ARTICLE

Quick and Accurate Counting of Rapeseed Seedling with Improved YOLOv5s and Deep-Sort Method

Chen Su, Jie Hong, Jiang Wang and Yang Yang[*]

College of Engineering, Huazhong Agricultural University, Wuhan, 430070, China

[*]Corresponding Author: Yang Yang. Email: yangyang@mail.hzau.edu.cn

ABSTRACT

The statistics of the number of rapeseed seedlings are very important for breeders and planters to conduct seed quality testing, field crop management and yield estimation. Calculating the number of seedlings is inefficient and cumbersome in the traditional method. In this study, a method was proposed for efficient detection and calculation of rapeseed seedling number based on improved you only look once version 5 (YOLOv5) to identify objects and deep-sort to perform object tracking for rapeseed seedling video. Coordinated attention (CA) mechanism was added to the trunk of the improved YOLOv5s, which made the model more effective in identifying shaded, dense and small rapeseed seedlings. Also, the use of the GSConv module replaced the standard convolution at the neck, reduced model parameters and enabled it better able to be equipped for mobile devices. The accuracy and recall rate of using improved YOLOv5s on the test set by 1.9% and 3.7% compared to 96.2% and 93.7% of YOLOv5s, respectively. The experimental results showed that the average error of monitoring the number of seedlings by unmanned aerial vehicles (UAV) video of rapeseed seedlings based on improved YOLOv5s combined with depth-sort method was 4.3%. The presented approach can realize rapid statistics of the number of rapeseed seedlings in the field based on UAV remote sensing, provide a reference for variety selection and precise management of rapeseed.

KEYWORDS

Rapeseed seedling; UAV; improved YOLOv5s; attention mechanism; real-time detection

1 Introduction

Rapeseed is the 2nd major source of vegetable oil in the world [1], has high oil content in its grains and appropriate fatty acid composition (65% oleic acid, 21.5% linoleic acid, and 8% linolenic acid), which can provide the body with the required oil [2]. Rapeseed is widely planted in China, so improving its yield has been the focus of research, where selection of high-yielding varieties and refined crop management are two important ways to improve the oil yield of rapeseed seeds [3]. For breeders and planters, the statistics of the number of crop seedlings are important features for testing seed quality, crop cultivation and field crop management in the field [4]. For rapeseed seedlings, the number of seedlings per unit area will also affect their subsequent growth, development and final yield [5]. Therefore, the estimation of the number of seedlings of rapeseed is significance for field rapeseed cultivation.

Previously, the population density of seedlings was often estimated by partitioning sampling, which was time-consuming and could not guarantee accuracy [6]. With the development of technology in recent year, unmanned aerial vehicles have many advantages in the field of agricultural phenotypic monitoring, such as flexible operation, low-altitude flight, high-definition image shooting [7]. Many researchers have used computer vision methods to count crops from photos taken by UAV. Zhou et al. [8] converted the red-green-blue (RGB) images to grayscale images and performed skeleton extraction of maize seedlings after separating them from the soil background using Otsu threshold segmentation to recognize maize number. The identification results of this method relate strongly to manually collected data (R = 0.77–086). Bai et al. [9] used peak detection algorithm to quickly count the number of equally spaced and overlapping crop seedlings. For the maize dataset, the proposed method obtained R^2 of 0.76 and relative root mean square error (RRMSE) of 4.44%. For the sunflower dataset, the method obtained R^2 of 0.89 and RRMSE of 4.29%. Liu et al. [10] used UAV to quickly and automatically count corn seedlings. The skeleton information of corn was extracted on the basis of the threshold segmentation method, and the corner detection model was used to identify the quantity of corn seedlings. The corn seedlings recognition rate reached 99.78%. Zhao et al. [11] developed practical remote sensing methods to assess mechanically seeded rape seedlings. The Otsu threshold method was used to segment rape plants and extract shape characteristics such as area, ellipse fitting and aspect ratio to build a regression model of seedling numbers. The obtained R^2 values were 0.845 and 0.867 for the two growth stages, demonstrating the feasibility of the method. From the above literatures, it can be seen that threshold segmentation is usually used to obtain binary images from the images acquired by UAV. Based on this, skeleton extraction, corner point detection and multiple linear fitting approach are used to identify the number of crops. However, the threshold segmentation method is susceptible to plant shading, light intensity, weeds and soil color, which make the final recognition effect not generalizable in other complex environments.

Deep learning as a modern method with enhanced learning capability means higher accuracy and has been widely used in farming and agriculture [12]. In the detection and counting of crop objects in complex environments, the two-stage object detection network Faster R-CNN [13] with high precision and the one-stage object detection network single shot multi-box detector (SSD) [14] and YOLO [15,16] series with fast processing speed have been proved to be advanced and effective. Faster R-CNN was used to achieve automatic detection of green tomatoes under fruit shading and variable light conditions with an average precision (AP) of 87.8% on the test dataset and reached high counting accuracy of 0.87 (R^2 = 0.87) [17]. A system for rapid estimation of maize seedling emergence, YOLOv3 and TOPSIS method were used to build the best model for identifying seedling location, spacing and size, and the system had good predictive performance for maize seedling number with 92% accuracy [18]. YOLOv4 was modified by improving the activation function, replacing non-maximal suppression and introducing attention mechanism to realize the identification of weeds and maize when UAV at higher flight altitudes and to lower training model labeling cost, with a mean average accuracy (map) of 86.9% [19]. Zhao et al. [20] improved YOLOv5 method by adding detection layer and adapting the confidence loss function to identify wheat spikes with small-size, high density and overlap in UAV images. The AP of wheat spike detection amounts to 94.1%, which was higher than the standard YOLOv5 by 10.8%. Although the two-stage and one-stage object detection methods and their improvements are accurate and effective ways for crop counting, it cannot effectively identify crops in single image of large area and small objects. It may also repeat counting for consecutive images with overlap.

The count of the number of small targets in a large area can be achieved by shooting video using UVA. The number of seedlings in the whole large field is counted by object detection paired with object tracking of consecutive video frames [21]. Li et al. [22] used an enhanced YOLOv5 model with squeeze and excitation network in combination with Kalman filter algorithm to achieve reliable tea bud counting in the field. The application of the model to tea bud counting experiments showed that the counting results of the test videos

were highly correlated with the manual counting results. Tan et al. [23] raised a cotton seedling tracking method and combined it with YOLOv4 to improve tracking speed and counting accuracy. The AP of the detection model was 99.1%. The multi-objective tracking accuracy (MOTA) and ID switching of the present tracking method were 72.8% and 0.1%, respectively. All test video count results relative error was 3.13%. Lin et al. [24] adopted improved YOLOv5, deep-sort and OpenCV programs to achieve efficient peanut video counting. The detection model had an F1 score of 0.92, the counting accuracy was 98.08%, and only one-fifth of the time for manual counting.

So far, most researches have focused on the detection of grain and fruit counts. Overlapping and smaller rapeseed seedlings are difficult to identify, and there is a certain gap in the study of rapeseed seedling counting using video. In this paper, a real-time rapeseed seedling UAV video counting model with improved YOLOv5s, deep-sort and OpenCV is proposed to obtain the number of rape seedlings efficiently. The subsequent portion of this article is as follows: Section 2 describes the material, the improved YOLOv5s model and the implementation process of rapeseed seedlings counting. Section 3 analyzes the effect of model improvement and the final rapeseed seedlings counting result. Section 4 concludes this study.

2 Material and Methods

2.1 Experiment and Image Acquisition

Experimental data were collected from Huazhong Agricultural University (Wuhan, Hubei Province, China; 30.5N, 114.5E). The experimental site was located in a north subtropical monsoonal atmosphere with an annual mean temperature of 16.5°C and 1950 h of sunshine. The cultivar planted was Chinese mixed rape, which was grown by strip sowing and manually seeded on September 30, 2022. The experimental site is shown as Fig. 1 with close-up and aerial images.

In order to collect information on rapeseed seedlings from various viewpoints, the DJI "Mavic 2" UAV was used to obtain rapeseed photos with a pixel of 5472 × 3648 at two heights of 3 m and 6 m in vertical, respectively. A total of 179 images containing rapeseed were collected, including 124 images taken at 3 m height and 55 images taken at 6 m height. These images were taken on October 20–22, 2022, under clear weather conditions between 10:00 and 2:00 local time. Most of the rapeseed seedlings had reached the 3-leaf stage under the time period of this shot.

2.2 Image Preprocessing

It is common for deep learning networks to compress the original image size, reduce the pixel count and increase the training speed. Excessive image resolution when compressed leads to loss of detail in relatively small objects, making it difficult for deep learning networks to discriminate the correct object [25]. For example, a rapeseed seedling occupies about 0.2% of the entire image pixels when the UAV is photographed at 6 m. Therefore, if the image is compressed when training a deep learning network, some detailed features of rapeseed will be lost, resulting in poor network performance.

