equipment development for accurate disease control technology.

With the recent rapid development of artificial intelligence in agriculture, deep learning, and vision technology have become widespread in silkworm breeding. Shi et al.[4] proposed a recognition method for silkworm species using MobileNet. Deep learning was applied to the identification of male and female silkworm pupae by Yu et al.[5]. Li et al.[6] conducted a study on the quality sorting mechanism of silkworm cocoons based on computer vision. He et al.[7] and Wen et al.[8] proposed a method for silkworm individual detection using object detection and semantic segmentation for precision feeding. In terms of diseased silkworm identification, Shi et al.[9] introduced an improved ResNet-based recognition model to achieve the identification of five types of diseased silkworms. Xia et al.[10] proposed a DenseNet-based silkworm disease identification model. Ding and Cheng[11] presented a method for image recognition of diseased silkworms based on the feature maps slicing and AlexNet architecture. However, the above studies on diseased silkworm identification are mainly based on image classification

Received date: 2022-08-15 **Accepted date:** 2023-04-19

Biographies: Hongkang Shi, PhD candidate, Research Associate, research interest: intelligent recognition of silkworms, Email: swushk@163.com; **Wenfu Xiao**, PhD, Associate Researcher, research interest: genetic breeding of silkworms, Email: wenfu_xiao1983@foxmail.com; **Linbo Li**, MS candidate, Research Associate, research interest: intelligent agricultural equipment, Email: bob870126@qq.com; **Jianfei Zhang**, MS, Researcher, research interest: intelligent equipment and technology of silkworm rearing, Email: 783890694@qq.com.

***Corresponding author: Shiping Zhu**, PhD, Professor, research interest, computer vision, deep learning, and intelligent detection. College of Engineering and Technology, Southwest University, Chongqing 400700, China. Tel: +86-23-68250803, Email: zspswu@126.com.

methods, which means that only one diseased or healthy silkworm can be present in each image, and thus fail to achieve localization of diseased silkworms under the condition of mixing healthy and diseased silkworms.

Object detection is an important branch of deep learning that allows a single image to contain multiple classes of objects at the same time, and possesses the ability to not only identify the class of each object in the image but also to predict the location of each object. The study of object detection-based diseased silkworm localization models can ensure the ability of this technique to be applied in real-world environments. Based on the detection process, object detection models can be divided into one-step and two-steps detection networks, with the former being more efficient and the latter being more accurate. In recent years, with the great potential shown by the field of Transformer, researchers have also proposed Vision Transformer-based object detection algorithms[12]. However, the exponential computational burden makes it highly challenging to train from scratch. Among the common one-step detection models, the YOLO series[13] is a highly representative model that has long been favored by researchers for its ability to maintain a high detection efficiency while having detection accuracy comparable to that of two-steps models. YOLOv5[14], the 5th generation version released in 2020, uses CSPNet[15] as the backbone for the feature extraction network and employs techniques such as Focus, SPP, and PANet to improve detection performance; therefore, it has received widespread attention since its release. Scholars have applied YOLOv5 to a variety of vision tasks, with promising results. For example, Qi et al.[16] proposed a tomato leaf disease detection model by adding SENet to YOLOv5. Ning et al.[17] proposed a face recognition method for dairy goats by adding SimAM to the feature extraction layer of YOLOv5. Yang et al.[18] derived a pig-counting algorithm based on SENet and YOLOv5. Xue et al.[19] developed an estrus detection method for parturient sows based on model compression and YOLOv5. Ma et al.[20] proposed a locust recognition in Ningxia grassland, Guyuan, Ningxia, China, by using Bi-FPN to enhance the capability of feature fusion and interaction of YOLOv5. Wang et al.[21] achieved the detection of the invasive weed Solanum rostratum Dunal seedlings by adding Convolutional Block Attention Module (CBAM) to YOLOv5. Li et al.[22] built a wheat ear detection algorithm based on YOLOv5 and image attention mechanism.

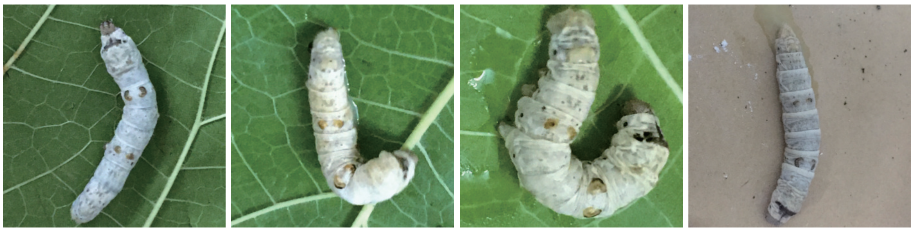
The above-mentioned studies illustrated that YOLOv5 has excellent detection performance in various vision tasks. Hence, a YOLOv5-based network was proposed for accurate and efficient detection of healthy and diseased silkworms. To this end, images of mixed healthy and diseased silkworms were collected from real environments, and the category and location of each silkworm were labeled to construct a detection image dataset. Then, a ConvNeXt-Attention-YOLOv5 (CA-YOLOv5) model was proposed based on the original YOLOv5s, in which the large kernel with depth-wise separable convolution (7×7 dw-conv) of ConvNeXt[23] was adopted to expand receptive fields, and the channel attention mechanism ECANet[24] was added to enhance the capability of feature extraction. Experiments showed that CA-YOLOv5 outperforms original YOLOv5, one-step algorithms such as CenterNet, EfficentDet, and Single Shot MultiBox Detector (SSD), as well as other improved YOLOv5 networks based on image attention mechanism and a lightweight backbone, like SENet-YOLOv5 and MobileNet-YOLOv5. This research can be a cornerstone for the research of disease precision control technology and equipment development, especially in silkworm breeding.

2 Materials and methods

2.1 Experimental data

2.1.1 Image acquisition

Among the many types of silkworm diseases, this study took diseased silkworms infected by nuclear polyhedrosis virus (NPV) as detection objects, as well as healthy silkworms. This disease has the greatest frequency and is highly infectious among silkworms, accounting for more than half of all silkworm diseases in China[25-26]. The appearance of healthy and NPV-infected silkworms is shown in Figure 1. It can be seen that the appearance of healthy silkworms is white and greenish, while silkworms infected with NPV show a dark yellow coloration, with pus flowing around the body.


a. Healthy silkworm

b. Diseased silkworms

Figure 1 Image examples of healthy and diseased silkworms

In order to obtain a dataset for model training and evaluation, in this study, all silkworm images were collected in a real environment. This was performed in May 2020 and May 2021 at the Institute of Sericulture, Sichuan Academy of Agricultural Sciences, Nanchong City, Sichuan Province, China, under natural indoor lighting, using a smartphone (iPhone 6S) with a resolution of 12 megapixels. When acquiring the image, the acquisition equipment was directed vertically downward, mulberry leaves were used as the image background, and several healthy silkworms and diseased silkworms were simultaneously placed on the background to simulate the coexistence of healthy and diseased silkworms in a real scenario when diseased silkworms need to be located and removed in time. A total of 4023 original images were collected, which included 5459 diseased silkworms and 4931 healthy silkworms. Some of the collected images are shown in Figure 2.

