Research Article

A Screen Location Method for Treating American *Hyphantria cunea* Larvae Using Convolutional Neural Network

Yan Gao, Ying Zhao [ID], Yujie Ji, Dongjie Zhao, Chong Wang, and Qun Sun [ID]

School of Mechanical and Automotive Engineering, Liaocheng University, Liaocheng, China

Correspondence should be addressed to Ying Zhao; zhaoying@lcu.edu.cn

Chemical control is the major approach to handle the American *Hyphantria cunea* issue; however, it often causes chemical pollution and resource waste. How to precisely apply pesticide to reduce pollution and waste has been a difficult problem. The premise of accurate spraying of chemicals is to accurately determine the location of the spray target. In this paper, an algorithm based on a convolutional neural network (CNN) is proposed to locate the screen of American *Hyphantria cunea*. Specifically, comparing the effect of multicolor space-grouping convolution with that of the same color space-grouping convolution, the better effect of different color space-grouping convolution is first proved. Then, RGB and YIQ are employed to identify American *Hyphantria cunea* screen. Moreover, a noncoincident sliding window method is proposed to divide the image into multiple candidate boxes to reduce the number of convolutions. That is, the probability of American *Hyphantria cunea* is determined by grouping convolution in each candidate box, and two thresholds (E and Q) are set. When the probability is higher than E, the candidate box is regarded as excellent; when the probability is lower than Q, the candidate box is regarded as unqualified; when the probability is in between, the candidate box is regarded as qualified. The unqualified candidate box is eliminated, and the qualified candidate box cannot exit the above steps until the number of extractions of the candidate box reaches the set value or there is no qualified candidate box. Finally, all the excellent candidate boxes are fused to obtain the final recognition result. Experiments show that the recognition rate of this method is higher than 96%, and the processing time of a single picture is less than 150 ms.

1. Introduction

As a world quarantine pest, American *Hyphantria cunea* seriously damages trees, fruit trees, and crops. It was first discovered in Dandong, Liaoning, China, in 1979 and spread rapidly from east to west. Notice no. 3 of 2018 issued by the State Forestry Administration shows that the current epidemic areas of American *Hyphantria cunea* involved 572 county-level administrative regions in 11 provinces (districts, cities). In addition, based on the survey results, adult American *Hyphantria cunea* is detected in a number of nonepidemic areas in 2018, and the diffusion situation is grim. The epidemic may spread in Changzhou, Jiangsu province, and Huanggang, Hubei province. In addition, the risk of transmission to Shanghai and Zhejiang province is existed.

Currently, the control methods of American *Hyphantria cunea* can be divided into physical control, chemical control, and biological control [1–4]. In terms of physical control, herb control and light trapping are widely used [2]. In biological control, natural enemies of *Hyphantria cunea*, biocontrol bacteria, and viruses are mainly used [3]. In chemical control, spraying chemical agents are mainly used [4]. Chemical control has the advantages of high efficiency, convenience, and wide application. However, it is easy to cause chemical pollution and waste of resources. The target application technology based on machine vision can effectively improve the spraying efficiency, reduce the dosage, and avoid chemical pollution.

Target application is one of the research studies focusing on fine pesticides, and many scholars have conducted meaningful studies on it [5–14]. Aiming at the

poor mechanization operation conditions within orchards in China and the ineffective spraying of fruit tree gaps during continuous spraying with traditional orchard sprayers, an automatic target-spraying control system is designed by Xu et al., which has a good practical value for the accurate targeting control as well as diseases and insect pests prevention in sparse orchards [5]. Underwood et al. designed a system that can perform automatic liquid delivery to designated crops using manipulator, fine nozzle, and other equipment, which can effectively reduce production cost and improve production efficiency [6]. Liu et al. designed a spray rod-based precise target-spraying system according to the agronomic requirements of plants with large row spacing and plant spacing, which has good practicability for target application operations with plant spacing above 15 cm in the field [8]. In the current agricultural practice, pesticides are usually applied evenly in the field. However, many insect pests and diseases show uneven spatial distribution. Moreover, the excessive pesticide will lead to pesticide pollution. To reduce the use of agriculture and meet people's demand for healthy food, Oberti et al. used a modular agricultural robot to selectively spray grapevine to study the spraying of target medicine [9]. Berenstein and Edan proposed a precise pesticide-spraying device that can deal with targets of irregular shape and variable size. The device includes a nozzle that can automatically adjust the spray angle, color camera, distance sensor, and other devices. Moreover, the device can spray specific targets on the site and significantly reduce the pesticide usage amount [13]. Based on the color, shape, and distribution of the screen of American *Hyphantria cunea* larvae, a new screen location algorithm is proposed based on the convolutional neural network. This algorithm can help the spraying robot to make rapid and accurate decisions and improve the spraying efficiency.

2. Algorithm Flow of the Screen Location of American *Hyphantria cunea* Larvae

The screen image samples are collected on the site of the American *Hyphantria cunea* disaster area. The screen data set of American *Hyphantria cunea* larvae is first created based on the color, texture, other characteristics, and distribution. Then, a multicolor space-based CNN architecture is proposed, and the screen data set of American *Hyphantria cunea* larvae is employed to train the model.

The screen of American *Hyphantria cunea* larvae is located by the sliding window method and CNN. The sliding window mechanism is mainly divided into the multisize sliding window mechanism and the multiscale sliding window mechanism. Based on the original mechanism of sliding window, nonoverlapping sliding window and non-overlapping regions are proposed in candidate boxes.

