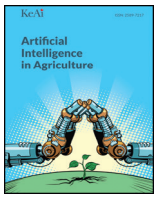
Artificial Intelligence in Agriculture 13 (2024) 117–127

Contents lists available at ScienceDirect

Artificial Intelligence in Agriculture

journal homepage: http://www.keaipublishing.com/en/journals/artificial-intelligence-in-agriculture/



UAV-based field watermelon detection and counting using YOLOv8s with image panorama stitching and overlap partitioning

Liguo Jiang [a], Hanhui Jiang [a], Xudong Jing [a], Haojie Dang [a], Rui Li [a], Jinyong Chen [c], Yaqoob Majeed [d], Ramesh Sahni [e], Longsheng Fu [a,b,*]

[a] College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
[b] Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi 712100, China
[c] Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou, Henan 450009, China
[d] Department of Biological & Agricultural Engineering, Texas A&M University, Dallas, TX 75252, United States
[e] Agricultural Mechanization Division, ICAR-Central Institute of Agricultural Engineering, Bhopal 462038, Madhya Pradesh, India

ARTICLE INFO

Article history:
Received 30 April 2024
Received in revised form 6 August 2024
Accepted 2 September 2024
Available online 12 September 2024

Keywords:
Watermelon yield estimation
Unmanned aerial vehicle
Object detection
Panorama stitching
Overlap partitioning

ABSTRACT

Accurate watermelon yield estimation is crucial to the agricultural value chain, as it guides the allocation of agricultural resources as well as facilitates inventory and logistics planning. The conventional method of watermelon yield estimation relies heavily on manual labor, which is both time-consuming and labor-intensive. To address this, this work proposes an algorithmic pipeline that utilizes unmanned aerial vehicle (UAV) videos for detection and counting of watermelons. This pipeline uses You Only Look Once version 8 s (YOLOv8s) with panorama stitching and overlap partitioning, which facilitates the overall number estimation of watermelons in field. The watermelon detection model, based on YOLOv8s and obtained using transfer learning, achieved a detection accuracy of 99.20 %, demonstrating its potential for application in yield estimation. The panorama stitching and overlap partitioning based detection and counting method uses panoramic images as input and effectively mitigates the duplications compared with the video tracking based detection and counting method. The counting accuracy reached over 96.61 %, proving a promising application for yield estimation. The high accuracy demonstrates the feasibility of applying this method for overall yield estimation in large watermelon fields.

1. Introduction

Yield estimation plays a crucial role in watermelon production, informing postharvest decisions related to resource allocation and marketing strategies (Ho et al., 2019). Accurate yield estimation of watermelons greatly benefits growers as it enables them to assess available yield for retailers (Dhanya et al., 2022). Despite its importance, watermelon yield estimation is generally carried out by manual counting due to the lack of an automatic and accurate yield estimation method (Kalantar et al., 2020). Consequently, there is a strong desire for an automated watermelon counting method that serves as an effective alternative to manual counting.

Unmanned aerial vehicles (UAVs) are a feasible option for data acquisition in vast watermelon fields due to their maneuverability and broad coverage capability. The data collection for the entire watermelon field can be completed within a short period of time (Miranda et al., 2019; Velusamy et al., 2021). Moreover, the low-altitude aerial perspective of UAVs allows for more intensive data compared with the perspective of satellite, facilitating a rapid and efficient monitoring of vast watermelon fields (Feng et al., 2022; Khokher et al., 2023). Furthermore, the images captured by UAVs equipped with RGB cameras often exhibit high resolution, showcasing the fine features of the watermelon fields with precision (Gao et al., 2023; Liao et al., 2023; Luna and Lobo, 2016). And the UAV images taken in different locations have multiple perspectives, preventing the loss of information due to watermelon vine occlusion (Cui et al., 2023). Additionally, geolocation information is tagged with UAV images, which facilitates subsequent operations such as localization and orthomosaic image generation (Xiao et al., 2022).