2.1.2 Image annotation and dataset

The acquired image size was 3224×3224 pixels, which is far beyond the input image size of YOLOv5. Bilinear interpolation was used to rescale all image sizes to 640×640 pixels. To train the object detection model of supervised learning, all images needed to be labeled first. The labeling tool LabelImg[27] was used to label the category and location of each silkworm. The labeling interface is shown in Figure 3, where each silkworm is represented by an external rectangular box with coordinates representing its location in the image, the label "H" represents a healthy silkworm, and "NP" represents a diseased silkworm infected by nuclear polyhedrosis.

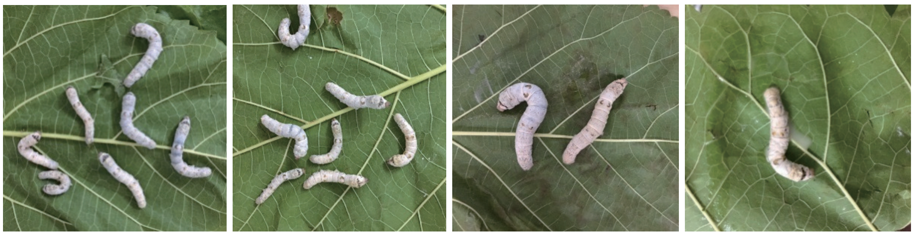


Figure 2 Examples of original images of silkworms

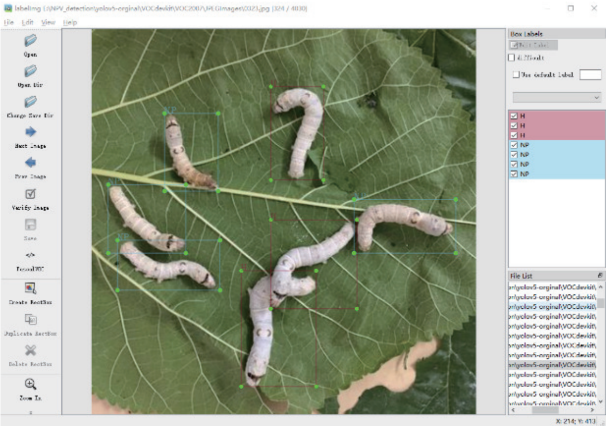


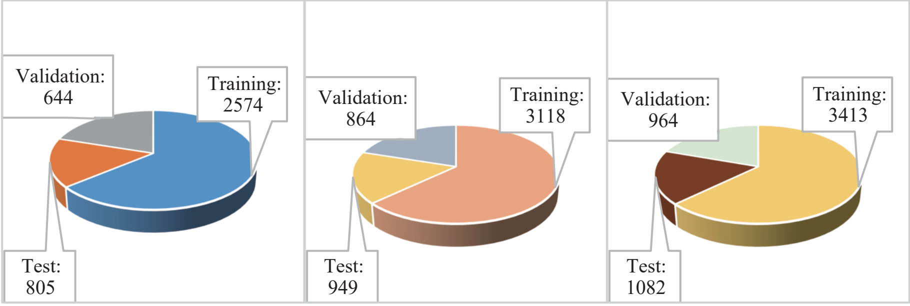
Figure 3 Labeling interface

After labeling, all images were divided into a training set and a test set in the ratio of 8:2. When model training, 20% of the images from the training set were randomly selected for the validation set. The number of images and detection objects in the training set, validation set, and test set are shown in Figure 4.

2.2 CA-YOLOv5 for detection of healthy and diseased silkworms

2.2.1 ConvNeXt-Attention block (CA-block) for feature extraction

The feature extraction capability of the object detection model determines the detection performance. YOLOv5 uses CSPNet as the backbone network to extract image features, and its basic structure is shown in Figure 5a, which mainly uses 3×3 convolutional layers to extract image features. The input image size for object detection is 640×640×3, which is far larger than the input size of 224×224×3 for general classification tasks. The detection model using CSPNet as the backbone network has limited receptive fields, which would result in insufficient feature extraction. Therefore, a ConvNeXt-Attention block (CA-block) was designed without changing the gradient transmission path of the original architecture. In the CA-block, the large kernel and depth-separable convolution from ConvNeXt were introduced to expand the receptive fields and reduce the number of parameters, and the image attention mechanism ECANet was further added to enhance the feature extraction capability. The structure of the CA-block is shown in Figure 5b.



a. Number of total images b. Number of healthy silkworms c. Number of diseased silkworms

Figure 4 Details of constructed dataset

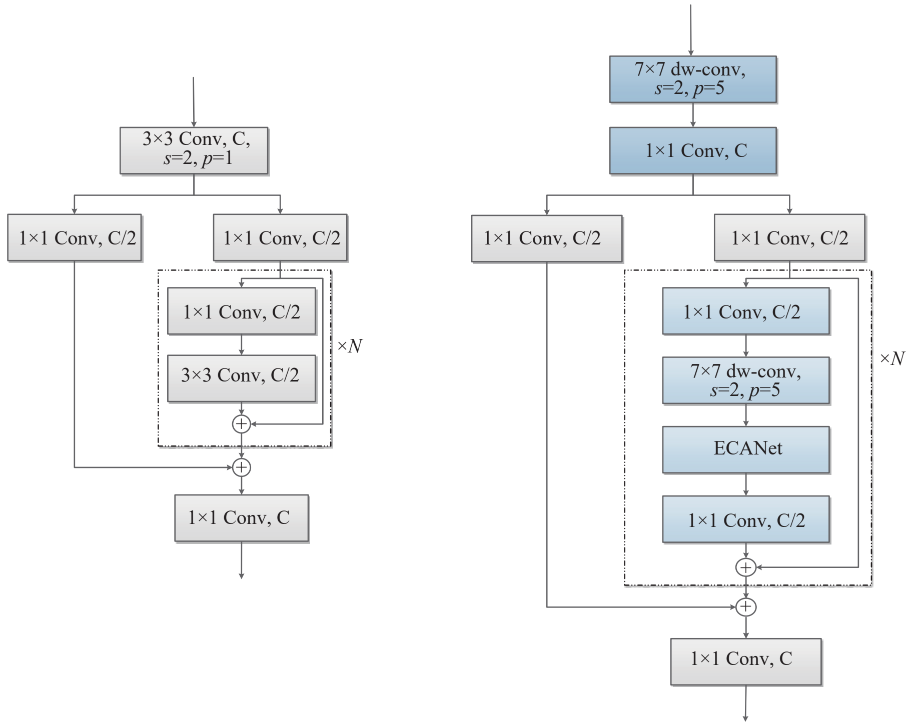
In Figure 5, 3×3 Conv represents convolutional operation using kernel size of 3×3, C refers to the number of channels, s means stride, and p is padding. The CA-block first pads 5 pixels for the width and height of the input feature map to keep the feature dimension, and extract image features by using a 7×7 depth-separable convolutional (7×7 dw-conv) and a 1×1 convolutional (conv) layer. Next, the feature map is compressed into 2 sub-feature maps using two 1×1 conv layers, and the number of channels in the sub-feature map is 1/2 of the original one. The subsequent feature extraction will be performed on one of the sub-feature maps with 1×1 conv, 7×7 dw-conv, and residual connection, respectively, with a number of loops executed in N. ECANet is embedded between the 7×7 dw-conv and 1×1 conv layers in the last loop. Another sub-feature map is connected as a residual block across to the features in the backward layer. Finally, the feature map is output after performing a 1×1 conv operation.