For the YOLOv5 network, the image pixels are compressed to 640 × 640 at default. Hence, the original image was sliced into multiple sub-images with a size of 1280 × 1280 resolution to guarantee that YOLOv5 network learns the feature information of rapeseed seedlings in the picture. When slicing the images, a certain overlap area between sub-images is needed to ensure the integrity of each rape seedling in the image to avoid wasting data. For the captured rapeseed images, this paper designs an image cutting method to ensure that the sub-images cover all the rapeseed seedlings of the original image. The process of creating the dataset is shown in Fig. 2.
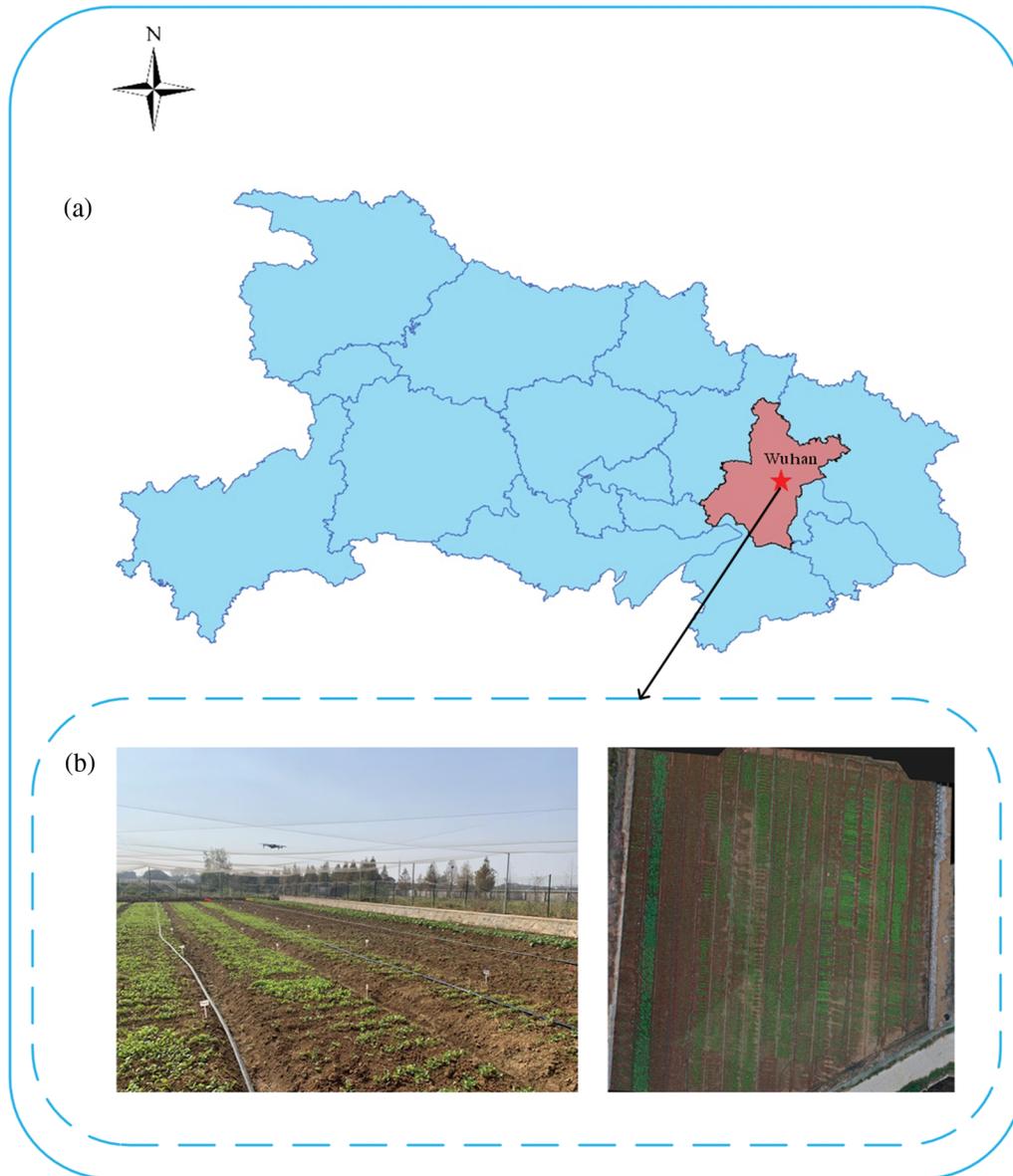
Figure 1: Overview of the rapeseed experimental site. (a) The star represents rapeseed growing locations (Wuhan City, Hubei Province; 30.5N, 114.5E). (b) Experimental scenes (strip sowing) and aerial stitching images

The formula for determining the overlap of sub-image is shown below:

$$
\begin{cases}
overlap\ rate_h = \dfrac{\max(\text{rectangle}_h)}{1280} \\[2mm]
overlap\ rate_w = \dfrac{\max(\text{rectangle}_w)}{1280} \\[2mm]
min(overlap\ rate_h, overlap\ rate_w)
\end{cases}
\tag{1}
$$
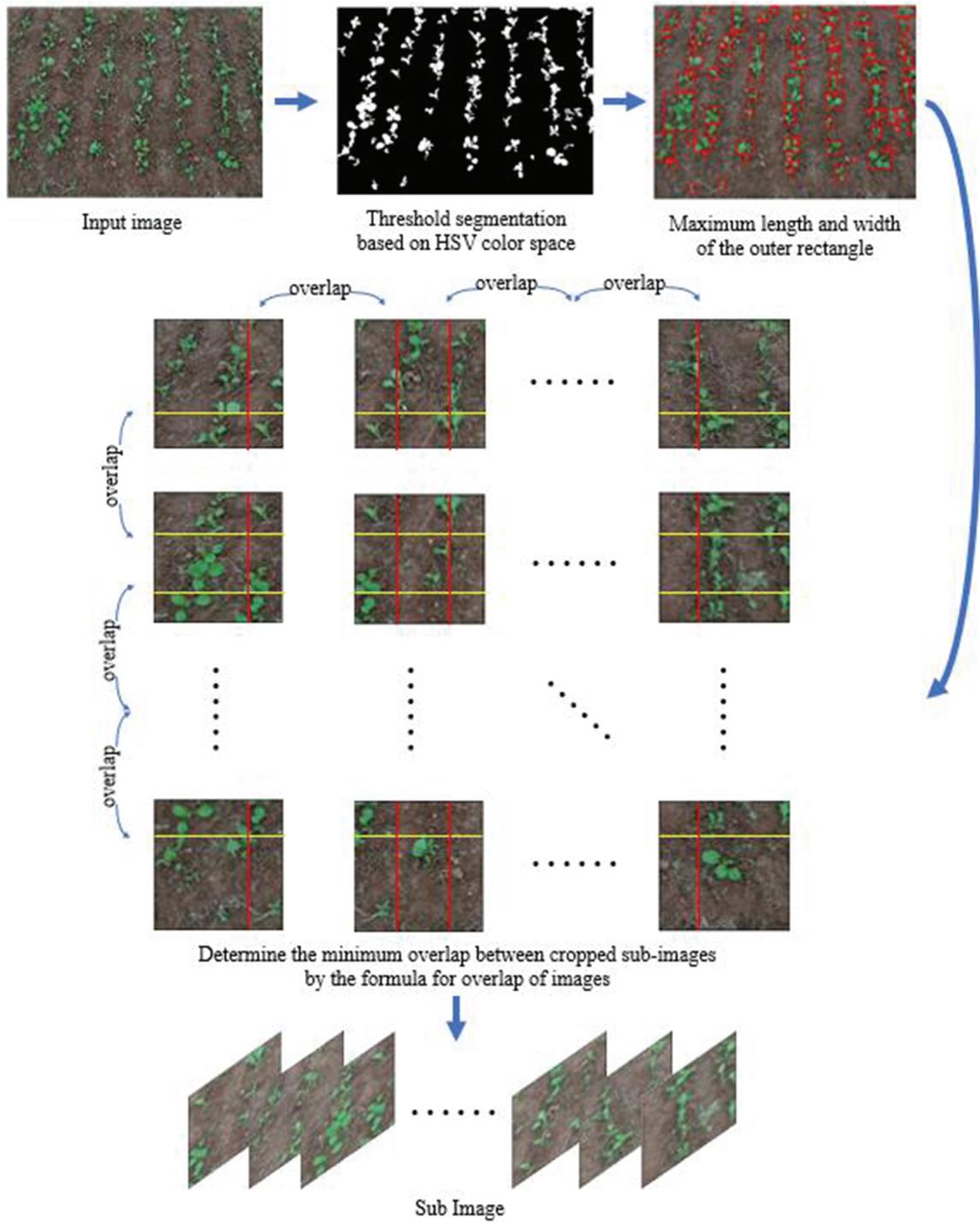
Figure 2: The creation steps of the image dataset

rectangle$_h$ and rectangle$_w$ indicate the length and width of the external rectangle in pixels. 1280 represents the resolution size of the original image length and width. *overlap rate$_h$* and *overlap rate$_w$* are the horizontal and vertical overlap between sub-images, respectively.

Based on the computed overlap region size, 179 original images were collected to generate sub-images. The generated sub-images were examined to obtain a sum of 1010 images, which 864 sub-images were taken at 3 m height and 146 sub-images were taken at 6 m height. The 1010 sub-images were annotated using the open source Labeling software in the form of rectangular boxes. To ensure the robustness of model training, 1010 rapeseed sub-images were enhanced with scaling, color change and distortion on the samples, and a total of 2000 images of rapeseed were obtained as the dataset. The dataset was partitioned into a training set (1600 images) and a test set (400 images). The training set includes 37,178 rapeseed seedlings, and the test set includes 8,532 rapeseed seedlings.

2.3 Process of Rapeseed Seedling Counting Realization

The general route of rape seedling counting in this paper is shown in Fig. 3. In total, there are four steps, firstly, the UAV acquires images and crops. The next step is to label the rape seedlings using Labeling software and generate the labeled text. In the third step, the images and the labeled text are used as data sets, and the improved YOLOv5s and deep-sort are trained to obtain the object detection weights and tracking weights, respectively. Finally, the UAV video is imported into deep-sort and the rape seedlings are counted with the help of OpenCV module.

2.4 Improved YOLOv5s

2.4.1 YOLOv5

YOLOv5 is a one-stage object detection algorithm. It usually consists of four parts, input terminal, backbone, neck and detection head. CSPDarknet is used as the backbone to efficiently extract feature information. The Panet+FPN structure is used to enable effective fusion of features to enhance the detection performance of objects with different sizes. The detection heads are used for object recognition and classification prediction. YOLOv5's network structure and component modules are shown in Fig. 4.

2.4.2 Improved YOLOv5s

In this study, the open source code YOLOv5s 5.0 was used to detect rapeseed seedlings. This version has speediest inference and minimum model version among the four official versions YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. YOLOv5s can identify smaller and with a certain degree of overlap rape seedlings. In this paper, the coordinated attention (CA) [26] mechanism was inserted to the backbone network and the standard convolution (SC) was replaced with GSConv [27] at the neck. CA attention mechanism considers image direction and position information as well as channel information with a simple structure. Its module structure is illustrated in Fig. 5.