The algorithm flow of the screen location of American *Hyphantria cunea* larvae is shown in Figure 1. First, the original image is sharpened in order to improve the image contrast and to enhance the recognition effect. Second, the maximum number of extraction N_m is set, and the initial
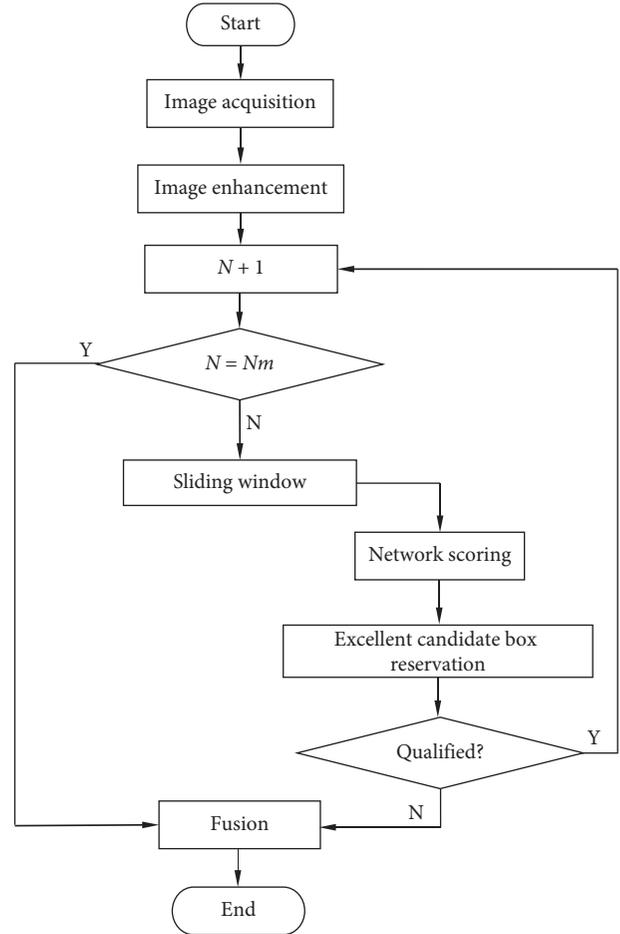

Figure 1: Flow chart of the screen-positioning algorithm of American *Hyphantria cunea* larvae.

value of N is zero. The image is divided into several candidate frames by the sliding window method. The CNN model of RGB and YIQ is used to score each candidate frame and classify it into three levels (excellent, qualified, and unqualified) based on the score. The excellent candidate frame is retained, and the unqualified candidate frame is eliminated. The region in the qualified candidate box is extracted and screened again, and the width and height of the window before each extraction candidate box are reduced to half of the original size. The extraction and screening process are completed when the number of candidate box extraction reaches the set value or there is no qualified candidate box. Finally, all the excellent candidate frames are fused, and the original image is painted with the outline of the screen of American *Hyphantria cunea* larvae.

3. Preparation of Data and Image Sharpening

3.1. Gathering Images.
The reticular screen of American *Hyphantria cunea* larvae is the object, and Canon 600D digital camera is used to take color pictures with a resolution of 960×720. As shown in Figure 2, the reticular screen picture of American *Hyphantria cunea* larvae is collected in the field.

(a)

(b)

(c)

(d)

FIGURE 2: The original image of the netting screen of American *Hyphantria cunea* larvae.

3.2. Sharpening the Image.

To get a better recognition effect, the original image is sharpened to improve the clarity of the image. Here, the Laplace operator $\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$ is employed to sharpen the original image, which is shown in Figure 3. The equation for updating the pixel value of each point of the original image is as follows:

$$
\begin{aligned}
g(x, y) = {} & 5 \cdot f(x, y) - f(x - 1, y) - f(x + 1, y) \\
& - f(x, y - 1) - f(x, y + 1).
\end{aligned}
\tag{1}
$$

3.3. The Data Set.

To train the CNN model, the collected images are employed to create the reticular screen data set of American *Hyphantria cunea* larvae. The training set contains 1318 images, and the test set contains 1318 images. First, the original image is sharpening. Then, the original image is divided into two categories (the reticular screen of American *Hyphantria cunea* larvae and the non-American *Hyphantria cunea* larvae) based on the manually intercepted local images. Two types of image samples are stored in different folders based on the JPG format. Some images in the data set are shown in Figure 4. Figures 4(a)–4(d) belong to the reticular screen of American *Hyphantria cunea* larvae, and Figures 4(e)–4(h) belong to the reticular screen of non-American *Hyphantria cunea* larvae.

4. Convolutional Neural Network (CNN)

4.1. History of CNN in Images.

In 1998, Lecun et al. proposed leNet-5 and used its opponents to write numbers for classification [15]. The network structure of leNet-5 is shown in Figure 5, which is consisted of an input layer, an output layer, three convolution layers, two pooling layers, and a full connection layer. Then, the development of convolutional neural network is at low ebb, and there is no major breakthrough for a long time. In 2012, Krizhevsky et al. used 1.2 million high-resolution images from the ImageNet LSVRC-2010 contest to train a large deep convolutional neural network, which is called AlexNet. On the test data, the recognition effect of AlexNet is obviously better than the previous advanced level [16].