Compared to the original block, the main advantage of the CA block is that it has larger receptive fields. Meanwhile, the image attention mechanism further enhances the feature extraction effect, thus ensuring the feature extraction capability of the model when detecting healthy and diseased silkworms.

2.2.2 ECANet

The image attention mechanism can enhance the extraction capability of key information and suppress interference information to ensure better detection results[28]. It mainly includes spatial attention mechanism and channel attention mechanism. Given that 7×7 dw-conv has been utilized to expand receptive fields in CA-block, which plays a role similar to that of spatial attention mechanism, channel attention mechanism ECANet is chosen to enhance the capability of feature extraction.

Specifically, ECANet was designed based on SENet[29]. Due to the advantage of one-dimensional convolution operation instead of fully connection operation, it is possible to achieve local cross-channel interaction without dimensionality reduction. Adding



a. Original block of CSPNet b. CA-block proposed by CA-YOLOv5

Note: 3×3 Conv represents convolutional operation using kernel size of 3×3; *C* refers to the number of channels; *s* means stride; *p* is padding; dw-conv is depth-separable convolutional; + represents feature addition; N refers to N times stacked and is set to 1, 3, 3, and 1, respectively, in the backbone. Same below.

Figure 5 Basic block of CSPNet and CA-YOLOv5

ECANet does not cause a significant increase in the number of parameters and computational burden. As illustrated in Figure 6, for given feature maps $X \in R^{W \times H \times C}$, where *W*, *H*, and *C* represent the width, height, and channel number of the feature map, respectively. ECA module first aggregates feature maps in spatial dimension using global average pooling according to the following equation:

$$Y = \mathrm{GAP}(X) \tag{1}$$

where, GAP refers to global average pooling and $Y \in R^{1 \times 1 \times C}$.

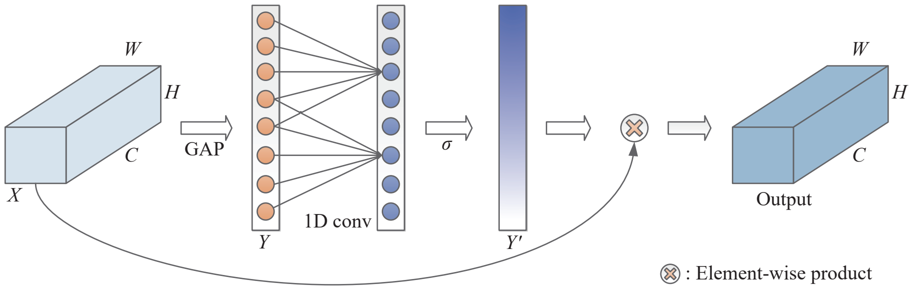
Then, a one-dimensional convolution and activation operation is adopted to obtain the attention weights, and the formula is as follows:

$$Y' = \delta(\mathrm{C1D}_k(Y)) \tag{2}$$

where, C1D refers to one-dimensional convolutional operation, subscript k is the kernel size, and δ means activation function, and the formula is as follows:

$$\delta(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

Subsequently, the broadcast operation was used to expand the shape of $Y' \in R^{1 \times 1 \times C}$ to $Y' \in R^{W \times H \times C}$. The refined feature map is finally obtained by multiplying Y' by input feature map X.



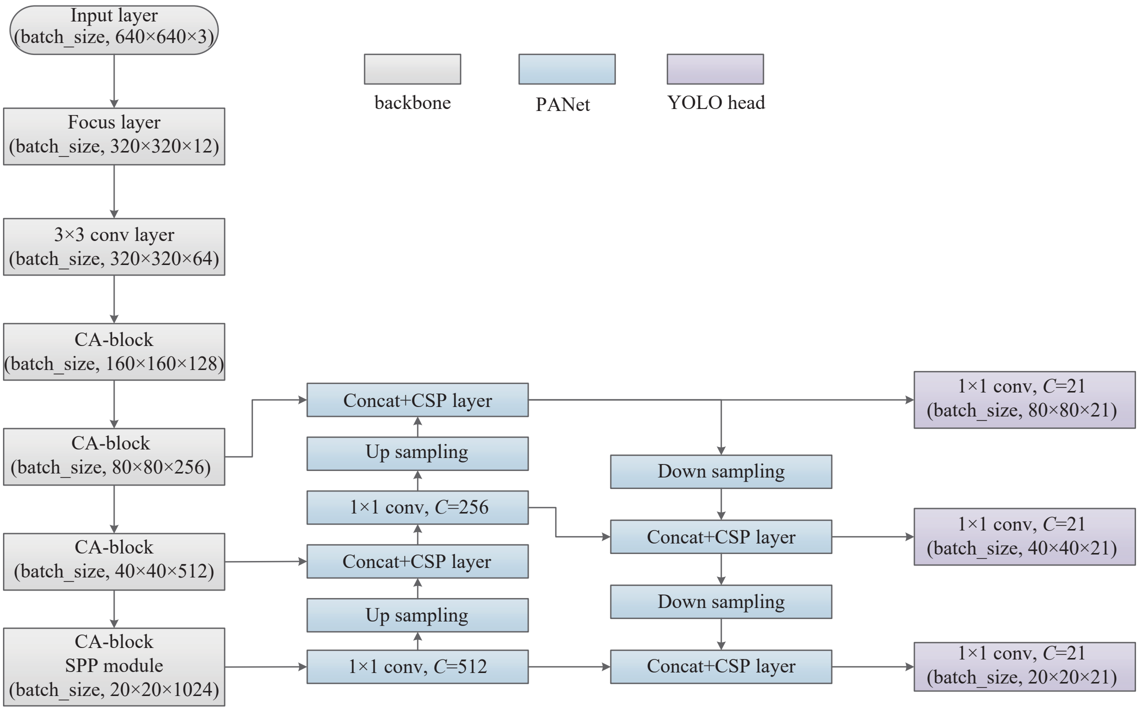
Note: *W*, *H*, and *C* represent the length, width, and channel number of the feature map; GAP means the global average pooling. *X* and *Y* are feature maps; 1D conv refers to one-dimensional convolution.