Two feature maps $C * H * 1$ and $C * 1 * W$ are obtained by 1-dimensional pooling the feature maps of $C * H * W$ in the x and y directions. The generated feature maps in both directions are encoded to obtain two attention maps. Both attention maps capture long-range dependencies in one spatial direction of the original feature map. Finally, the result is obtained by multiplying the two attention maps and the original feature map.

GSConv module has low complexity and low computational cost. GSConv computational cost is about 60% of the SC, and also reduces the negative impact of low extraction ability of depth-wise separable convolution. Its structure is illustrated in Fig. 6.
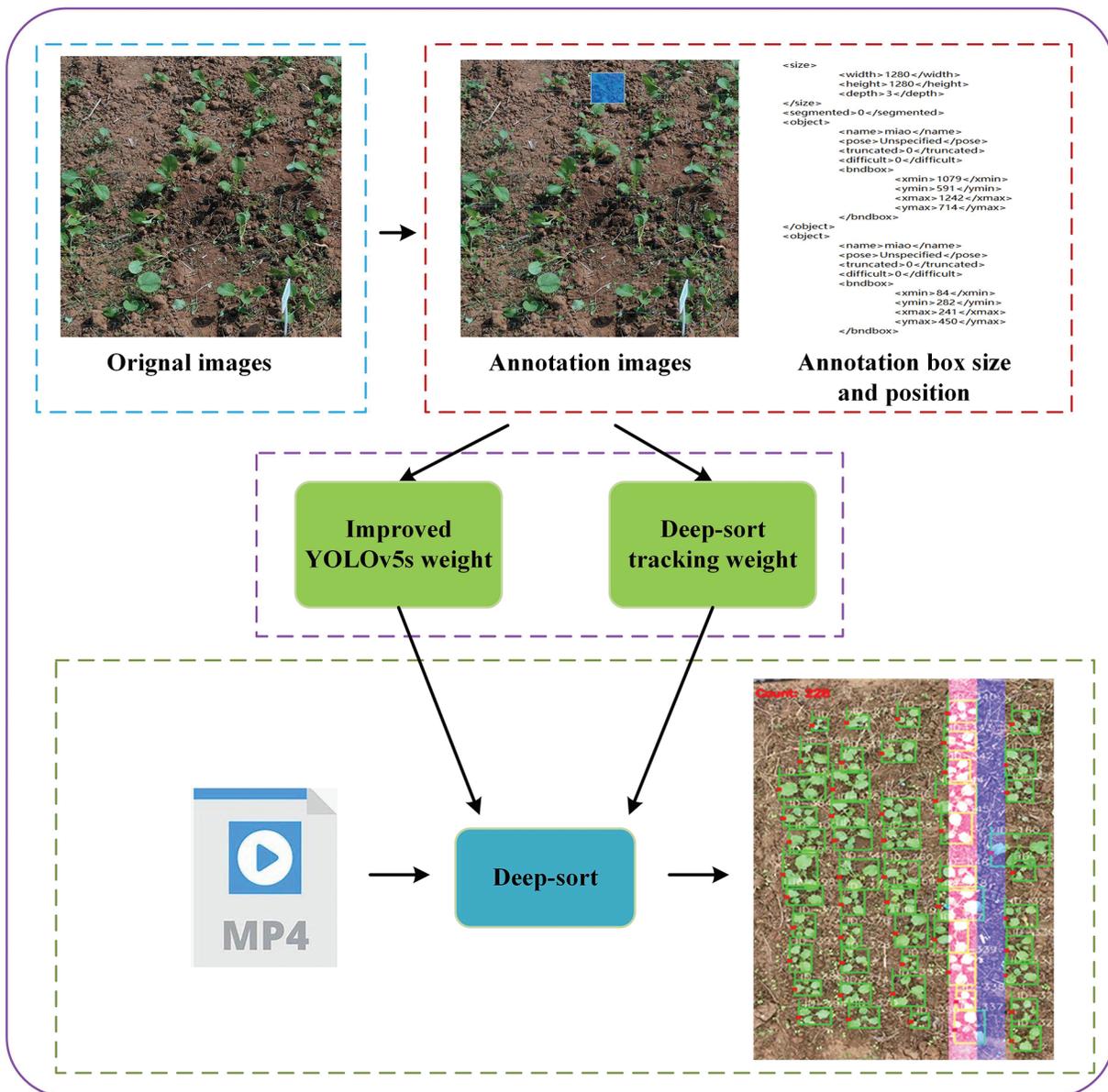
Figure 3: Diagram of the overall implementation process. Please increase the size of the very small letter; they are unreadable

If all stages of the model use GSConv, the network layer of the model will be deeper, which will aggravate the resistance of data flow and significantly increase the inference time. Therefore, a better choice is to use GSConv in the neck, because the feature layer has become slender (the channel dimension reaches its maximum and the width-height dimension reaches its minimum) and there is less redundant repetitive information.

The addition of both modules can further improve the feature extraction capability of the backbone network and guarantee the inference time. The improved YOLOv5s network configuration is displayed in Fig. 7.
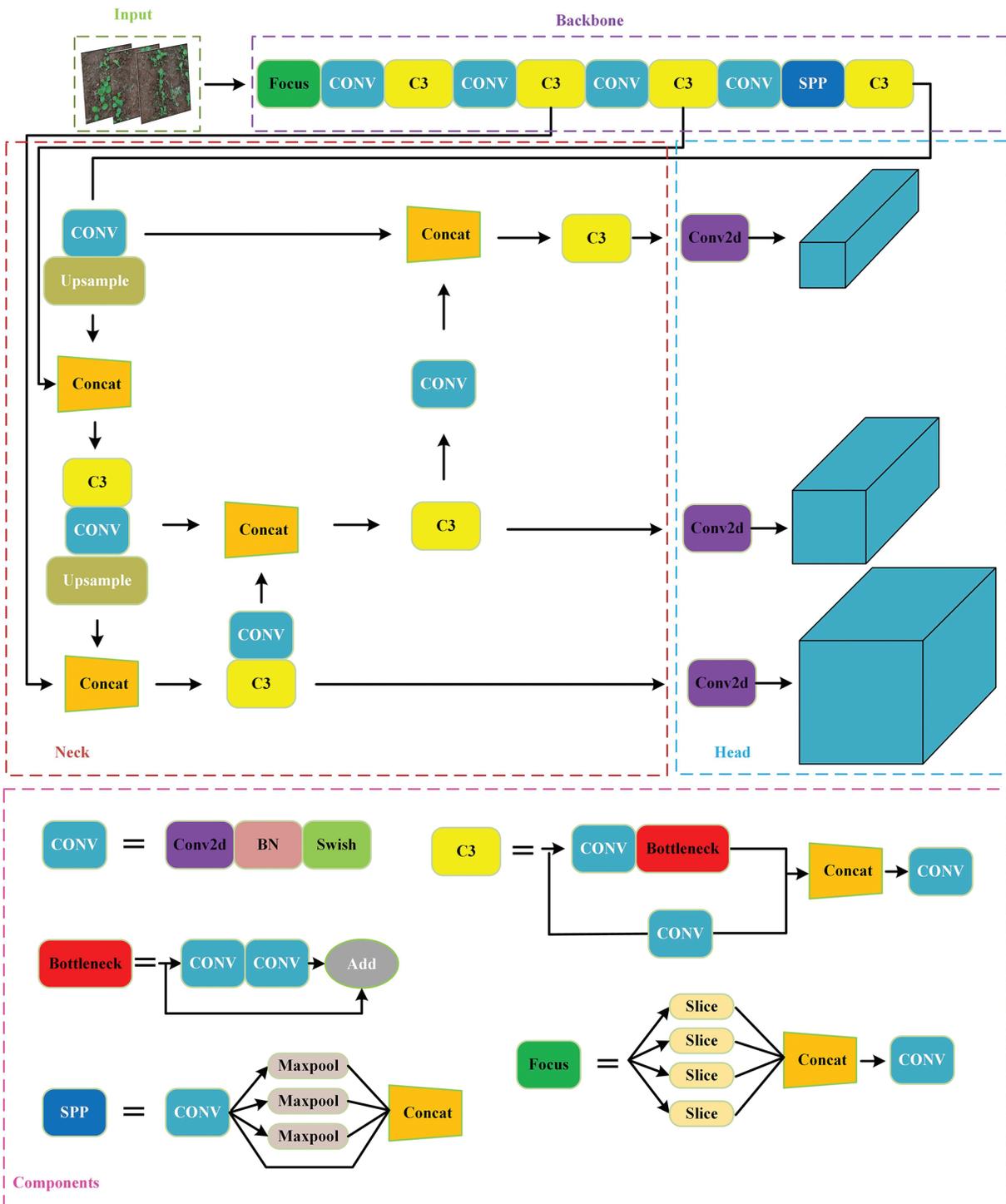
Figure 4: The YOLOv5 network structure diagram with four common blocks, input, backbone, neck and prediction heads, and five basic components CONV, Bottleneck, SPP, C3, Focus
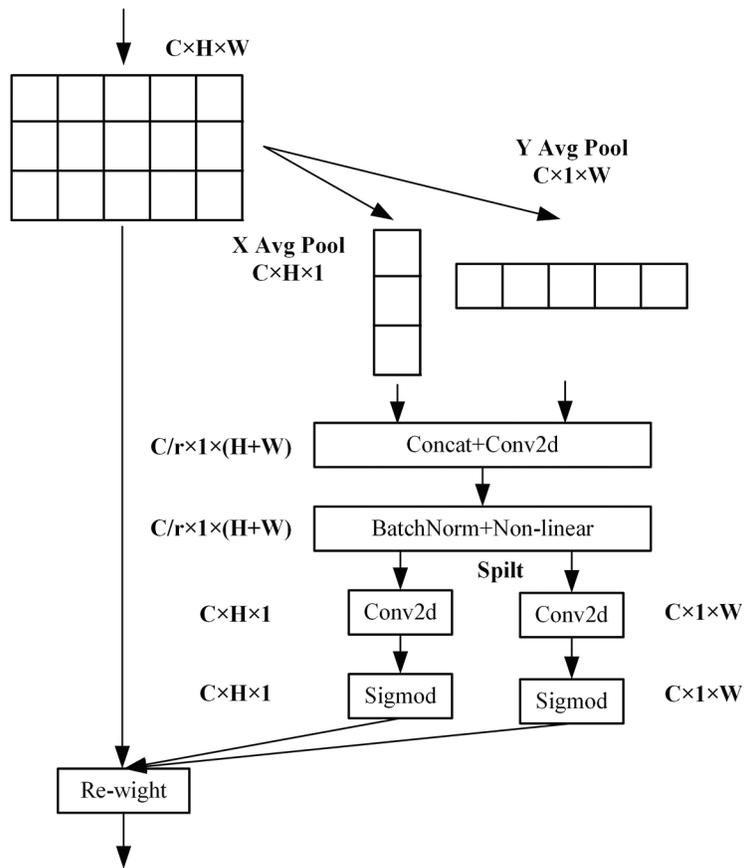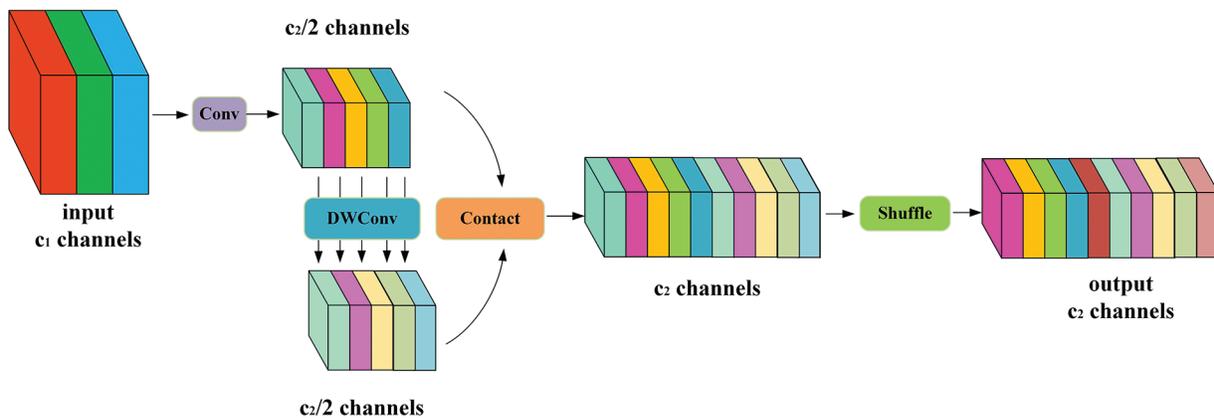
Figure 5: Schematic of the CA module