Accurate detection of watermelons in field is a prerequisite for watermelon yield estimation using UAVs (Milioto et al., 2017). However, the watermelons are relatively small compared to watermelon fields, while the field environment shares similar textures and colors with the watermelons, posing a challenge for detection. In recent years, several studies have been conducted by researchers on fruit detection. Various approaches, including K-means clustering

* Corresponding author at: College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China.
E-mail address: fulsh@nwafu.edu.cn (L. Fu).

https://doi.org/10.1016/j.aiia.2024.09.001

(Jiao et al., 2020), Bayesian classifiers and mathematical morphology (Hsu et al., 2019) have been utilized for fruit detection in field. These conventional methods primarily rely on colour, shape, or texture and frequently necessitate the utilization of fixed thresholds for detection (Fu et al., 2021). Therefore, these approaches are typically suitable for environments with simple backgrounds and struggle to achieve robust detection under complex environmental conditions (Gao et al., 2022).

Recent advancements in deep learning have achieved promising outcomes in object detection, improving both the precision and robustness (Song et al., 2023; Zhao et al., 2017). This technique enables autonomous learning of both superficial and profound features, resulting in enhanced adaptability of detection. One object detection algorithm based on deep learning is You Only Look Once (YOLO), which strikes a balance between accuracy and speed (Fu et al., 2021). YOLO redefines object detection as a regression problem and uses a single-stage algorithm to achieve efficient detection and accurate results (Tripathi et al., 2022). Within the agricultural sphere, Gao et al. (2022) developed an apple detecting method based on YOLOv4-tiny and achieved a mean average precision (mAP) of 99.35 %. Nan et al. (2023) developed a new WGB-YOLO network based on YOLO, increasing mAP for multi-level pitaya detection to 86.00 % in the target pick line. YOLO performs well in natural field environments and can accurately detect fruits (Tang et al., 2023). Utilizing YOLO for watermelon detection has the potential to enhance accuracy.

Counting is a crucial step for yield estimation of watermelons by UAVs after watermelon detection. There are two primary methods for fruit counting. The first method is a video tracking based detection and counting method (VTDCM), which involves establishing correlations between same targets across consecutive video frames (Gao et al., 2022). Liu et al. (2019) developed a mango fruit counting system based on the Kalman filter to track all detected fruits and reached an R^2 value of 0.88 compared to the actual number. Although the VTDCM has achieved acceptable results, this method requires a lot of computing resources. Meanwhile, the high accuracy of the VTDCM highly relies on the accuracy of detection. However, the small size of watermelon targets in aerial perspective leads to unsatisfactory tracking results (Guo et al., 2023). Some other studies have achieved good results by using other field markers instead of fruits for detection and tracking, but watermelon fields lack similar markers, making this method unsuitable (Gao et al., 2022). The second method is inspired by the concept of You Only Look Twice (Van Etten, 2018) and treats watermelon object detection as the detection of small objects within a wide-ranging environment. This technique uses panorama stitching and overlap partitioning to count watermelons. This paper proposes a panorama stitching and overlap partitioning based detection and counting method (PSOPDCM) and compares it with the VTDCM.

This work aims to construct a field watermelon detection and counting method based on YOLOv8 with panorama stitching and overlap partitioning. The remaining of the work is organized as follows: In Section 2, the materials and methods are described in terms of specific method. In Section 3, the results are presented and discussed. Lastly, in Section 4, conclusions of this work are described.

2. Materials and methods

2.1. Data acquisition

The field data of watermelon was collected in Fufeng County, Baoji City, Shaanxi Province, China (34°25′25″N, 107°58′18″E) during the 2022 harvest season. The watermelon field covers a vast area, approximately 15 m wide and 150 m long. However, due to the field of view limitation of the UAV camera, a single image cannot fully cover the entire field, making it insufficient for estimating the total yield by a single image. To overcome this limitation and obtain an estimate of the overall yield, videos were recorded to the entire field. The data was captured by a DJI Mavic 2 Pro (SZ DJI Technology Co., Ltd., Shenzhen, China) drone, equipped with a Hasselblad L1D-20c RGB camera. The drone was positioned at an altitude of 10 m above the field, with the camera pointing vertically downward during acquisition of videos. To ensure consistency and comparability of the data, the videos were recorded at a resolution of 3840 × 2160 pixels with a frame rate of 30 frames per second. For the three watermelon fields shown in Fig. 1, three sets of videos were recorded and labeled as A, B and C, respectively. Video A was used to construct the object detection dataset and test the object detection model, while videos B and C were used to evaluate the counting methods. All videos were recorded under natural daylight conditions, including natural interference and overlap.

Fig. 1. The scene of the watermelon fields. The yellow area represents the shooting area of video A, which is used for dataset construction. The two red areas represent the shooting areas of videos B and C, which are utilized for counting evaluation.