Figure 6 Schematic diagram of ECANet

2.2.3 Structure of CA-YOLOv5

YOLOv5 is available in four different models, YOLOv5s, ~l, ~x, and ~m, and the main difference between them is the number of convolutional layers and convolutional kernels. To ensure the detection efficiency, YOLOv5s was selected to design CA-YOLOv5 in this study.

The structure of CA-YOLOv5 is shown in Figure 7, which is basically the same as YOLOv5 and consists of the backbone, PANet, and YOLO head. The CA-block is used in the backbone and PANet to replace the original feature extraction block. In addition to CA-block, the backbone also includes a Focus layer, 3×3 convolutional layer, and Spatial Pyramid Pooling (SPP) module. For an input image with a size of 640×640×3, the Focus layer was first adopted to obtain tensor with a size of 320×320×12. As shown in Figure 8, the principle of the Focus layer is to stack the input image after sampling the pixels at certain intervals, to reduce dimensionality and enrich the semantic information of the input image. After a 3×3 convolution layer, four stages of CA-block were performed to extract image features. An SPP module is embedded in the last stage of the CA-block to obtain feature maps with diversified resolutions by using different sizes of pooling kernels. The principle of the SPP module is shown in Figure 9.


Input layer (batch_size, 640×640×3)
Focus layer (batch_size, 320×320×12)
3×3 conv layer (batch_size, 320×320×64)
CA-block (batch_size, 160×160×128)
CA-block (batch_size, 80×80×256)
CA-block (batch_size, 40×40×512)
CA-block SPP module (batch_size, 20×20×1024)
backbone
PANet
YOLO head
Concat+CSP layer
Up sampling
1×1 conv, C=256
Concat+CSP layer
Up sampling
1×1 conv, C=512
Down sampling
Concat+CSP layer
Down sampling
Concat+CSP layer
1×1 conv, C=21 (batch_size, 80×80×21)
1×1 conv, C=21 (batch_size, 40×40×21)
1×1 conv, C=21 (batch_size, 20×20×21)

Note: CA-block: ConvNeXt-Attention block; SPP: Spatial Pyramid Pooling; CSP layer refers to the loop structure of CA-block.

Figure 7 Structure of CA-YOLOv5

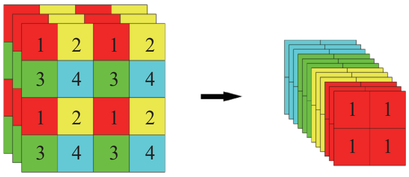


Figure 8 Schematic diagram of the focus layer

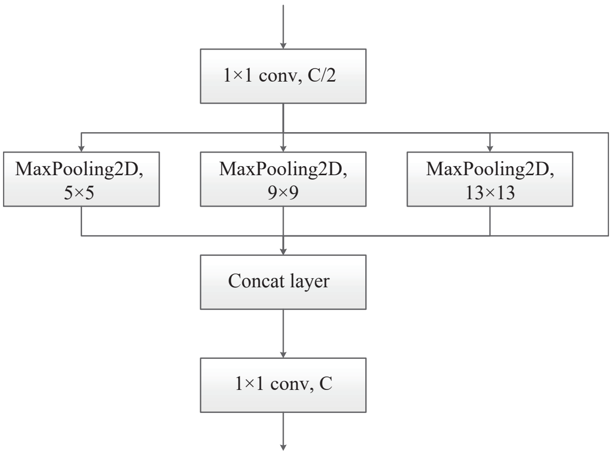


Figure 9 Principle of SPP module

After the backbone extraction, three effective feature maps with dimensions of 80×80, 40×40, and 20×20 were input to PANet for a series of up- or down-sampling for feature fusion, in order to enhance the feature representation capability of the network. In PANet, CSP layer refers to the loop structure of CA-block. Subsequent to PANet fusion, three feature maps were input to YOLO head directly for obtaining prediction results using a 1 × 1 convolution operation. The number of convolution kernels in the YOLO head is related to the number of categories of dataset, and its calculation formula is

$$K = (N+5)\times N_c \tag{4}$$

where, K means the number of filters in the last convolution operation, N refers to the number of detection object categories; N_c is the number of anchor boxes. This research includes healthy and diseased silkworms, and the number of anchor boxes in each YOLO head was 3, it can be concluded that K is 21 by using Equation (4).

In CA-YOLOv5, the Batch Normalization operation and SiLU function were also used after each convolutional layer to accelerate model convergence. The specific calculation formula of SiLU is as follows:

$$\mathrm{SiLU} = x\cdot\delta = \frac{x}{1+e^{-x}} \tag{5}$$

2.3 Experimental environment and evaluation indicators

2.3.1 Experimental environment

All experiments were operated on a Dell Precision 5820 workstation with an Intel® Core i7-9800X processor, and RTX2080Ti GPUs, with 11 GB memory, and the CUDA-10.0 computing platform. The operating system was Windows 10 Professional (64 bits), the programming language was Python3.7, the programming environment was Jupyter Notebook, and the deep learning framework was TensorFlow-GPU 1.14. The toolkits used include Numpy, Keras, PIL, etc.

When training CA-YOLOv5, YOLOv5, and YOLOv5-based models, the hyper-parameters of model training included the number of epochs was 300, and the mini-batch size was 8. The learning rate was initially 0.001 via cosine decay unless otherwise stated subsequently, with a momentum of 0.9, and a weight decay of 0.05[30]. The GIOU_loss[31] was used as a bounding boxes regression loss function and the Adam was used as the optimizer. The Mosaic data enhancement[32] was used to improve the robustness of the model in the first 200 iterations of training. The IoU value was 0.5, and the confidence threshold of the prediction result was 0.3. The anchor boxes proposed by YOLOv5 were used for data encoding and prediction result decoding. When training other compared models, the number of epochs, the mini-batch size, and the optimizer were the same as that of YOLOv5-based models, while the initial learning rate and loss function were referenced from the original paper and its code on Github (Available at https://github.com/bubbliiiing?tab=repositories).