Figure 6: The calculation process of the GSConv

2.5 Deep-Sort

Deep-sort [28] is conducted on the sort object tracking improvement. It introduces a deep learning model to extract the appearance features of the object, so that the nearest neighbor matching can be carried out in the real-time object tracking process, thus reducing the number of ID transformations. Its workflow is shown in Fig. 8.
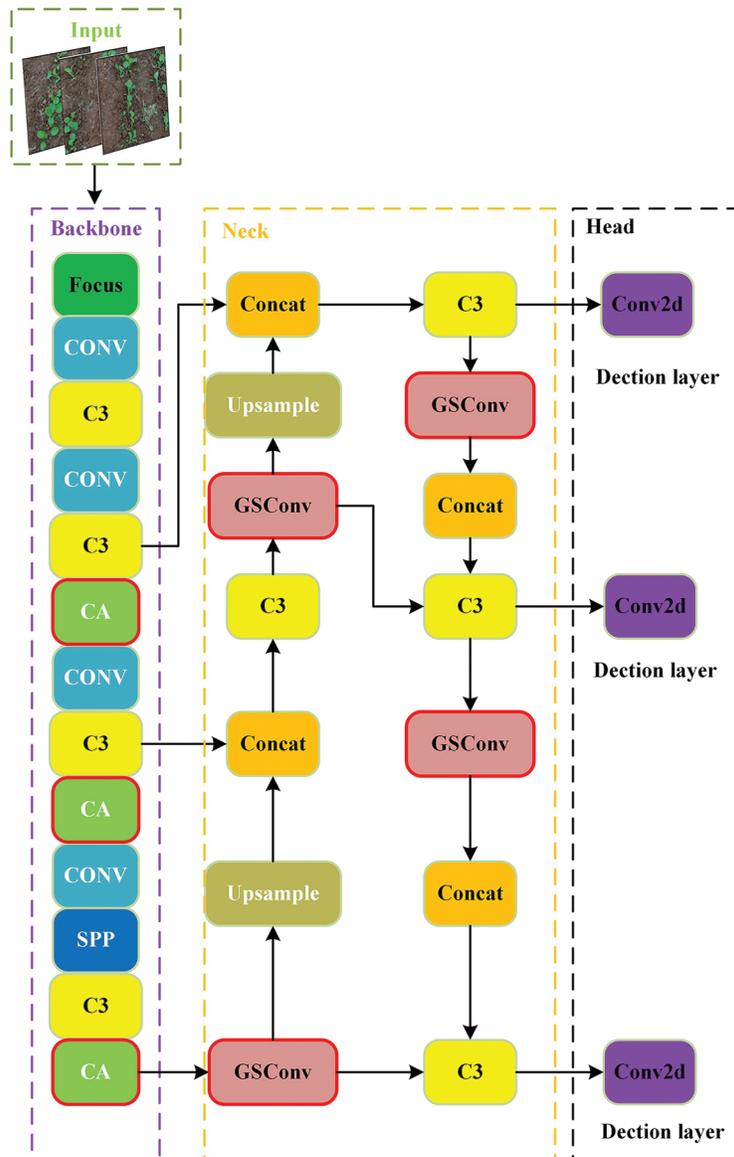
Figure 7: The configuration of improved YOLOv5s, with the improved part marked by a red bolded line

The overall process is as follows:

(1) Predicted state: the tracks generated in the previous iteration are predicted by Kalman Filter (KF), and the data states are Confirmed and Unconfirmed.

(2) First matching: the tracks from step 1 and the detections detected in this round of YOLOv5s are sent to Cascade for matching, resulting in three state results unmatched tracks, unmatched detections, and matched tracks.

(3) Second Matching: the detections in step 2 will be missed, so they are combined with the undetected tracks in step 1 and matched with intersection over union (IOU) in order to produce the new three state results unmatched tracks, unmatched detections, matched tracks.

(4) Failure processing object: unconfirmed and confirmed of the unmatched tracks but the state of the age exceeds the threshold are set to delete.

(5) Output the results and prepare data for the next round: merge the tracks from three sources. Merge the matching Tracks from step 3 and step 4. The unmatched detections in step 3 create new Tracks. The confirmed tracks in step 4 that are not over age. The three are combined as the output of this round and also as the input of the next iteration, continue with step 1.
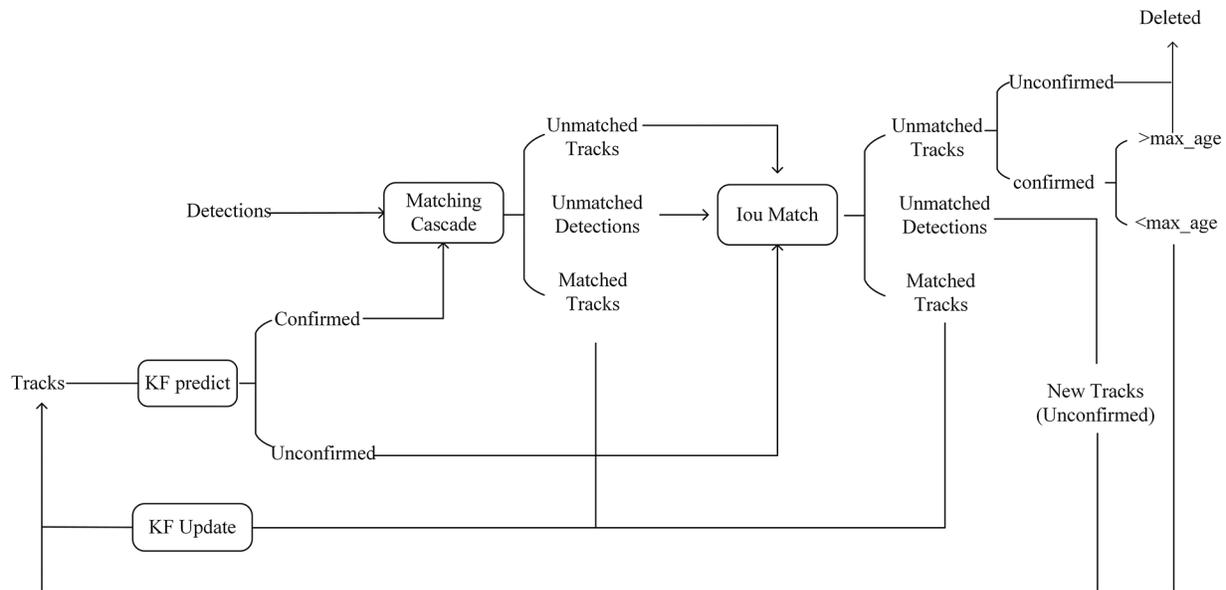

Figure 8: Deep-sort flow chart

2.6 Evaluation Indicators

In this paper, model performance is evaluated by accuracy (P), recall rate (R), average accuracy (AP), F1 score (F1), frames per second (FPS), and model parameter size (Mb). FPS representative can handle images per second.

Precision represents the proportion of the number of predicted results that are correct to the total predicted results.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

True positive (TP) expresses the number of rape seedlings correctly identified, false positive (FP) denotes other samples incorrectly identified as rape seedlings. Recall is the proportion of the number of correctly predicted to the actual positive samples.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

False negative (FN) represents the number of rapeseed seedlings not detected by the model. AP refers to the accuracy and under line area of the recall curves plotted at different confidences.

$$AP = \int_0^1 Precision * Recall \, dt \tag{4}$$

The higher the AP value, the better the detection effect. The F1 score is regarded as a weighted average of the precision and recall, and a larger F1 value means a better model.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{5}$$

3 Results and Discussion

3.1 Detection Performance

All projects in this article were written on Windows 10, Pytorch 1.13 and Python 3.9. The training of the model was performed on a computer platform consisting of GPU-NVIDIA and GeForce RTX 3090Ti (24 GB memory), CPU i7-12700F processor. Mosaic data enhancement was used to the process images during the training phase. The images were compressed to 640 * 640 size and then input to improved YOLOv5s. The improved YOLOv5s was executed for 300 generations, with a total of 50 runs per generation, putting in a total of 32 images each generation. The improved YOLOv5 is compared with three single-stage object detection models (YOLOv4, YOLOv3, SSD) and a two-stage object detection model (Faster-RCNN) to evaluate the performance. More details of the hyperparameter settings are displayed in Table 1.