2.2. Dataset building

Dataset is the foundation of object detection model. Frames were extracted from the captured watermelon video A (1 min 22 s) based on the video timestamp, at a frequency of one frame every 1.5 s, resulting in 55 initial images with a resolution of 3840 × 2160 pixels in JPEG format.

Considering the pixels of watermelons (approximately 30 × 40 pixels), watermelons are too small to be detected in the initial images. To address this issue, the dataset was overlap-partitioned to increase the proportion of watermelon pixels in the images. Each initial image was overlap partitioned into 45 patches with 20 % overlap at a resolution of 640 × 640 pixels (Fig. 2). The annotation was completed by labelImg (https://github.com/tzutalin/labelImg), whereby all watermelon targets in the patches were manually annotated with a label “watermelon” using rectangular bounding boxes (Fig. 3). The annotation file was saved in XML format. After annotation, a dataset comprising 2475 images, containing watermelon targets and their respective annotated files, was established.

Small number of training set may cause overfitting or nonconvergence of object detection model training (Jiang et al., 2023). Data augmentation was thus employed to expand the watermelon dataset. Various augmentations were utilized, including brightness transformation with rates of 0.8 and 1.2 (Fig. 4b and Fig. 4c), vertical image mirroring (Fig. 4d), horizontal image mirroring (Fig. 4e), and Gaussian blur (Fig. 4f). A dataset containing 14,850 watermelon images was established using these augmentation strategies. To validate the accuracy of the model, the dataset is divided into a training set and a validation set in a ratio of 4:1.

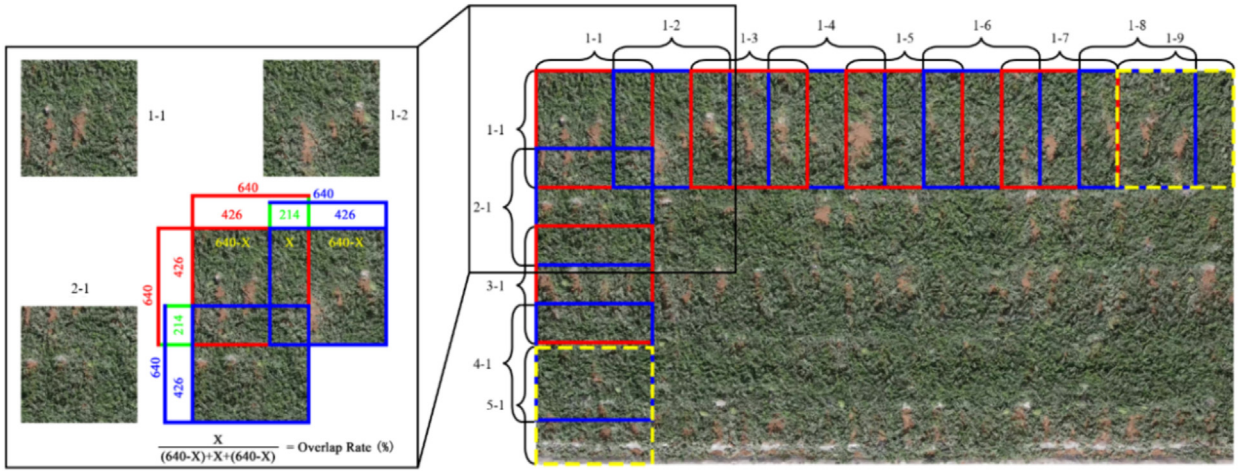


Fig. 2. Schematic diagram of the overlap partitioning algorithm. The right side shows an example of the patches arranged in rows and columns of the initial image. The process of dividing a 3840 × 2160 image into 45 sub-images of 640 × 640 is shown. The patch with red borders overlaps the patch with blue borders by an overlap rate of 20 %. In case where the cropping pixels do not fit precisely with the edge of the image, the cropping box will be adjusted forward by the corresponding number of pixels, as illustrated by the yellow box. The left side shows an enlarged view of the cropping position with an overlap rate of 20 % and examples of cropped sub-images.