2.3.2 Evaluation indicators

The precision, recall, F1-score, average precision (AP), and mean average precision (mAP) were used for model evaluation. The specific calculation formulas are as follows, respectively:

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}}\times 100\% \tag{6}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}}\times 100\% \tag{7}$$

$$\text{F1}-\text{score}=\frac{2\cdot\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}} \tag{8}$$

$$\text{AP}=\sum_{1}^{N}\int_{0}^{1}\text{Precision(Recall)d}R\times 100\% \tag{9}$$

$$\text{mAP}=\frac{\sum_{1}^{N}\int_{0}^{1}\text{Precision(Recall)d}R}{N}\times 100\% \tag{10}$$

where, True Positives (TP) refers to the object being detected as a positive sample and the test result is correct, False Positives (FP) means the object is detected as a positive sample, but the actual object is a negative sample, False Negatives (FN) represents object is detected as a negative sample, but the object is a positive sample. AP is the area enclosed by the precision and recall curve (*P-R* curve). The mAP is the average of all APs. In this study, objects are healthy silkworms and diseased silkworms, respectively, so the value of *N* is 2.

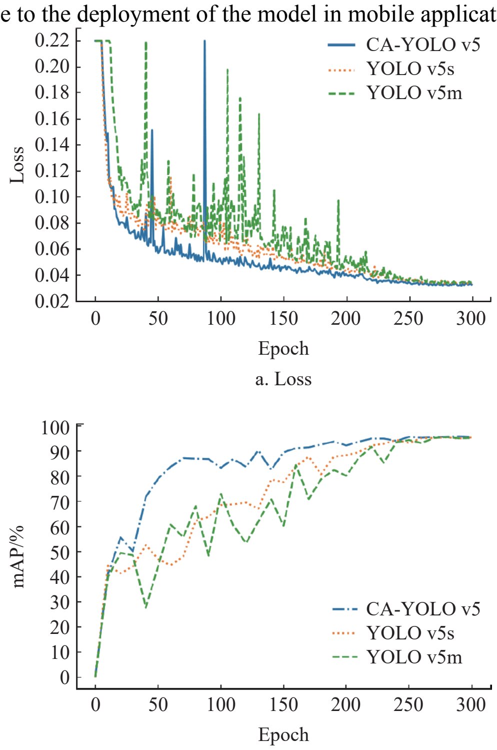
3 Results and discussion

3.1 Comparison of the detection effect with YOLOv5

In this section, CA-YOLOv5, YOLOv5s, and ~m were compared by training and testing in the same environment, and loss values of each iteration and mAP values per 10 iterations of three models on the validation set were recorded for evaluating their convergence speed and effects.

Figure 10 shows the loss and mAP curves of three compared models on the validation set. The loss and mAP values of CA-YOLOv5, YOLOv5s, and YOLOv5m can be seen to fluctuate greatly in the initial stage of training, but they tend to stabilize after roughly 200 iterations as the training continues. It is worth noting that the loss values of CA-YOLOv5 are significantly lower than that of the other 2 models during most of the training epochs, and the corresponding mAP values are higher than 2 original models, which indicates the best convergence was obtained on the validation set.

Table 1 lists the number of parameters, model size, and detection speed of three compared models. It can be seen that CA-YOLOv5 not only has a significant reduction in the number of parameters and model size compared with the 2 original models but also has a significant advantage in detection speed, which is more favorable to the deployment of the model in mobile applications.



a. Loss

b. mAP

Note: mAP: mean Average Precision.

Figure 10 Loss and mAP curves of three models

Table 1 Parameters, model size, and detection speed of three compared models

Model	Parameters (Million)	Model size/MB	Detection speed /(images·s^{-1})
YOLOv5s	7.08	27.4	22.3
YOLOv5m	21.01	81.1	14.6
CA-YOLOv5	**3.95**	**15.6**	**23.0**

After model training, the model weights that reached the best mAP value on the validation set were used as the test model, and the corresponding evaluation metrics of three models were calculated.

Table 2 lists the detection results of three compared models on the test set. It can be established that, compared with YOLOv5s, CA-YOLOv5 shows significant advantages in Recall, Precision, and F1-score, and brings an improvement of 1.35% in mAP value. Compared with YOLOv5m, CA-YOLOv5 only lags behind in Precision of “H”, but has significant advantages in other evaluation indicators, such as Recall and F1-score of “H”, and also brings 1.39% improvement in mAP value. Therefore, CA-YOLOv5 has obvious advantages over the original YOLOv5, both in terms of detection speed and detection accuracy. Moreover, given that the results of YOLOv5m are comparable to those of YOLOv5s, it can be concluded that the use of more convolutional layers and convolutional kernels has a limited effect on improving the detection accuracy on the dataset of this study.

Table 2 Detection results of three compared models on the test set

Model	Recall/%		Precision/%		F1-socre		mAP/%
	H	NP	H	NP	H	NP	
YOLOv5s	88.09	88.26	85.57	90.87	0.87	0.90	95.11
YOLOv5m	83.56	90.11	**89.60**	88.80	0.86	0.89	95.07
CA-YOLOv5	**90.94**	**90.76**	89.52	**92.29**	**0.90**	**0.92**	**96.46**

Note: H means healthy silkworms; NP means diseased silkworms infected by nuclear polyhedrosis; mAP: mean Average Precision. Same below.

3.2 Comparison of the detection effect with one-step models

In order to verify the detection effectiveness of CA-YOLOv5, in this section, several classical one-step networks, including CenterNet[33], EfficientDet[34], YOLOv4[32], SSD[35] and YOLOX[36], were trained and tested in the same environment.

Table 3 reports the number of parameters, model size, and detection speed of several single-step models, from which it can be concluded that the parameters and model size of CA-YOLOv5 are only larger than those of EfficientDet. In terms of detection speed, CA-YOLOv5 outperforms other one-step detection models.

Table 3 Parameters, model size, and detection speed of several one-step models

Model	Parameters (Million)	Model size/MB	Detection speed $/(\text{images}\cdot\text{s}^{-1})$
CenterNet	32.71	125.0	14.6
EfficientDet	**3.88**	**16.1**	15.2
YOLO v4	64.01	244.0	11.7
SSD	23.88	91.1	9.7
YOLO X	8.96	34.7	16.4
CA-YOLOv5	3.95	15.6	**23.0**

Table 4 lists the detection results of the compared one-step models on the test set. It can be seen that the performance of CA-YOLOv5 is inferior to EfficientDet and SSD only in Recall of “NP”, lower than CenterNet in Precision value of “H”, and the same as Efficient and YOLOX in F1-score. Meanwhile, it shows an obvious advantage in other evaluation metrics, especially in the AP and mAP values. It can be concluded that CA-YOLOv5 has the best detection performance compared to selected one-step detection models.