Table 1: Model training hyperparameters

Class	Batch	Optimizer	Learning rate	Decay(/Iterations)	Iterations
Proposed	32	SGD	0.01	$5 * 10^{-4}$	300
YOLOv5s	32	SGD	0.01	$5 * 10^{-4}$	300
YOLOv4	16	SGD	0.01	$5 * 10^{-4}$	300
YOLOv3	32	SGD	0.01	$5 * 10^{-4}$	300
SSD	32	SGD	$2 * 10^{-3}$	$5 * 10^{-4}$	300
Faster-RCNN	16	Adam	$1 * 10^{-4}$	0	300

The divided 400 pictures were used as a test set and sent into the trained model. The curve of change of Precision, Recall, AP, and Loss values is plotted in Fig. 9. As can be seen that the network converges fast, and Precision, Recall and AP show a rapid growth trend within about 100 generations. The final Precision, Recall, AP, and Loss values are 0.982, 0.974, 0.991, and 0.132, respectively.

The detection of the improved YOLOv5s on rapeseed seedlings is shown in Fig. 10. The improved YOLOv5s can exactly identify rapeseed seedlings taken at different heights in the presence of occlusion and overlap.

In order to further compare the interest extracted by the three detection heads of rapeseed at different heights in the improved YOLOv5s. Grad-CAM [29] was taken as a feature visualization tool to visualize the area of interest for pictures taken at a height of 3 m and 45 degrees, pictures taken vertically at a height of 3 m, and two pictures taken vertically at a height of 6 m. Fig. 11 shows the features extracted by the 80 * 80 detection head are finer and it is easier to identify small objects. The area of interest is concentrated on the rape seedlings for the 6 m high shot images. The 40 * 40 detection head is more accurate for 3 m oblique 45° images of rape seedlings, and the area of interest falls on the rape seedlings.

The 20 * 20 inspection head focuses more on the edge and the whole. All three detection heads can identify the area of rapeseed seedlings as the area of interest.
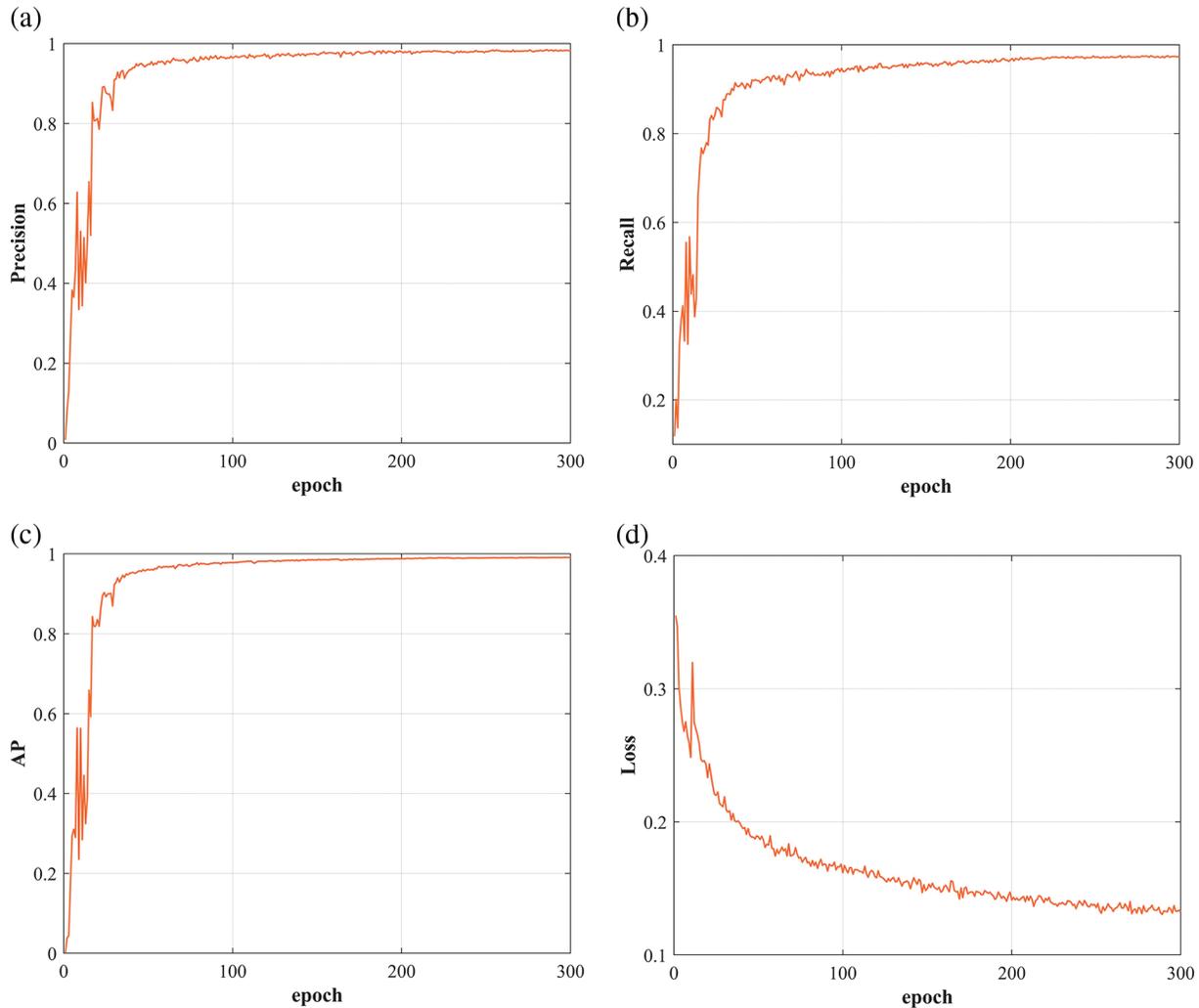

Figure 9: Plotting of (a) precision, (b) recall, (c) AP and (d) loss curves when training the improved YOLOv5s

3.2 Comparison of YOLOv5s with Other Object Detection Models

Building a small, accurate and fast seedling detection model for rapeseed facilitates rapid calculation of seedling numbers and is also easy to deploy on UAV platforms. In this section, the performance of detecting rapeseed seedlings based on YOLOv5s, YOLOv4, YOLOv3, SSD, Faster-RCNN and the proposed method were compared. To ensure consistency and comparability of the results, the performance of each model was performed using the same training dataset set as well as the test dataset, and the results of the test set were shown in Table 2.
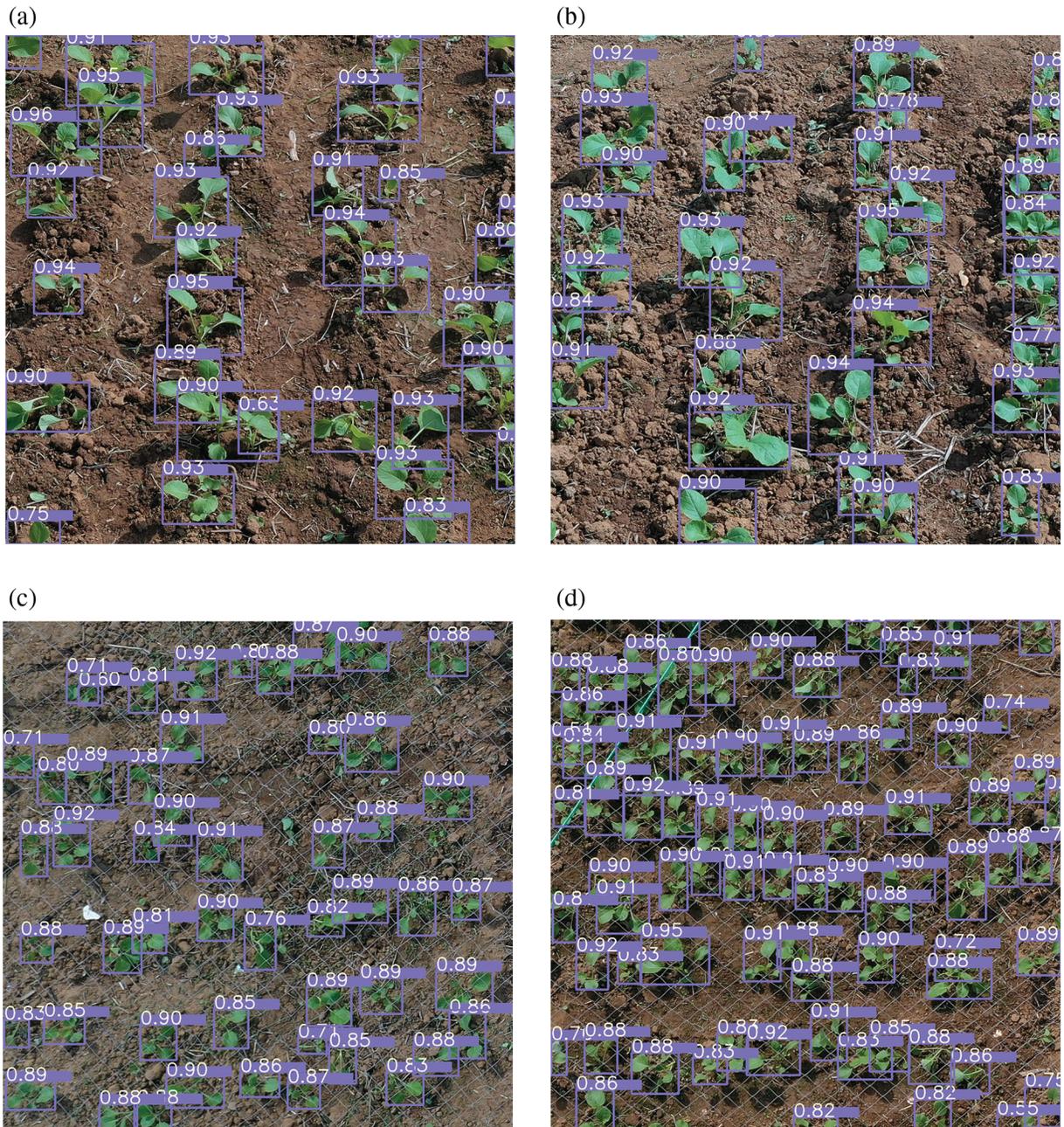
Figure 10: Improved YOLOv5s model prediction result, (a) and (b) for 3 m vertical shot, (c) and (d) for 6 m vertical shot