The convolutional neural network later became the focus of research in the field of image, and researchers proposed many improvement methods to enhance network performance [17–21]. In 2013, Zeiler and Fergus proposed ZF Net and introduced a new visualization technique [17]. In 2014, Simonyan and Zisserman proposed VGG Net, which extended the depth of the network to 19 layers. The experimental results showed that the depth has an important impact on the network performance [18]. In the same year, Szegedy et al. proposed Google Net to introduce the inception structure while deepening the network to replace the traditional operation of convolution and activation function [19]. By constant improvement and innovation, the network gets deeper and deeper, the architecture gets more and more complex, and the accuracy gets higher and higher. He and
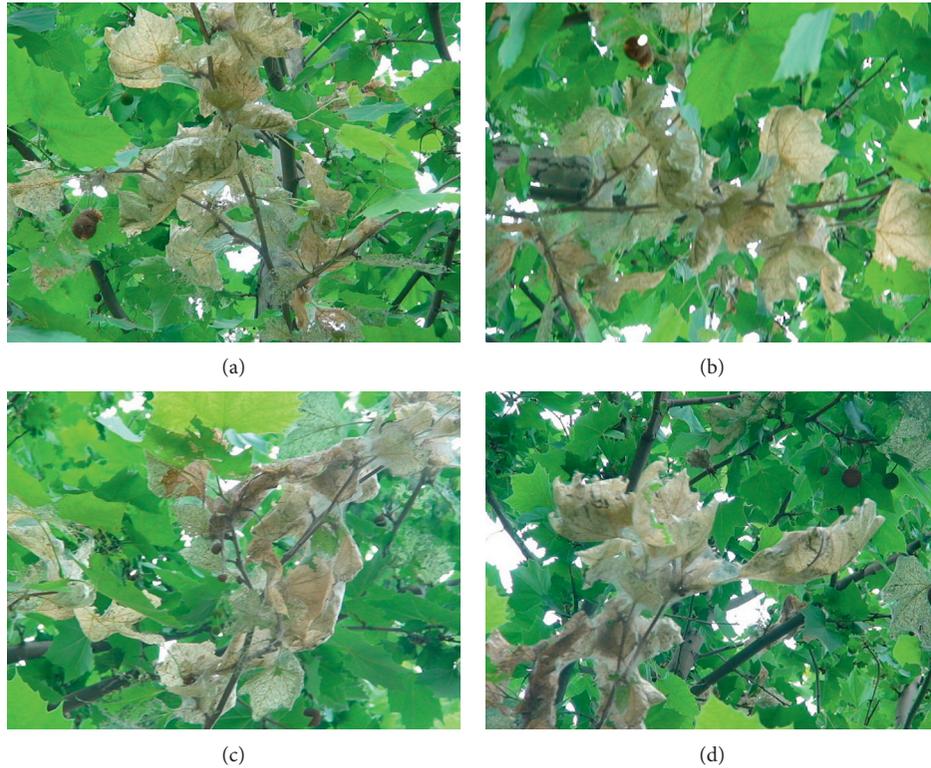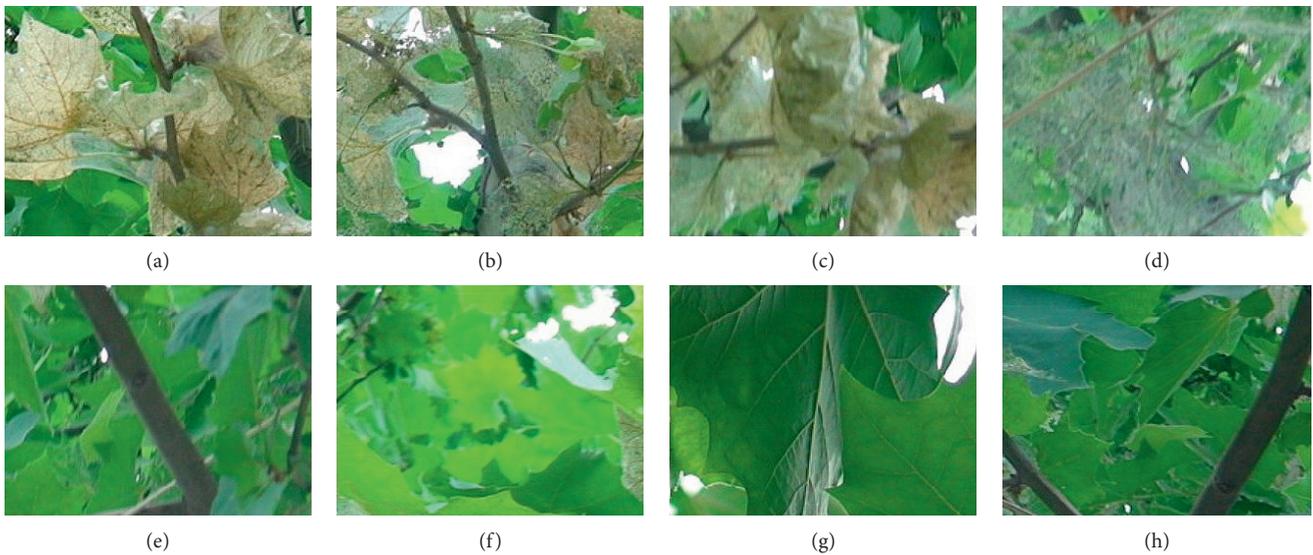
(a)

(b)

(c)

(d)

Figure 3: Image sharpening effect.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Figure 4: Netting data set of American *Hyphantria cunea* larvae.

others proposed ResNet in 2015 [20], and Huang et al. proposed DenseNet in 2016 [21].

In addition to accuracy, speed is also an important performance index of the convolutional neural network. In 2017, Howard et al. proposed an efficient MobileNets model for mobile and embedded visual applications [22]. MobileNets reduces the requirement of the CNN for hardware and plays an important role in promoting the wide application of the CNN.

4.2. CNN Based on Multicolor Space. Packet convolution first appeared in AlexNet [16], which can greatly reduce training parameters. However, the use of packet convolution is limited to the same color space, mostly RGB space. Different color spaces have different advantages, while a variety of color space mix can make up for the lack of a single-color space. This paper proposes a CNN architecture based on different color space-grouping convolutions, as shown in Table 1, taking the RGB image and YIQ image as examples. The input is the
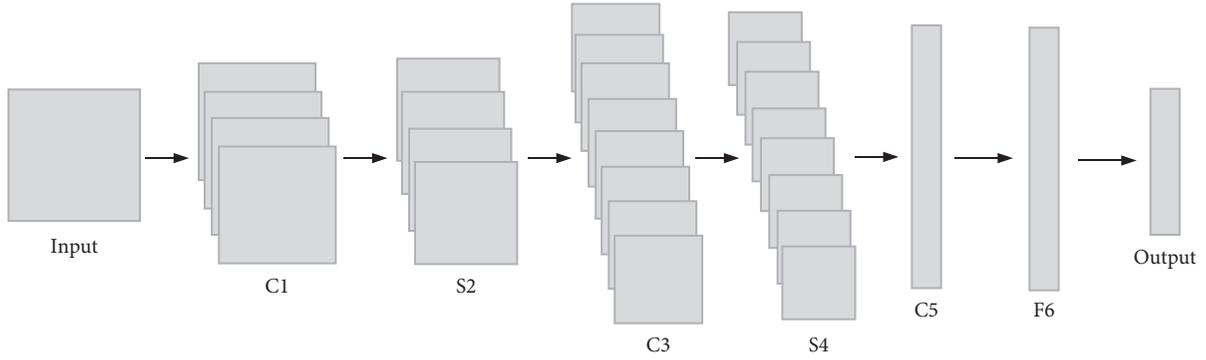
FIGURE 5: LeNet-5 network structure diagram.