2.3. YOLOv8 model

YOLO redefines object detection as a regression problem, enabling fast and accurate detection by simultaneously predicting bounding boxes and class probabilities. In this paper, YOLOv8 detect model (https://github.com/ultralytics/ultralytics) is employed, which uses a Cross Stage Partial Network (CSPNet) as its backbone to extract image features. This setup reduces redundant gradient information and floating-point operations per second (FLOPS). The neck network structure of the YOLOv8 algorithm, which combines the concepts of FPN (Feature Pyramid Network) and PAN (Path Aggregation Network), is modeled as a network structure of FPN + PAN (Lin et al., 2017; Liu et al., 2018), expanding the receptive field before being passed through the feature aggregation architecture. The head of the YOLOv8 algorithm adopts the Decoupled-Head structure, which separates the classification and detection heads and utilizes the idea of Distributional Focal Loss (Liu et al., 2024). The Network architecture diagram of YOLOv8 is shown in Fig. 5. YOLOv8 provides YOLOv8s, YOLOv8n, YOLOv8m, YOLOv8l, and YOLOv8x, differentiated by parameters such as depth_multiple and width_multiple. While more complex models yield higher accuracy, they also result in slower detection speed and larger model size (Li et al., 2022; Nepal and Eslamiat, 2022).

2.4. Network training

Both the training and testing of models were conducted on a computer with AMD Ryzen 7 5800X CPU, NVIDIA GeForce RTX 3080Ti 12 GB GPU, and 64 GB memory. The software utilized included CUDA 11.7 and cuDNN 8.2.2. For training the watermelon detection models, the YOLOv8 architecture was utilized within the PyTorch framework, implementing transfer learning. Transfer learning is a method in machine learning that involves leveraging a pretrained model for training on another task. During training, the input image size was set to 640 × 640 pixels, with a batch size of 16. The learning rate was set to 0.01, and the training was conducted at 300 epochs.

2.5. Process of the VTDCM

In order to complete the video-based counting, video tracking is employed. DeepSORT (Simple Online and Realtime Tracking with a Deep Association Metric) is an extension of the original SORT (Simple Online and Realtime Tracking) algorithm, which integrates deep learning features to enhance tracking performance (Bewley et al., 2016; Wojke et al., 2017).

In this paper, a combination of YOLOv8 and DeepSORT is employed to complete the counting step. This approach utilizes YOLOv8 for object

Fig. 3. Examples of annotated watermelons. (a) The original image. (b) The annotated image. The blue box represents the bounding box of the watermelon target, covering the entire area of the target, and closely adhering to edges.