The improved YOLOv5s had better precision, recall, AP, F1 score and model parameter size than those obtained using the original YOLOv5s, by adding the CA attention mechanism and replacing the SC with the GSConv module. The precision and recall rate of detecting the rapeseed seedlings were 1.9% and 3.7% higher than traditional YOLOv5s, respectively. The improved YOLOv5s model parameters size compared to the original YOLOv5 were 6% smaller by replacing with SC modules. The inference time of the improved YOLOv5s was 12 fps slower than YOLOv5s due to the added CA attention mechanism consuming some computational resources.
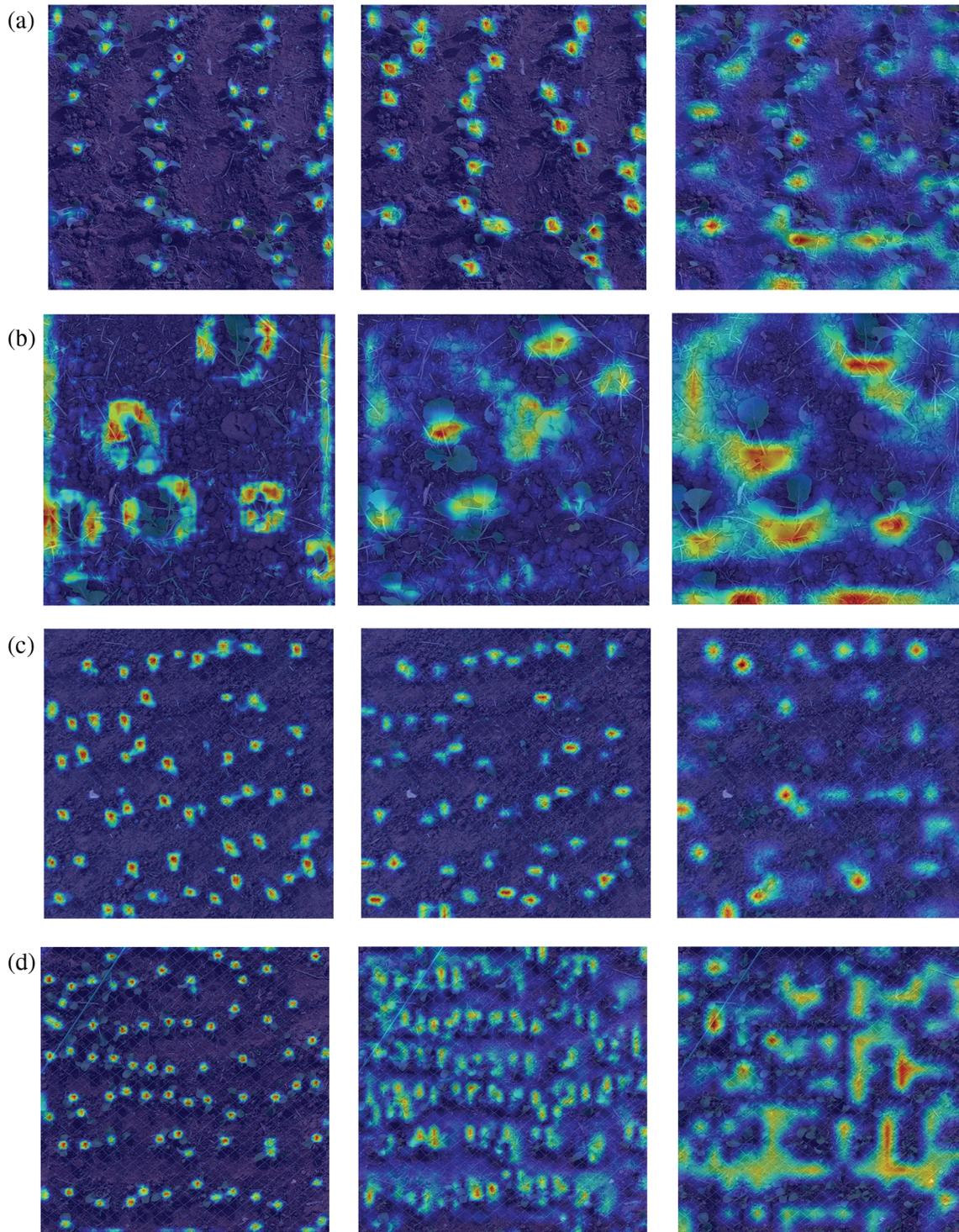
Figure 11: Visualization of feature maps generated by 3 detection heads for images at different shooting heights, (a) and (b) 3 m vertical shot, (c) and (d) 6 m vertical shot

Table 2: Comparison of different methods based on test dataset

Method	Precision	Recall	AP@0.5	F1	FPS	Model size (Mb)
Proposed	**0.982**	**0.974**	**0.991**	**0.977**	116	**13.2**
YOLOv5s	0.962	0.937	0.978	0.950	**128**	14.0
YOLOv4	0.896	0.808	0.869	0.849	43	250.5
YOLOv3	0.935	0.900	0.951	0.917	62	240.6
SSD	0.836	0.526	0.779	0.646	59	92.8
Faster-RCNN	0.681	0.866	0.819	0.762	33	110.7

YOLO series has better performance in terms of accuracy, recall, AP and F1 score. The performance of SSD and Faster-RCNN in terms of detection performance and inference speed was not satisfactory compared with the YOLO series, which is similar to the results reported by [30]. Therefore, they were not applicable in the detection of small objects in the present study of rape seedlings. SSD and Faster-RCNN have less than satisfactory performance in detection performance and inference speed, with 21.2% and 17.2% lower AP metrics obtained compared to the improved YOLOv5s, respectively, and thus are not applicable in the detection of small objects of rape seedlings in this study. Considering that detection accuracy is an important indicator affecting object tracking, and smaller models can be better deployed on equipment in the future. Therefore, the improved YOLOv5s, which can also meet the requirements in terms of inference speed, can be better applied in counting the number of rapeseed seedlings.

Fig. 12 shows the detection results of the six methods at two heights, where the confidence and IOU values were set to 0.3. The improved YOLOv5s detected smaller and overlapping rapeseed seedlings. The standard YOLOv5, YOLOv4 and YOLOv3 showed missing detection in images collected at 3 m height with overlapping rapeseed seedlings. SSD and Faster-RCNN methods had more misses for smaller rapeseed seedlings when detecting images collected at 6 m height.

3.3 Counting Accuracy

The results of the final rapeseed seedling counting were achieved using functions from the Opencv library, and the final results are displayed in Fig. 13. As the UAV forward, the red dots successively pass the pink and light purple lines and the number is added by one.

A total of 15 videos with a resolution of 1920 * 1080 were recorded using the UAV at 1:00 pm on October 20, 2022. 15 videos were counted manually and mechanically. The results of the comparison were illustrated in Fig. 14. The fitted lines of the 15 sample points were very close to the 1:1 line. The methods used were very close to the results of manual techniques, and their average error was 4.3%.