TABLE 1: CNN network configuration.

Stage	RGB YIQ	Output quantity
Conv1	3×3 9, stride 1 2×2 max pool, stride 2	$16 \times 16 \times 18$
Conv2	3×3 18, stride 1 2×2 max pool, stride 2	$8 \times 8 \times 36$
Conv3	3×3 36, stride 1 2×2 max pool, stride 2	$4 \times 4 \times 72$
Global average pool		72
Full connection, Softmax		2

32×32 image, and the output is the probability of the existence of the screen of American *Hyphantria cunea* larvae; the detailed process is described as follows:

4.2.1. Input Layers. First, since this network model can only process fixed-size images, the bilinear interpolation algorithm is used to scale the image to meet the input size requirements. Generally, more practical network models need to be trained with a large amount of data, but there are few picture samples of the screen of American *Hyphantria cunea* larvae. To solve this problem, random clipping, flip, saturation, and brightness are performed on the image samples in the training of the model to make up for the insufficient number of image samples and improve the performance and generalization ability of the model.

4.2.2. Conv1 Layer. The Conv1 layer is a convolution layer, where the images are transferred from RGB space to YIQ space before convolution and are convolved, respectively, in RGB space and YIQ space.

In this layer, the convolution kernel of size 3×3 and the zero-complementing strategy are adopted and the step size is set as 1. Then, 9 feature graphs of size 32×32 are obtained by the convolution of a single-color space. The size of the convolution kernel determines the size of a neuron's sensory field. If the convolution kernel is too small, effective local features cannot be extracted; if the convolution kernel is too large, the complexity of feature extraction may far exceed its

representation ability. Therefore, the proper convolution kernel is crucial to improve the performance of the CNN. Based on verification, the application of the 3×3 convolution kernel has the better effect, which is constituted by 9 weight w, corresponding to the pixel value of the image block x. The convolution operation can be described as

$$y = \sum_{i=1}^{9} w_i x_i + b, \tag{2}$$

where y denotes the output of the convolution operation and b denotes the offset item, which is employed to better fit the data. The terminal value of w and b is determined by the network training. The convolution kernel traverses the whole image to get a feature graph. Since there are 9 different convolution kernels and 2 images, 18 feature graphs are finally obtained.

For the 18 feature graphs obtained by convolution, a 2×2 pooling window with a step size of 2 can be used for maximum pooling to reduce the amount of data. The result of pooling should be further calculated with the activation function *ReLU* and then stored as the eigenvalue in the feature graph. The expression of *ReLU* is shown as

$$y = \begin{cases} x, & x > 0, \\ 0, & x \leq 0. \end{cases} \tag{3}$$

18 feature images of size 16×16 are output in this layer; the number of training parameters is as follows: $2 \times (3 \times 3 \times 3 \times 9 + 9) = 504$.

4.2.3. Remaining Convolutional Layer. The setup is similar to the Conv1 layer except the number of convolution kernels. With the number of convolutional layers increasing, the number of convolutional kernels increases exponentially by 2. The extracted features are more complete and accurate, but the processing time will increase correspondingly. The number of convolutional layers is determined by the actual situation, and the convolution in this case will be completed when the accuracy and timeliness are comprehensively considered.

4.2.4. Global Average Pool. The global mean pooling layer is used to replace the fully connected layer which is connected with the last convolutional layer. It is proposed by Network

in Network [23] that the number of neurons can be obtained by averaging each feature graph to the last convolutional layer. Since a total of 72 characteristic images are output from the Conv3 layer, the layer has 72 neurons.

4.2.5. Output Layers. The output layers are the full connection layer and connected the global average pool. In this case, the sample only has two classes (American *Hyphantria cunea* and non-American *Hyphantria cunea*). Therefore, the output layer contains two neurons. The data of the Conv3 layer are accumulated and calculated to obtain two values V_1 and V_2, and the specific calculation process can be described as

$$V_i = \sum_{j=1}^{72} w_{ij} x_j, \quad i = 1, 2, \tag{4}$$

where w denotes weight and x denotes the value of neuron in the output layer. The final value of w is determined by network training. The activation function Softmax is employed to calculate the target probability and nontarget probability, which can be developed as

$$S_i = \frac{e^{V_i}}{1 + e^{-V_i}}, \quad i = 1, 2. \tag{5}$$

The first value is the target probability, which is set as the final output value of the program. When the probability is higher than 99%, it is considered that the image is the screen image of American *Hyphantria cunea*, otherwise it is not.

Based on the above method, the commonly used color spaces such as RGB and YIQ are selected. The effects of single-color space convolution, multicolor space grouping convolution, and single-color space grouping convolution are tested by the screen sample database of American *Hyphantria cunea* larvae. The results are shown in Table 2. Suppose a total of N images are involved in the test, in which N_1 images are identified as screen images, N_2 images are truly screen images in N_1, and N_3 screen images are not identified as screen images. The detection rate = $(N_2/N_1) \times 100\%$ and the omission rate = $(N_3/N) \times 100\%$ are shown in Table 2. The network architecture of single-color space convolution is similar to that of a single group in Table 1, and it can be seen from Table 2 that RGB and YIQ perform well in the convolution of single-color space and better than in other convolution methods of color space. However, the number of parameters is much higher than the grouping convolution. The detection rate and recognition rate of the grouping convolution are similar to those of the convolution with a single-color space in the case of parameter reduction. Therefore, the grouping convolution is selected.