Table 4 Detection results of compared one-step models on the test set

Model	Recall/%		Precision/%		F1-socre		AP/%		mAP/%
	H	NP	H	NP	H	NP	H	NP	
CenterNet	71.87	72.64	**91.42**	85.06	0.80	0.78	83.61	75.27	79.44
EfficientDet	89.15	**92.98**	90.48	91.37	**0.90**	**0.92**	93.75	95.68	94.71
YOLO v4	86.83	86.23	87.85	88.35	0.87	0.87	93.85	94.59	94.22
SSD	89.99	**92.98**	86.88	89.42	0.88	0.91	94.30	96.46	95.38
YOLO X	90.62	91.40	90.15	91.66	**0.90**	**0.92**	95.75	96.54	96.14
CA-YOLOv5	**90.94**	90.76	89.52	**92.29**	**0.90**	**0.92**	**96.02**	**96.90**	**96.46**

3.3 Comparison of the detection effect with other improved YOLOv5

In order to verify the improvement of YOLOv5 by the method proposed in this paper, in this section, two types of YOLOv5-based improvement models were utilized to train and test in the same environment. One of them is based on image attention mechanism and the other is based on a lightweight backbone.

Table 5 lists the parameters, model size, and detection speed of several improved YOLOv5 models. The results show that the number of parameters, model size, and detection speed of CA-YOLOv5 are only slightly lower than those of MobileNetv3-YOLOv5s.

Table 6 lists detection results of several improved YOLOv5 models on the test set, from which it can be obtained that CA-YOLOv5 is only lower than MobileNetv3-YOLOv5s and SENet-YOLOv5s in Precision, while it shows significant advantages in other evaluation metrics. This leads to the conclusion that CA-YOLOv5 has better overall detection performance.

Table 5 Parameters, model size, and detection speed of several improved YOLOv5

Model	Similar studies	Parameters (Million)	Model size/MB	Detection speed
SENet-YOLOv5s	Qi et al.[16] and Yang et al.[18]	7.11	27.6	22.5
ECA-YOLOv5s	Cao et al.[37]	7.09	27.5	18.6
CBAM-YOLOv5s	Wang et al.[21] and Lu et al.[38]	7.13	27.8	21.7
GhostNet-YOLOv5s	Xu et al.[39]	5.61	22.2	20.1
MobileNet v3-YOLOv5s	Zhang et al.[40]	**3.57**	**14.1**	**23.7**
CA-YOLOv5	Proposed in this study	3.95	15.6	23.0

Table 6 Detection results of several improved YOLOv5 models on the test set

Model	Recall/%		Precision/%		F1-socre		AP/%		mAP/%
	H	NP	H	NP	H	NP	H	NP	
SENet-YOLOv5s	87.67	87.34	89.17	**92.47**	0.88	0.90	95.27	96.42	95.85
ECA-YOLOv5s	89.25	87.99	88.32	92.43	0.89	0.90	95.09	96.58	95.84
CBAM-YOLOv5s	88.20	86.41	86.38	90.95	0.87	0.89	93.83	95.85	94.84
GhostNet-YOLOv5s	89.36	88.72	87.51	89.72	0.88	0.89	94.47	95.73	95.10
MobileNet v3-YOLOv5s	77.45	84.75	**89.96**	87.84	0.83	0.86	92.74	93.91	93.33
CA-YOLOv5	**90.94**	**90.76**	89.52	92.29	**0.90**	**0.92**	**96.02**	**96.90**	**96.46**

3.4 Visualization of detection results

In this section, six images on the test set were selected for visualization, and their labeled results and the detection results of CA-YOLOv5 were also visualized, as shown in Figure 11.

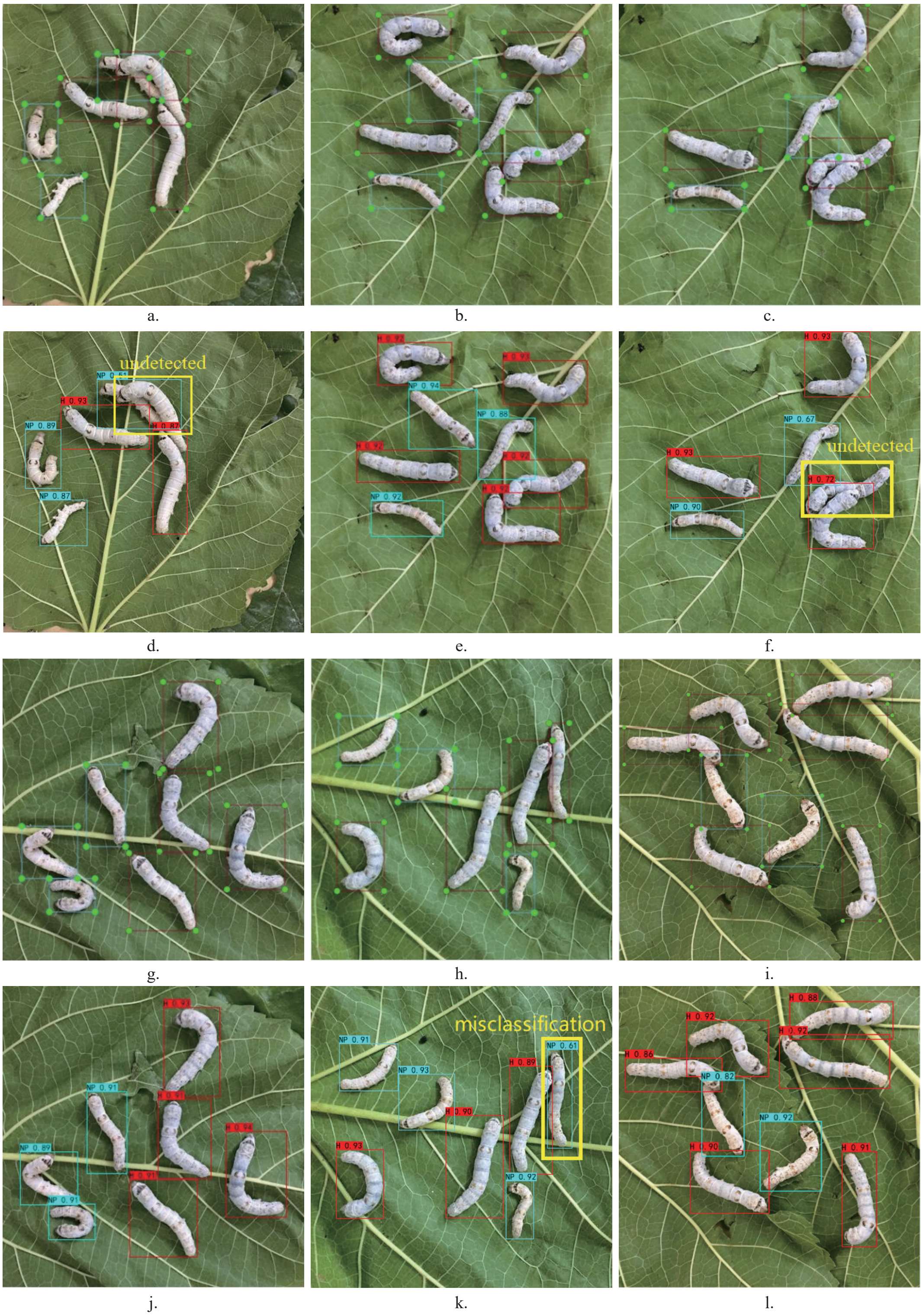
Figures 11a-11c and Figures 11g-11i are original images labeled by the LabelImg. In these figures, the brown boxes (software default settings) represent healthy silkworms, and the light green boxes (software default settings) represent diseased silkworms. Figures 11d-11f and Figures 11j-11l are detection results using the trained CA-YOLOv5. In these figures, the blue boxes represent diseased silkworms, and the red boxes represent healthy silkworms. The number in the upper left corner of the box is the probability value predicted by the model. The yellow boxes are labeled manually which represent undetected or misclassification.