The average relative error for video-based tea bud counting was 8.12% [22], for cotton counting based on single-stage detector and optical flow method was 3.13% [23] as well as for peanut counting based on YOLOv5 and deep-sort was 1.92% [24]. It can be found from the related literature that there are some fluctuations in the counting errors for different plants counting. The error of 4.3% is acceptable considering the existence of false detection when detecting the number of oilseed rape seedlings with overlapping.
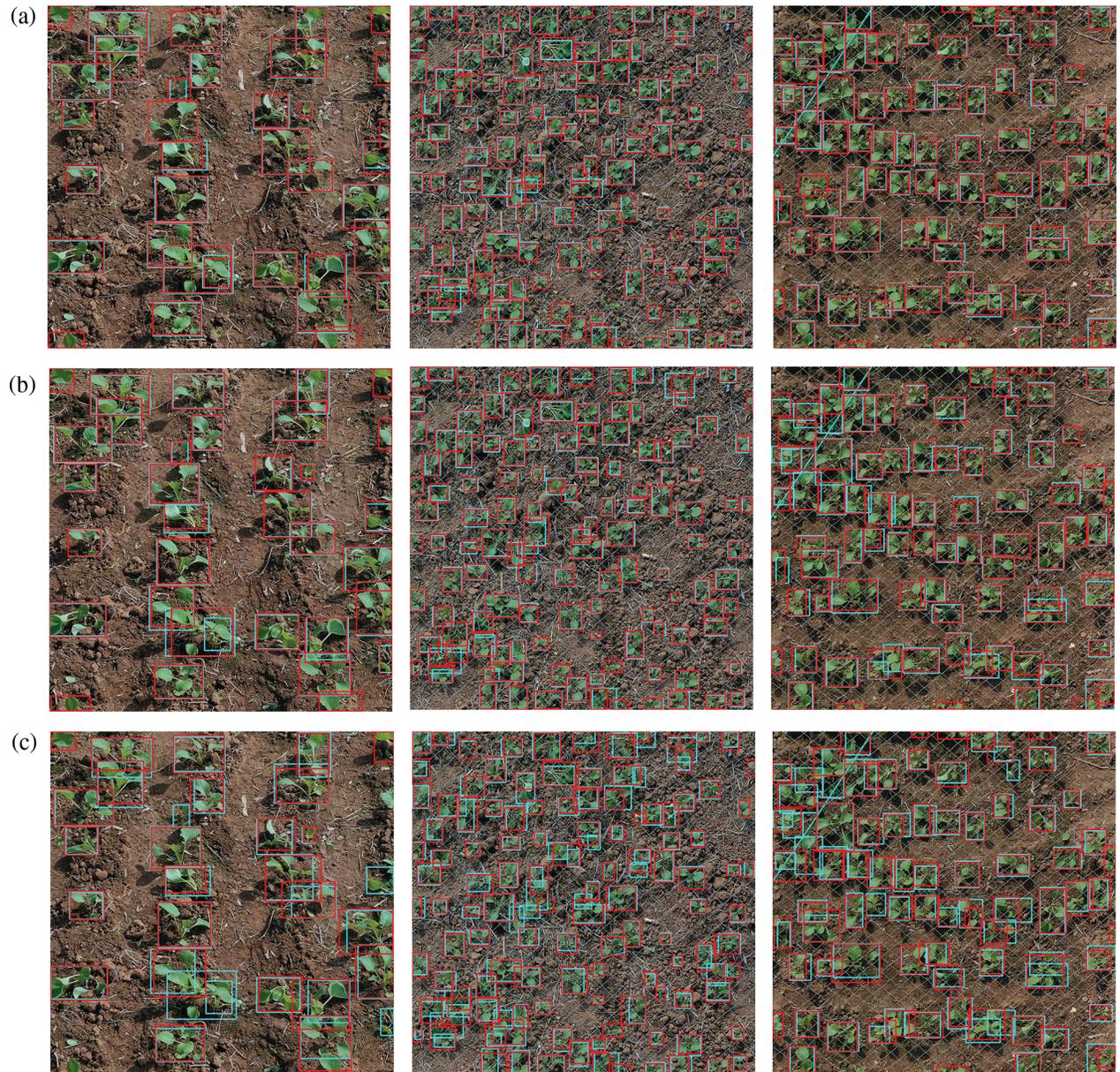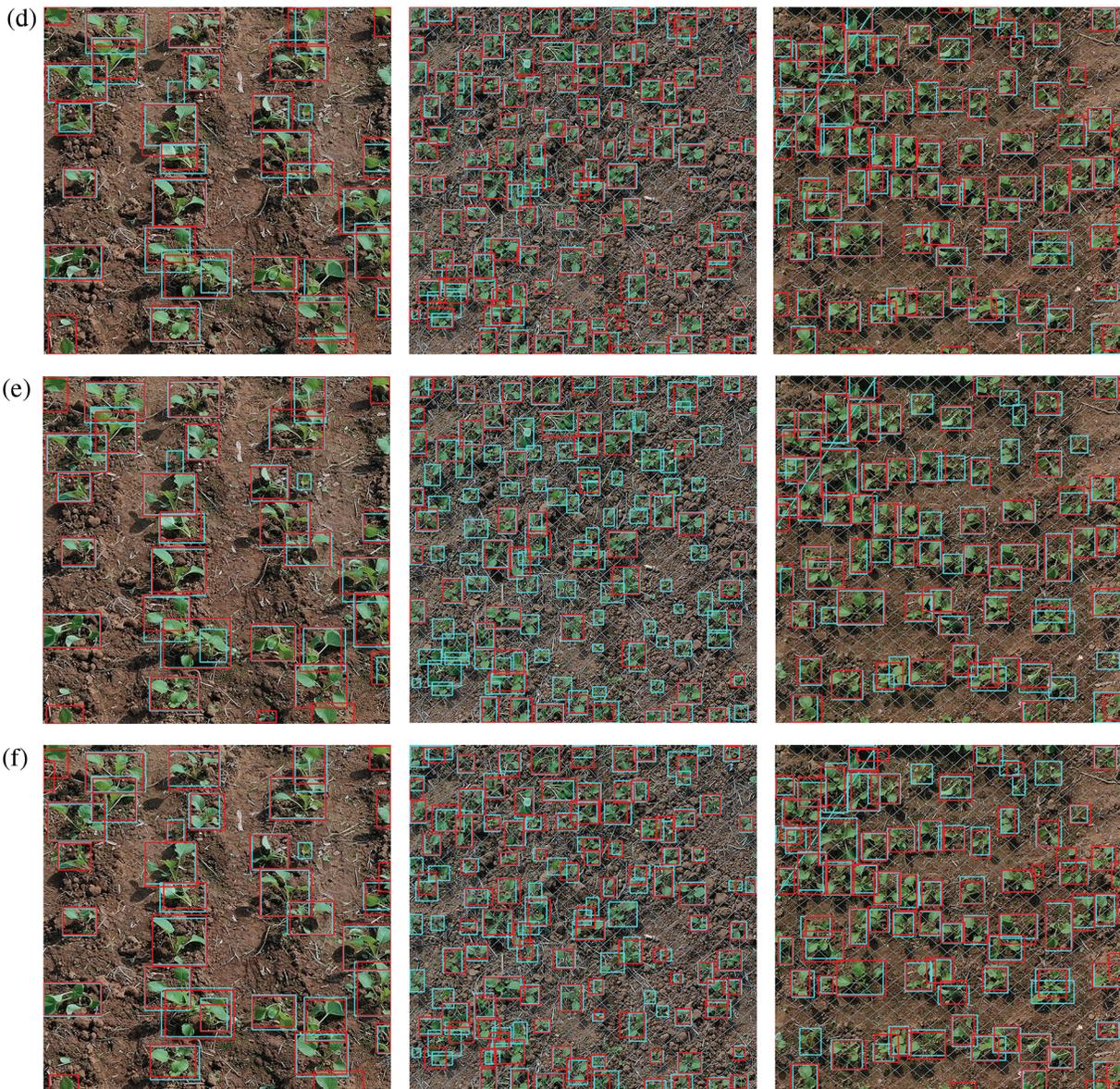
(a)

(b)

(c)

Figure 12: (Continued)

Figure 12: Detection results of 6 methods of (a) improved YOLOv5s, (b) YOLOv5s, (c) YOLOv4, (d) YOLOv3, (e) SSD and (f) Faster-RCNN, blue boxes are labels, red boxes are detection results, the first column is 3 m shooting high image, the last two columns are 6 m high shooting images

These showed that the use of UAV to detect rapeseed seedlings is an efficient and accurate method, which can save labor costs and the proposed real-time rapeseed video counting model can be applied to early rapeseed seedlings detection. However, counting accuracy sometimes affected by detection accuracy, as shown in Fig. 15, weeds are identified as seedlings when they are very similar to seedlings, while small seedlings are sometimes missed. Therefore, the accuracy of detection is a factor limiting the accuracy of counting.
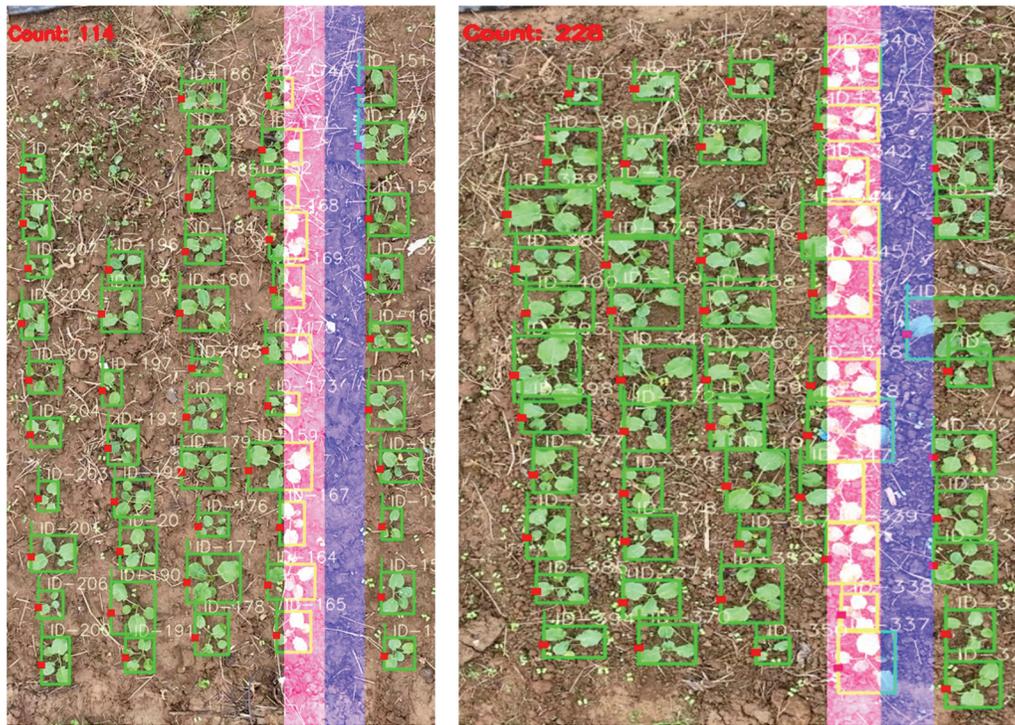
Figure 13: Video calculation of the number of rapeseed seedlings (the green boxes are the detected rapeseed seedling, and the red dots on the green box are used as bump detection points)
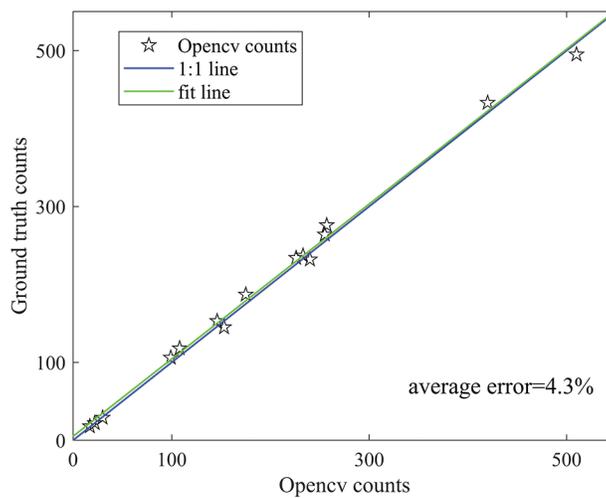