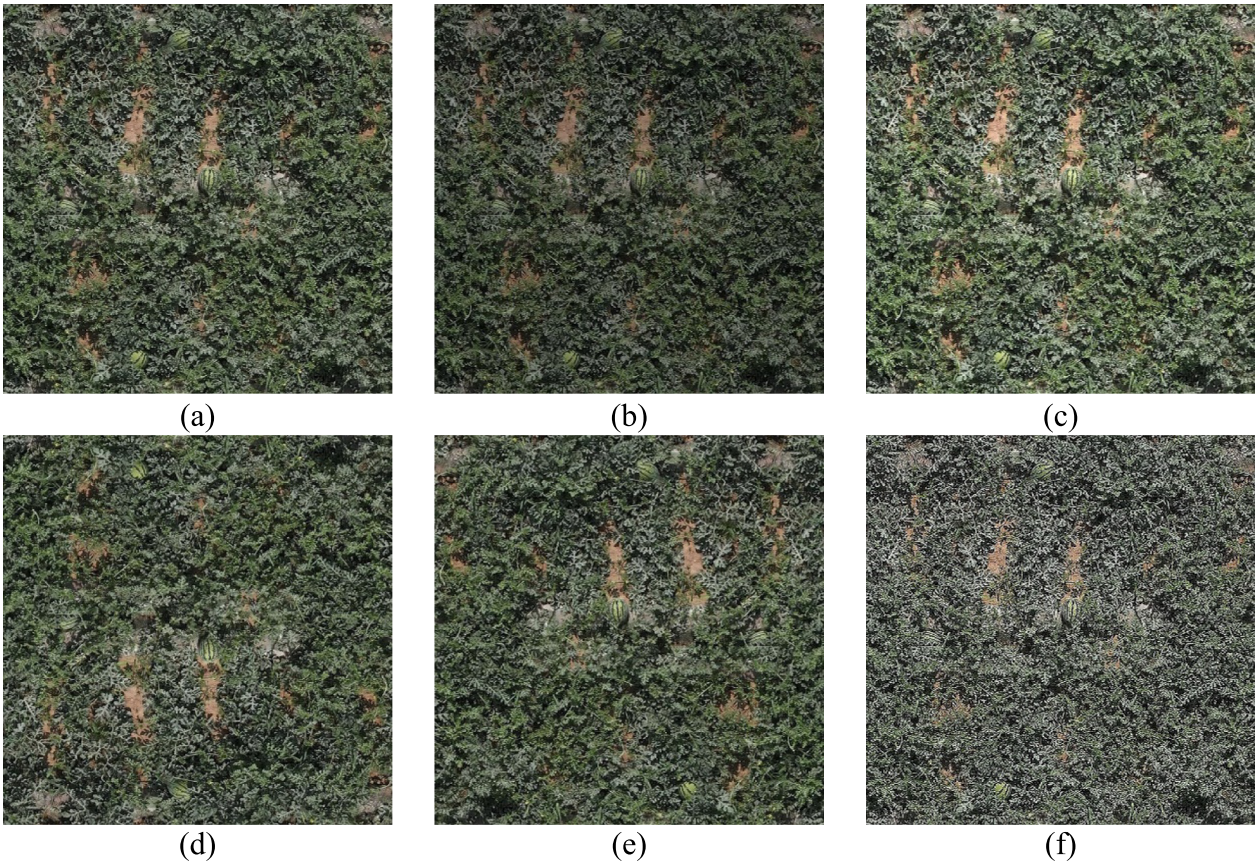


Fig. 4. Data augmentation. (a) Original image, (b) Brightness 0.8, (c) Brightness 1.2, (d) Mirroring in vertical, (e) Mirroring in horizontal, (f) Gaussian blur.

detection. YOLOv8 performs object detection on video frames, detecting objects and providing bounding boxes and category information for the detected targets. The subsequent step involves DeepSORT, which employs the extracted target information to track them using a fusion of Kalman filtering and the Hungarian algorithm. Each target is assigned a unique ID as it is tracked. In this way, the combined efforts of YOLOv8 and DeepSORT enable detection and tracking of multiple targets in videos, with the ability to associate and track these targets across consecutive frames. The detection model used in the VTDCM is trained based on the initial images extracted from video A.

2.6. Process of the PSOPDCM

2.6.1. Panorama stitching

To ensure accurate detection and avoid redundant detection of watermelons in overlapping areas of adjacent images, a panorama stitching process is employed on UAV images to reconstruct a comprehensive view of the watermelon field. The Pix4Dmapper software (Pix4D S.A., Prilly, Switzerland) is utilized to generate orthomosaic images of the watermelon fields. This process helps reduce any potential geometric distortion of the watermelon targets that may arise from the image stitching process.

The process of creating orthomosaic images involves performing aerial triangulation, extracting and matching feature points, images rectifying, and images fusing (Fig. 6). The execution of aerial triangulation is the first step. This process facilitates the determination of precise positions and orientations of the acquired images, enabling subsequent accurate alignment. Next, feature points are extracted and matched across the images. The feature points are key elements for matching and identifying specific locations in images characterized by unique structures such as brightness changes, edges, corners, and textures. By identifying distinctive visual landmarks and finding corresponding feature points in different images, precise alignment is achieved. Following the feature point extraction and matching, the images undergo rectification. This step aims to remove distortions caused by variations in camera perspectives. By rectifying the images, a geometrically accurate representation of the watermelon field is achieved. Finally, the rectified images are fused together to create the orthomosaic. Using drone orthomosaic technology, images captured from different angles or locations harmoniously merge to construct a panoramic view of the watermelon field.

2.6.2. Overlap partitioning

The resolution of the panoramic image is too large to be fed into the object detection network. To achieve watermelon detection in panoramic images, the 20 % overlap partitioning algorithm is utilized. Initially, a panoramic image is partitioned into sub-images with 20 % overlap. These sub-images undergo the detection process individually, yielding location information of all watermelons in each sub-image. Subsequently, the detection results of sub-images are effectively merged and reconstructed, resulting in the panoramic image detection outcomes.