As shown in Figures 11e, 11i, and 11j, all silkworms were correctly detected, which illustrates the performance of CA-YOLOv5. There is one undetected silkworm in Figures 11d and 11f, respectively, owing to overlapping. A silkworm was misclassification in Figure 11k, due to its body shape being relatively small.

Through visualization, it was found that whether the postures of silkworms are horizontal, vertical, diagonal, or even twisted, CA-YOLOv5 can predict the location boundary of silkworms accurately, not to mention the category of silkworms. However, when 2 silkworms are overlapping, detection reliability may decrease remarkably. Therefore, future studies should mainly concentrate on the detection of overlapping silkworms in higher density, and the early stage of disease when there are only minor differences between healthy silkworms and diseased silkworms.

4 Conclusions

In this study, a detection method for healthy and diseased silkworms was proposed using object detection technology. An image dataset for object detection was constructed, and the most of images in the dataset contain both healthy and diseased silkworms.



a. b. c.

d. e. f.

g. h. i.

j. k. l.

Note: a-c and g-i are original images labeled by the LabelImg. In a-c and g-i, brown boxes represent healthy silkworms, and light green boxes represent diseased silkworms. d-f and j-l are detection results using the trained CA-YOLOv5. In d-f and j-l, the blue boxes represent diseased silkworms and the red boxes represent healthy silkworms.

Figure 11 Visualization of detection results of silkworms

State-of-the-art deep learning architectures, such as YOLOv5, ConvNeXt, and ECANet, were architecturally improved to design a CA-YOLOv5 network to effectively and accurately detect healthy and diseased silkworms in mixed conditions. Based on the results, the following specific conclusions can be drawn:

1) An object detection-based method was proposed for identifying and locating healthy and diseased silkworms. which can predict location information of diseased silkworms in mixed conditions, so as to provide reference for accurate disease control research and equipment development;

2) CA-YOLOv5 was proposed based on YOLOv5s, in which the large kernel with 7×7 dw-conv of ConvNeXt was adopted to expand receptive fields, and ECANet was added to enhance the capability of feature extraction. Experiments showed that the overall performance of CA-YOLOv5 is not only better than the original YOLOv5 but also better than several one-step models (such as

Single Shot MultiBox Detector (SSD), EfficientDet, CenterNet, and so on), as well as improved YOLOv5 based on image attention mechanism and lightweight backbone (like SENet-YOLOv5, MobileNet-YOLOv5 and so on).

Nevertheless, this study still has some deficiencies: 1) The silkworm density of the dataset in this experiment was less than that in real-rearing environments, and only partial growth stages of a single silkworm variety were imaged and detected. Hence, the diversity and richness of the dataset need to be enhanced; 2) An image-based approach was proposed to detect the diseased silkworm, which only considered the visual features of healthy and diseased silkworms. Other characteristics between healthy and diseased silkworms, such as morphological and behavioral characteristics, were not considered. Due to the visual features depending on image acquisition, which can be affected by the environment, acquisition equipment, and rearing mode, further research on multi-feature fusion detection is needed to meet the practical requirements based on this study. ·

In future works, we plan to enrich the diversity of the dataset, for example, by including more silkworm species and growth stages. We also aim to focus on behavioral differences between healthy silkworms and diseased silkworms in the early infection stages and explore the early detection of diseased silkworms by fusing behavioral and visual characteristics

Acknowledgements

The authors acknowledge the financial support provided by the Sichuan Science and Technology Program, China (Grant No. 2023NSFSC0498) and the National Modern Agricultural Industrial Technology System Special Project, China (Grant No. CARS-18).

[References]

[1] Jiang L, Zhao P, Xia Q Y. Research progress and prospect of silkworm molecular breeding for disease resistance. Acta Sericologica Sinica, 2014; 40(4): 571–575. (in Chinese)

[2] Xu A Y, Qian H Y, Sun P J, Liu M Z, Lin C L, Li G, et al. Breeding of a new silkworm variety Huakang 3 with resistance to *Bombyx mori* Nucleopolyhedrosis. Acta Sericologica Sinica, 2019; 45(2): 201–211. (in Chinese)

[3] Shi H K, Jiang M, Li L B, Wu J M, Ye J J, Ma Y, et al. Design of young silkworm feeding machine with spiral lifting system and its production test. Acta Sericologica Sinica, 2018; 44(6): 891–897. (in Chinese)

[4] Shi H K, Tian Y Y, Yang C, Chen Y, Su S Y, Zhang Z Y, et al. Research on intelligent recognition of silkworm larvae species based on convolutional neural network. Journal of Southwestern University (Natural Science Edition), 2020; 42(12): 34–45. (in Chinese)

[5] Yu Y D, Gao P F, Zhao Y Z, Pan G Q, Chen T. Automatic male and female identification of silkworm pupae based on deep convolutional neural network. Acta Sericologica Sinica, 2020; 46(2): 197–203. (in Chinese)

[6] Li S J, Sun W H, Liang M, Shao T F, Shen J. Design of real-time cocoon species detection system based on deep learning. Shanghai Textile Science & Technology, 2021; 49(11): 53–55,58. (in Chinese)

[7] He R M, Zheng K F, Wei Q Y, Zhang X B, Zhang J, Zhu Y H, et al. Identification and counting of silkworms in factory farm using improved Mask R-CNN model. Smart Agriculture, 2022; 4(2): 163–173. (in Chinese)

[8] Wen C M, Wen J, Li J H, Luo Y Y, Chen M B, Xiao Z P, et al. Lightweight silkworm recognition based on Multi-scale feature fusion. Computers and Electronics in Agriculture, 2022; 200: 107234.