Figure 14: Counting results of rapeseed seedlings using deep-sort method on 15 video datasets

3.4 Number Detection Platform for Rapeseed Seedlings Based on PyQt5

In order to more conveniently apply the proposed method to rapeseed seedling detection, PyQt5 is used to construct a visual detection platform that can identify the number of rapeseed seedlings in pictures or videos, as shown in Fig. 16. The platform includes the selection weights for YOLOv5 model training, model initialization, image detection and video counting detection.
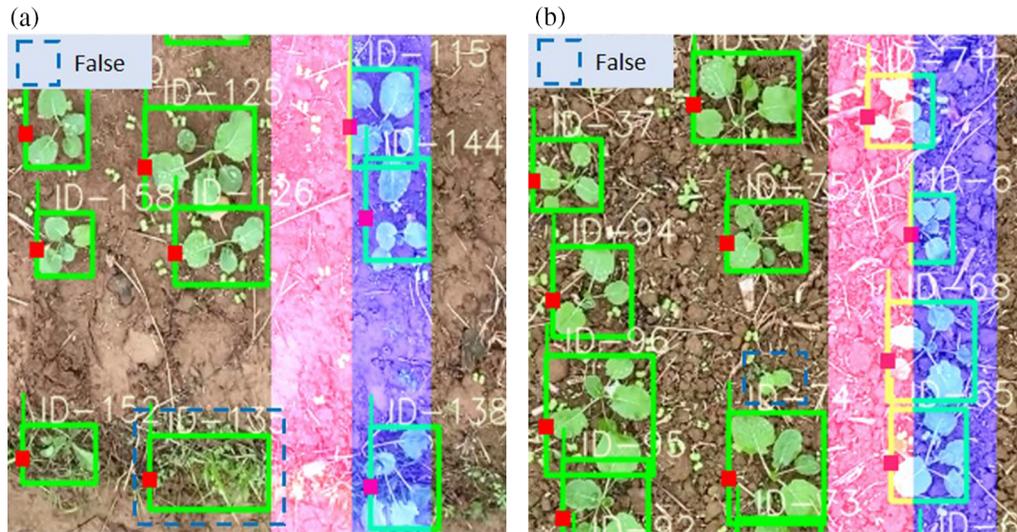
Figure 15: Example of mis-tracking. (a) The weed was detected as a seedling and the count was increased (b) the seedling was not detected and the count was decreased

Figure 16: Number detection platform for rapeseed seedlings based on PyQt5

The operation process of the entire detection platform is: first step is to click the select weight button, and the address of the weight will be recorded. Second step is to build out the model structure and initialize the model according to the selected weight. In the third step, clicking the image recognition or video recognition buttons, the image or video will be displayed in the window on the right. The total number of rapeseed seedlings from the result output window below to meet the rapid detection of rapeseed seedlings.

4 Conclusion

In this research, a rapeseed seedling detection and counting method was developed, which used a one-stage object detection network to identify rapeseed and deep-sort to track object. A new cropping method was proposed for high-resolution rapeseed images. In order to recognize smaller rape seedlings with some overlap more accurately identify, YOLOv5s was improved by adding a CA attention mechanism in the backbone and replacing the standard convolution using GSConv in the neck. The accuracy of the improved YOLOv5s to identify the overlapping, dense and small rape was improved and the precision and recall are 0.982 and 0.974, which were 1.9% and 3.7% better than the original YOLOv5s. A field

test was conducted to identify the number of rape seedlings for 15 videos recorded by the UAV. The average error between the method used and manual counting was 4.3%. The results showed that the improved YOLOv5 and deep-sort can be applied to practical rape seedling counting, which was conducive to breeding programs and accurate crop management. In the future, we will consider applying the method to other crops in an effort to provide some assistance to crop breeding and crop management.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Chen Su; data collection: Chen Su, Jiang Wang and Jie Hong; analysis and interpretation of results: Chen Su; draft manuscript preparation: Yang Yang and Chen Su. All authors reviewed the results and approved the final version of the manuscript.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Ortega Ramos, P. A., Coston, D. J., Seimandi Corda, G., Mauchline, A. L., Cook, S. M. (2022). Integrated pest management strategies for cabbage stem flea beetle (*Psylliodes chrysocephala*) in oilseed rape. *GCB Bioenergy, 14(3),* 267–286.

2. Ashkiani, A., Sayfzadeh, S., Shirani Rad, A. H., Valadabadi, A., Hadidi Masouleh, E. (2020). Effects of foliar zinc application on yield and oil quality of rapeseed genotypes under drought stress. *Journal of Plant Nutrition, 43(11),* 1594–1603.

3. Snowdon, R. J., Iniguez Luy, F. L. (2012). Potential to improve oilseed rape and canola breeding in the genomics era. *Plant Breeding, 131(3),* 351–360.

4. Liu, S. Y., Baret, F., Allard, D., Jin, X. L., Andrieu, B. et al. (2017). A method to estimate plant density and plant spacing heterogeneity: Application to wheat crops. *Plant Methods, 13(1),* 1–11.

5. Momoh, E., Zhou, W. (2001). Growth and yield responses to plant density and stage of transplanting in winter oilseed rape (*Brassica napus* L.). *Journal of Agronomy and Crop Science, 186(4),* 253–259.

6. Luck, G. W. (2007). A review of the relationships between human population density and biodiversity. *Biological Reviews, 82(4),* 607–645.

7. Bao, W. X., Liu, W. Q., Yang, X. J., Hu, G. S., Zhang, D. Y. et al. (2023). Adaptively spatial feature fusion network: An improved UAV detection method for wheat scab. *Precision Agriculture, 24(3),* 1154–1180.

8. Zhou, C. Q., Yang, G. J., Liang, D., Yang, X. D., Xu, B. (2018). An integrated skeleton extraction and pruning method for spatial recognition of maize seedlings in MGV and UAV remote images. *IEEE Transactions on Geoscience and Remote Sensing, 56(8),* 4618–4632.

9. Bai, Y., Nie, C. W., Wang, H. W., Cheng, M. H., Liu, S. B. et al. (2022). A fast and robust method for plant count in sunflower and maize at different seedling stages using high-resolution UAV RGB imagery. *Precision Agriculture, 23(5),* 1720–1742.

10. Liu, S. B., Yin, D. M., Feng, H. K., Li, Z. H., Xu, X. B. et al. (2022). Estimating maize seedling number with UAV RGB images and advanced image processing methods. *Precision Agriculture, 23(5),* 1–29.

11. Zhao, B. Q., Zhang, J., Yang, C. H., Zhou, G. S., Ding, Y. C. et al. (2018). Rapeseed seedling stand counting and seeding performance evaluation at two early growth stages based on unmanned aerial vehicle imagery. *Frontiers in Plant Science, 9,* 1362.

12. Albahar, M. (2023). A survey on deep learning and its impact on agriculture: Challenges and opportunities. *Agriculture, 13(3),* 540.

13. Girshick, R. (2015). Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448. Santiago.

14. Zhai, S. P., Shang, D. R., Wang, S. H., Dong, S. S. (2020). DF-SSD: An improved SSD object detection algorithm based on DenseNet and feature fusion. *IEEE Access, 8,* 24344–24357.

15. Bochkovskiy, A., Wang, C. Y., Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

16. Redmon, J., Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767.

17. Mu, Y., Chen, T. S., Ninomiya, S., Guo, W. (2020). Intact detection of highly occluded immature tomatoes on plants using deep learning techniques. *Sensors, 20(10),* 2984.

18. Liu, M. G., Su, W. H., Wang, X. Q. (2023). Quantitative evaluation of maize emergence using UAV imagery and deep learning. *Remote Sensing, 15(8),* 1979.

19. Pei, H. T., Sun, Y. Q., Huang, H., Zhang, W., Sheng, J. J. et al. (2022). Weed detection in maize fields by UAV images based on crop row preprocessing and improved YOLOv4. *Agriculture, 12(7),* 975.

20. Zhao, J. Q., Zhang, X. H., Yan, J. W., Qiu, X. L., Yao, X. et al. (2021). A wheat spike detection method in UAV images based on improved YOLOv5. *Remote Sensing, 13(16),* 3095.

21. Luo, W. H., Xing, J. L., Milan, A., Zhang, X. Q., Liu, W. et al. (2021). Multiple object tracking: A literature review. *Artificial Intelligence, 293(2),* 103448.

22. Li, Y., Ma, R., Zhang, R. T., Cheng, Y. F., Dong, C. W. (2023). A tea buds counting method based on YOLOv5 and kalman filter tracking algorithm. *Plant Phenomics, 5(8),* 0030.

23. Tan, C. J., Li, C. Y., He, D. J., Song, H. B. (2022). Towards real-time tracking and counting of seedlings with a one-stage detector and optical flow. *Computers and Electronics in Agriculture, 193(1),* 106683.

24. Lin, Y. D., Chen, T. T., Liu, S. Y., Cai, Y. L., Shi, H. W. et al. (2022). Quick and accurate monitoring peanut seedlings emergence rate through UAV video and deep learning. *Computers and Electronics in Agriculture, 197(1),* 106938.

25. Cui, Z. Y., Wang, X. Y., Liu, N. Y., Cao, Z. J., Yang, J. Y. (2020). Ship detection in large-scale SAR images via spatial shuffle-group enhance attention. *IEEE Transactions on Geoscience and Remote Sensing, 59(1),* 379–391.

26. Shi, C. C., Lin, L., Sun, J., Su, W., Yang, H. et al. (2022). A lightweight YOLOv5 transmission line defect detection method based on coordinate attention. *2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC)*, pp. 1779–1785. Chongqing, China.

27. Li, H. L., Li, J., Wei, H. B., Liu, Z., Zhan, Z. F. et al. (2022). Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles. arXiv preprint arXiv:2206.02424.

28. Wojke, N., Bewley, A., Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. *2017 IEEE International Conference on Image Processing*, pp. 3645–3649. Beijing, China.

29. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. et al. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision*, pp. 618–626. Venice.

30. Naftali, M. G., Sulistyawan, J. S., Julian, K. (2022). Comparison of object detection algorithms for street-level objects. arXiv preprint arXiv:2208.11315.