Multicolor space grouping convolution is better than single-color space grouping convolution on the whole. Since the omission rate of single-color space grouping convolution is higher, it is not conducive to prevent and control American *Hyphantria cunea*. Compared to the results of multiple packet convolution, RGB and YIQ packet convolutions have the higher detectable rate and lowest omission rate. After comprehensive consideration of the experimental results, RGB and YIQ grouping convolutions are finally selected for judgment.

5. Image Positioning

5.1. Image Positioning Based on Candidate Frames. At present, the image localization algorithm based on convolution can be divided into two categories. One is the combination of the candidate frame and classifier. The image is first divided into several blocks according to certain criteria, and candidate frames are generated. Then, each candidate frame is selected, and the area is convoluted in each candidate frame [24–27]. The other is the recognition probability and position coordinate value of the object directly generated in the whole image range [28–30]. By comparison, the accuracy of the former is better, and the latter is faster.

In general, the distribution of the net curtain of the American *Hyphantria cunea* larvae is irregular, the environmental interference is relatively greater, and the accuracy of direct recognition in the whole image range is low. Therefore, the combination of the candidate box and convolutional neural network is adopted in this paper. The accuracy of candidate frame extraction affects the accuracy of object location and algorithm speed. Thus, the researchers proposed many candidate frame extraction algorithms [25, 31, 32], among which the sliding window, selective search, and regional suggestion network are the most common. The selective search processing speed is slow, and the regional suggestion network method requires a large amount of prior knowledge. For the actual situation of American *Hyphantria cunea* larvae, the sliding window method is finally selected. That is, in the image coordinate system, the rectangular window moves in accordance with a certain law and intercepts the subimages in the window. The rectangular window is called the sliding window; the size and outer contour of the intercepted subimages are called candidate boxes.

The mechanism for sliding windows is divided into multisize sliding windows and multiscale sliding windows. A multisize sliding window uses a plurality of sliding windows of different sizes to slide over the entire image in equidistant steps to extract candidate frames. On the basis of this method, Li et al. proposed an artificial object detection algorithm based on the sliding window [33]. The principle of a multiscale sliding window is based on an image pyramid, which requires scaling the image at different scales. Then, the fixed-size sliding window will move across the entire image to extract candidate frames. Teutsch and Krugerused used this method to quickly detect moving vehicles [34].

5.2. Inconsistence with the Sliding Window. A noncoincident sliding window is proposed based on the original mechanism of the sliding window. The significant advantage of the noncoincident sliding window is that the process of extracting the candidate frame is combined with the process of screening the candidate frame by the CNN model, which greatly reduces the processing time. The specific process is shown in Figure 6. First, a suitable sliding window size is

TABLE 2: Data comparison of different color spaces.

| Mode	Color space	Detectable rate (%)	Omission rate (%)	Number of parameters
	RGB	99.271	0.607	29,918
	YUV	97.857	0.379	29,918
Convolution of a single-color space	HSV	92.329	2.352	29,918
	XYZ	97.002	0.227	29,918
	YIQ	99.278	0.227	29,918
	RGB and HSV	96.601	0.986	15,338
	RGB and YUV	99.091	0.607	15,338
	RGB and XYZ	99.267	0.835	15,338
	RGB and YIQ	99.091	0.607	15,338
Multicolor space grouping convolution	HSV and YUV	98.844	3.035	15,338
	HSV and XYZ	98.327	1.821	15,338
	HSV and YIQ	98.182	0.986	15,338
	YUV and XYZ	99.091	0.607	15,338
	YUV and YIQ	98.736	0.455	15,338
	XYZ and YIQ	98.188	0.835	15,338
	RGB and RGB	99.264	0.986	15,338
	YUV and YUV	99.450	0.759	15,338
Convolving in a single-color space	HSV and HSV	92.701	3.414	15,338
	XYZ and XYZ	97.846	0.607	15,338
	YIQ and YIQ	99.814	1.138	15,338
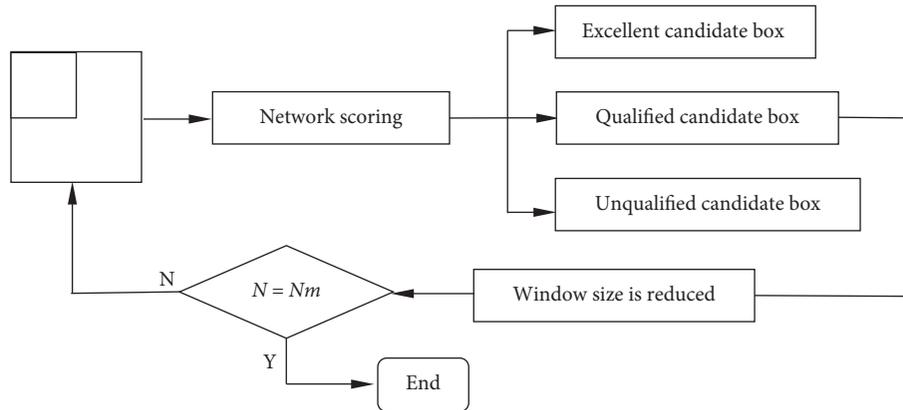
FIGURE 6: Flow chart of noncoincident sliding window.