[9] Shi H K, Xiao W F, Huang L, Hu C W, Hu G R, Zhang J F. Research on recognition of silkworm diseases based on Convolutional Neural Network. Journal of Chinese Agricultural Mechanization, 2022; 43(1): 150–157. (in Chinese)

[10] Xia D Y, Zhen Y, Cheng A J. Development and application of silkworm disease recognition system based on mobile App.Beijing: In: 10th International Conference on Image and Graphics (ICIG 2019), Cham: Springer, 2019; pp.471-482. doi: 10.1007/978-3-030-34110-7_39.

[11] Ding J Y, Cheng A J. An improved similarity algorithm based on deep hash and code bit independence. In: 2019 4th International Conference on Insulating Materials, Material Application and Electrical Engineering. 2019; 440: 032079. doi: 10.1088/1755-1315/440/3/032079.

[12] Nicolas C, Francisco M, Gabriel S, Nicolas U, Alexander K, Sergey Z. End-to-end object detection with transformers. arXiv e-Prints archive, 2020. arXiv: 2005.12872.

[13] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement. arXiv e-Prints archive, 2018. arXiv: 1804.02767.

[14] Glenn J. yolov5. Git code. Available: https://github.com/ultralytics/yolov5. Accessed on [2022-03-20].

[15] Wang C Y, Liao M H Y, Yeh I, Wu Y H, Chen P Y, Hsieh J W. CSPNet: A new backbone that can enhance learning capability of CNN. arXiv e-Prints archive, 2019. arXiv: 1911.11929.

[16] Qi J T, Liu X N, Liu K, Xu F R, Guo H, Tian X L, et al. An improved YOLOv5 model based on visual attention mechanism: Application to recognition of tomato virus disease. Computer and Electronics in Agriculture, 2022; 194: 106780.

[17] Ning J F, Lin J Y, Yang S Q, Wang Y S, Lan X Y. Face recognition method of dairy goat based on improved YOLO v5s. Transactions of the CSAM, 2023; 54(4): 331–337. (in Chinese)

[18] Yang Q M, Chen S B, Huang Y G, Xiao D Q, Liu Y F, Zhou J X. Pig counting algorithm based on improved YOLO v5n. Transactions of the CSAM, 2023; 54(1): 251–262. (in Chinese)

[19] Xue H X, Shen M X, Liu L S, Chen J X, Shan W P, Sun Y W. Estrus detection method of parturient sows based on improved YOLO v5s. Transactions of the CSAM, 2023; 54(1): 263–270. (in Chinese)

[20] Ma H X, Zhang M, Dong K B, Wei S H, Zhang R, Wang S X. Research of locust recognition in Ningxia grassland based on improved YOLO v5. Transactions of the CSAM, 2022; 53(11): 270–279. (in Chinese)

[21] Wang Q F, Cheng M, Huang S, Cai Z J, Zhang J L, Yuan H B. A deep learning approach incorporating YOLO v5 and attention mechanisms for field real-time detection of the invasive weed *Solanum rostratum* Dunal seedlings. Computers and Electronics in Agriculture, 2022; 199: 107194.

[22] Li R, Wu Y P. Improved YOLO v5 wheat ear detection algorithm based on attention mechanism. Electronics, 2022; 11(11): 1673.

[23] Liu Z, Mao H, Wu C Y, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022; pp.11966–11976. doi: 10.1109/CVPR.52688.2022.01167.

[24] Wang Q L, Wu B G, Zhu P F, Li P H, Zuo W M, Hu Q H. ECA-Net: Efficient channel attention for deep convolutional neural networks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020; 11531–11539. doi: 10.1109/CVPR42600.2020.01155.

[25] Qin Q, Dong Z Q, Lei X J, Cao M Y, Tang L, Shi M N, et al. Characteristic analysis of resistance to BmNPV in CVDAR strain of silkworm. Acta Microbiologica Sinica, 2019; 59(12): 2390–2400. (in Chinese)

[26] Dong Z Q, Lei X J, Qin Q, Zhang X L, Tang L, Shi M N, et al. Mechanism analysis of Anti-BmNPV resistant strain NC99R. Chinese Journal of Biotechnology, 2020; 36(1): 100–108. (in Chinese)

[27] Tzutalin D. LabelImg. Git code. 2015. Available:https://github.com/tzutalin/labelImg. Accessed on [2021-09-16].

[28] Zhang Z Z, Lan C L, Zeng W J, Jin X, Chen Z. Relation-aware global attention. arXiv e-prints archive, 2022. arXiv: 2201.03545.

[29] Hu J, Shen L, Albanie S, Sun G, Wu E H. Squeeze-and-Excitation Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020; 42(8): 2011–2023.

[30] Loshchilov I, Hutter F. SGDR: Stochastic Gradient Descent with warm restarts. arXiv e-prints archive, 2017. arXiv: 1608.03983.

[31] Rezatofighi H, Tsoi N, Gwak J Y, Sadeghian A, Reid I, Savarese S. Generalized intersection over union: A metric and a loss for bounding box regression. In: P2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach: IEEE, 2019; pp.658–666. doi: 10.1109/CVPR.2019.00075.

[32] Bochkovskiy A, Wang C, Liao H M. YOLOv4: Optimal speed and accuracy of object detection. arXiv e-prints archive, 2020. arXiv: 2004.10934.

[33] Zhou X, Wang D, Krähenbühl P. Objects as Points. arXiv e-prints archive, 2019. arXiv: 1904.07850.

[34] Tan M X, Pang R M, Le Q V. EfficientDet: Scalable and efficient object detection. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern

Recognition (CVPR), Seattle: IEEE, 2020; pp.10778–10787. doi: 10.1109/CVPR42600.2020.01079.

[35] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y, et al. SSD: Single Shot MultiBox Detector. arXiv preprint arXiv: 1512.02325.

[36] Ge Z, Liu S, Wang F, Li Z, Sun J. YOLOX: Exceeding YOLO series in 2021. arXiv e-prints archive, 2021. arXiv 2107.08430.

[37] Cao Y Y, Chen J, Zhang Z C. A sheep dynamic counting scheme based on the fusion between an improved-sparrow-search YOLOv5x-ECA model and few-shot deepsort algorithm. Computers and Electronics in Agriculture, 2023; 206: 107696.

[38] Lu S L, Chen W K, Zhang X, Karkee M. Canopy-attention-YOLOv4-based immature/mature apple fruit detection on dense-foliage tree architectures for early crop load estimation. Computer and Electronics in Agriculture, 2022; 193: 106696.

[39] Xu L J, Wang Y H, Shi X S, Tang Z L, Chen X Y, Wang Y C, et al. Real-time and accurate detection of citrus in complex scenes based on HPL-YOLOv4. Computers and Electronics in Agriculture, 2023; 205: 107590.

[40] Zhang L, Zhou X H, Li B B, Zhang H X, Duan Q L. Automatic shrimp counting method using local images and lightweight YOLOv4. Biosystems Engineering, 2022; 220: 39–54.