determined. The window slides over the entire image with its width and height as the step size in the x-axis and y-axis directions to extract the candidate frame, and the area in the candidate frame is noncoincident with each other. Each candidate frame is processed by the method shown in Section 4.2, and the candidate box is output for the presence of the white moth larvae screen. Two thresholds E and Q are set, and the grade is denoted as G. When $G > E$, the candidate box is excellent. Most of the areas are target areas; when $Q < G < E$, the candidate box is qualified, and some of them are target areas; when $G < Q$, the candidate box is unqualified, and the target area is not considered to exist. The excellent candidate box is retained, and the unqualified candidate box is eliminated. Since only a small part of the target area is contained in the qualified candidate box, to accurately locate, it is necessary to narrow the sliding window and then conduct window sliding again to extract the candidate box from the area in the qualified candidate box, repeating the cycle for many times. Once there is no

qualified candidate box or the number of candidate box extraction reaches the set value, the cycle is no longer conducted. Once the width and height of the sliding window are reduced to one-half of the original, the candidate areas generated by the sliding window in the qualified box do not overlap each other. Therefore, one-half is chosen as the reduction proportion of the sliding window.

By continuously looping, the candidate frames become smaller and smaller, and there are more and more excellent candidate frames. All the excellent candidate frames are combined to obtain the target outline frame. The specific comparison between the noncoincident sliding window and the existing sliding window mechanism is shown in Table 3. To better show the advantages of not overlapping sliding window, the width (w) of the sliding window is used as the x-axis direction in the three sliding window mechanisms. The step size is four times of candidate frame extraction with the height (h) of the sliding window as the step size in the y-axis direction. The sliding window of the multisize sliding

TABLE 3: The comparison of different sliding window mechanisms.

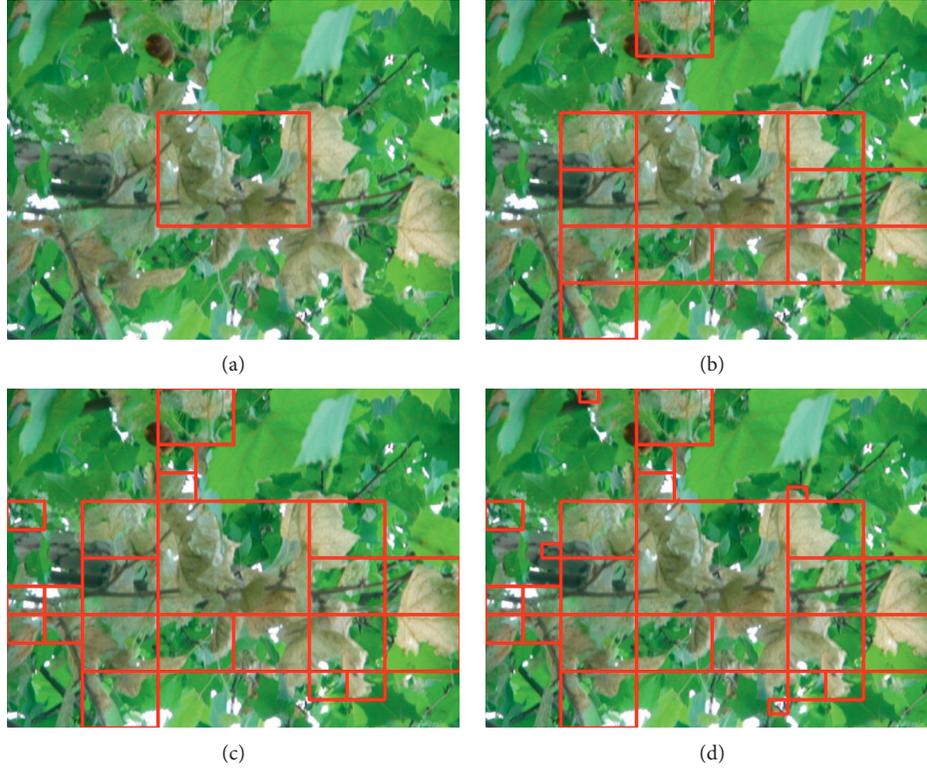| Sliding window mechanisms	Size of sliding window ($w \times h$)	The size of images	Step size	Extraction times of candidate box (n)	Number of candidate boxes
Multisize sliding window		960×720	x: w y: h	4	877
Multiscale sliding window	40×30		x: w y: h	4	877
Noncoincident sliding window		960×720	x: w y: h	4	100~300

(a)

(b)

(c)

(d)

FIGURE 7: The extraction and screening candidate box.

window is reduced by one-half of its original width and height, and the candidate frame is slid in the original image by sliding windows of four sizes. The image of the multiscale sliding window is reduced by one-half of its original width and height, and the candidate frame is slid on the images of 4 different scales with a sliding window of a fixed size. The initial sliding window size of the noncoincident sliding window is set to be a range value, and the number of candidate frames extracted by each image is experimentally verified to be between 100 and 300. From the number of candidate frames, the candidate frame extraction algorithm proposed in this paper greatly reduces the number of candidate frames compared to the other two sliding window mechanisms.

5.3. Screening Results of the Net Curtain of American Hyphantria cunea Larvae. The size of the original image is 960×720. The specific process of the candidate frame extraction and screen is shown in Figure 7. The red rectangular frame is an excellent candidate frame. A total of four candidate frame extraction and screening are performed, and a total of 26 excellent candidate frames are obtained. The size of the first candidate frame extraction and screening window is 320×240, and the sliding window traverses the entire image to obtain 9 candidate frames. As shown in Figure 7(a), the candidate frame with probability higher than 99% is excellent and less than 1% is unqualified. To be qualified, after screening by the CNN model, one excellent candidate frame and 8 qualified candidate frames are obtained. The second candidate frame extraction and screening window size is 160×120, and the sliding window traverses the entire image to obtain 32 candidate frames. After screening, 12 excellent candidate frames, thirteen qualified candidate frames, and seven unqualified candidate frames are obtained from Figure 7(b). The third candidate frame extraction and screening window size is 80×60, and 52 candidate frames are obtained. After screening, nine excellent candidate frames, sixteen qualified candidate frames, and 27 unqualified candidate frames are obtained, as shown in Figure 7(c). The fourth candidate frame extraction and screening window size is 40×30, and 64 candidate frames
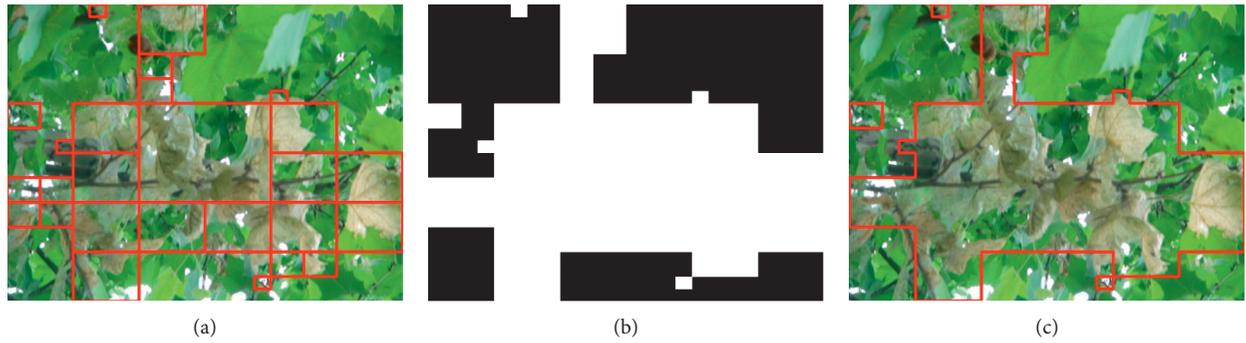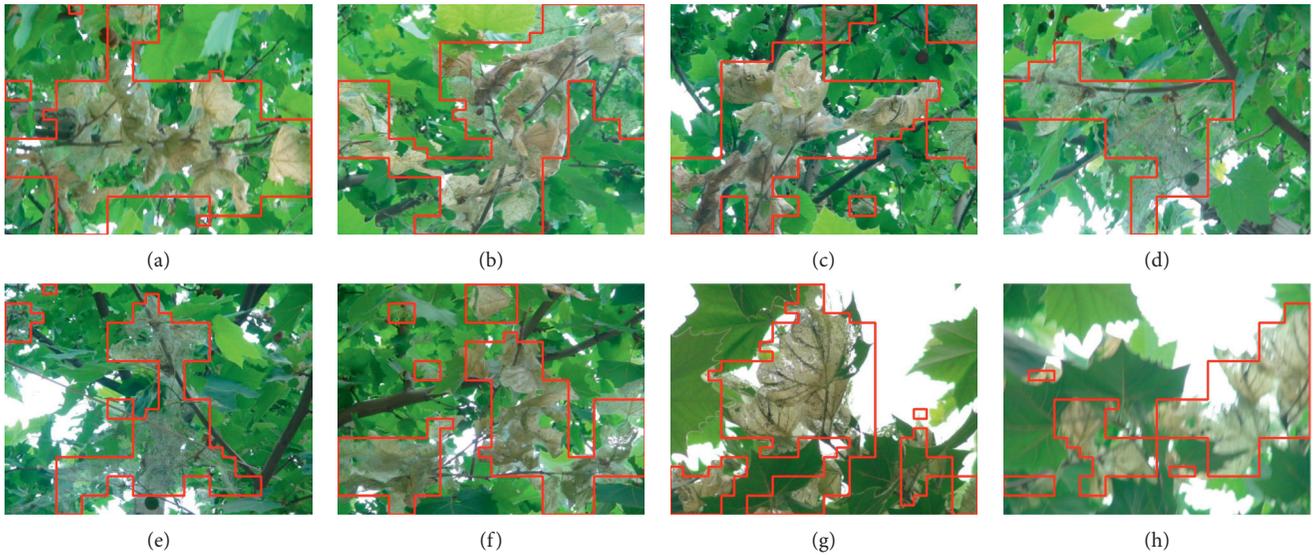
Figure 8: Fusion results.

Figure 9: The resulting graph of algorithm processes.

Table 4: Analysis of single image recognition results.

Image	Recognition rate (%)	Misjudgment rate (%)	Processing time (ms)
a	99.367	3.633	120
b	97.656	2.344	86
c	98.295	4.705	130
d	97.973	2.016	110
e	97.361	1.639	137
f	96.835	2.165	119
g	99.190	8.818	149
h	97.927	5.062	147

are obtained. After screening, four excellent candidate frames are obtained, as shown in Figure 7(d).

After finishing the screen, all the excellent candidates are merged. The specific process is shown in Figure 8. First, a pure black image of the same size as the original image is set, and the size of all the excellent candidate frames in Figure 8(a) is copied on the image and set to white. As shown in Figure 8(b), extraction of the white outline frame and drawing it on the original image are performed as shown in Figure 8(c) to obtain the final processing result.

Using the above process, more images are processed, and the processing results are as shown in Figure 9. It can be seen that the algorithm can obtain ideal results in different processing scenarios. The single picture recognition rate, false positive rate, and processing time are shown in Table 4. The recognition rate refers to the ratio of the identified screen area to the total screen area. The false positive rate refers to the ratio of the screen area defined in the target area except the screen area to the total screen area. It can be seen that the recognition rate is above 96%, and the processing time is less than 150 ms. The false positive rate is slightly

higher when the background light intensity is higher, and the rest are less than 5%.

6. Conclusion

In this paper, a new screen analysis method for American *Hyphantria cunea* larvae is proposed based on the CNN. A CNN architecture is proposed based on multicolor space. Meanwhile, the RGB and YIQ packet convolution methods are selected for judgment. The sliding window is divided to avoid the convolution in the whole image range and improve the processing precision. Based on the image, a new candidate frame extraction algorithm is proposed which is named the noncoincident sliding window method. The image is divided into several candidate frames. The volume convolution of RGB and YIQ space is used in each candidate frame. The product result is output in the form of probability, and two thresholds are set. The result higher than the high threshold is directly considered as excellent, and that lower than the low threshold is removed. The candidate frame in the middle region is again divided by the non-coincident sliding window method. The above process is repeated until the process completed. The number of candidate frame extractions reaches the set value or ends when there is no qualified candidate frame. The final recognition result can be obtained by merging the excellent candidate frames. It is verified that the recognition rate of the method is higher than 96%, and the single image processing time is less than 150 ms.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] L. Luo, X. Wang, Z. Yang, and L. Cao, "Research progress on the control technology of white moth in the USA," *Chinese Journal of Environmental Entomology*, vol. 40, no. 4, pp. 721–735, 2018.

[2] Y. Zhang, Z. Xue, Z. Wang et al., "Control experiment of American *Hyphantria cunea*," *Jiangsu Forestry Science and Technology*, vol. 41, no. 6, pp. 24–27, 2014.

[3] Y. Bi, Q. Qin, Z. Wang et al., "An auto-contamination trap-strips system for biological control of *Hyphantria cunea* (lepidoptera: noctuidae), an invasive pest in China," *International Journal of Pest Management*, vol. 64, no. 3, pp. 230–235, 2018.

[4] T. Gokturk, S. Kordali, and A. U. Bozhuyuk, "Insecticidal effect of essential oils against fall webworm (*Hypantria cunea* drury (lepidoptera: arctiidae))," *Natural Product Communications*, vol. 12, no. 10, 2017.

[5] L. Xu, H. Zhang, H. Zhang et al., "Development and test of automatic target spray control system 4 for orchard sprayer," *Journal of Agricultural Engineering*, vol. 30, no. 22, pp. 1–9, 2014.

[6] J. P. Underwood, M. Calleija, Z. Taylor et al., "Real-time target detection and steerable spray for vegetable crops," in *Proceedings of the International Conference on Robotics and Automation: Robotics in Agriculture Workshop*, pp. 26–30, Seattle, WA, USA, May 2015.

[7] Y. Song, H. Sun, M. Li, and Q. Zhang, "Technology application of smart spray in agriculture: a review," *Intelligent Automation & Soft Computing*, vol. 21, no. 3, pp. 319–333, 2015.

[8] X. Liu, Y. Li, M. Li et al., "Design and test of precise target application system for rod sprayer," *Journal of Agricultural Machinery*, vol. 47, no. 3, pp. 37–44, 2016.

[9] R. Oberti, M. Marchi, P. Tirelli et al., "Selective spraying of grapevines for disease control using a modular agricultural robot," *Biosystems Engineering*, vol. 146, pp. 203–215, 2016.

[10] R. Berenstein, *A Human-Robot Cooperative Vineyard Selective Sprayer*, Ben-Gurion University of the Negev, Beersheba, Israel, 2016.

[11] M. Laursen, R. Jørgensen, H. Midtiby et al., "Dicotyledon weed quantification algorithm for selective herbicide application in maize crops," *Sensors*, vol. 16, no. 11, p. 1848, 2016.

[12] B. S. Faiçal, H. Freitas, P. H. Gomes et al., "An adaptive approach for UAV-based pesticide spraying in dynamic environments," *Computers and Electronics in Agriculture*, vol. 138, pp. 210–223, 2017.

[13] R. Berenstein and Y. Edan, "Automatic adjustable spraying device for site-specific agricultural application," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 2, pp. 641–650, 2018.

[14] C. Zhai, C. Zhao, X. Wang et al., "Research progress on precise control method of orchard air spray," *Journal of Agricultural Engineering*, vol. 34, no. 10, pp. 1–15, 2018.

[15] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 1097–1105, Lake Tahoe, NV, USA, December 2012.

[17] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proceedings of the 2014 European Conference on Computer Vision*, pp. 818–833, Springer, Zurich, Switzerland, September 2014.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556.

[19] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, IEEE, Boston, MA, USA, June 2015.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[21] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings*

of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708, Honolulu, HI, USA, July 2017.

[22] A. G. Howard, M. Zhu, B. Chen et al., "Mobilenets: efficient convolutional neural networks for mobile vision applications," 2017, https://arxiv.org/abs/1704.04861.

[23] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, https://arxiv.org/abs/1312.4400.

[24] R. Girshick, "Fast R-CNN," in *Proceedings of the 2015 IEEE International Conference on Computer Vision*, pp. 1440–1448, IEEE, Santiago, Chile, December 2015.

[25] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2015.

[26] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proceedings of the 2017 IEEE international Conference on Computer Vision*, pp. 2961–2969, Venice, Italy, October 2017.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *Proceedings of the 2014 European Conference on Computer Vision*, pp. 346–361, Springer, Zurich, Switzerland, September 2014.

[28] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, IEEE, Las Vegas, NV, USA, June 2016.

[29] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot multibox detector," in *Proceedings of the European Conference on Computer Vision*, pp. 21–37, Springer, Amsterdam, Netherlands, October 2016.

[30] C. Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "DSSD: deconvolutional single shot detector," 2017, https://arxiv.org/abs/1701.06659.

[31] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154–171, 2013.

[32] C. L. Zitnick and P. Dollár, "Edge boxes: locating object proposals from edges," in *Proceedings of the 2014 European Conference on Computer Vision*, pp. 391–405, Springer, Zurich, Switzerland, September 2014.

[33] L. Li, Y. Liu, Z. Wang et al., "Detection algorithm of artificial targets in remote sensing images based on sliding window," *Journal of Xiamen University: Natural Science Edition*, vol. 53, no. 6, pp. 792–796, 2014.

[34] M. Teutsch and W. Kruger, "Robust and fast detection of moving vehicles in aerial videos using sliding windows," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 26–34, Boston, MA, USA, June 2015.