(1) Image partitioning

The panoramic image is partitioned horizontally and vertically, generating sub-images with a resolution of 640 × 640 pixels. Upon loading the panoramic image, its resolution is assessed, and the number of horizontal and vertical partitioning times is calculated based on the resolution of panoramic image and the resolution of 640 × 640 pixels. To account for any potential discrepancy between the partitioning box and the image edges, adjustments are made when the partitioning box cannot fit precisely against the panoramic image edge. In such

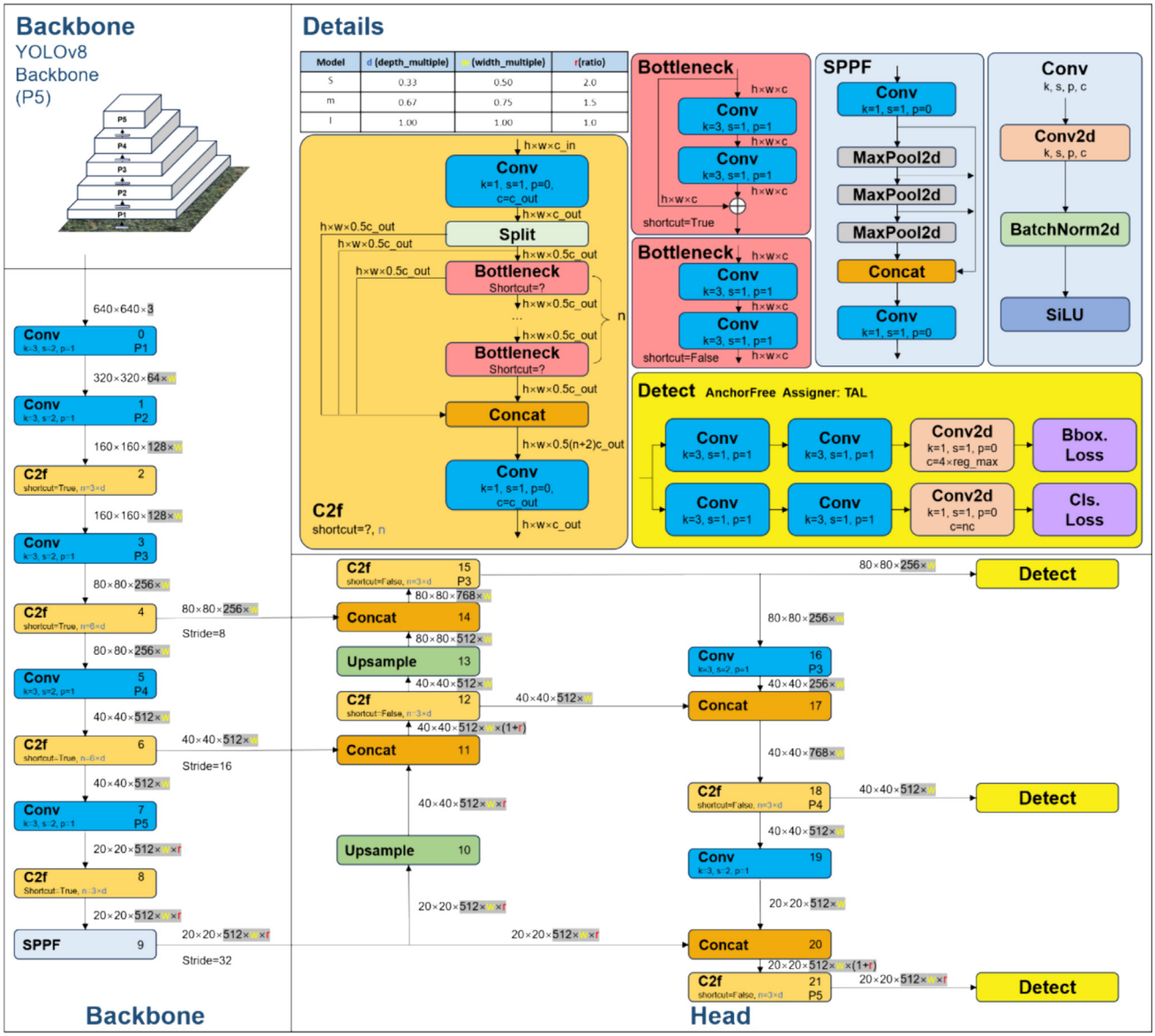


Fig. 5. Network architecture diagram of YOLOv8 (https://github.com/RangeKing).

cases, the partitioning box is shifted forward by the appropriate number of pixels. Starting from the top left corner of the image, the starting and ending positions of partitioning are calculated based on the index of the current sub-image. The specific region corresponding to the current sub-image is cropped using the slice procedure and subsequently saved to the output folder (Fig. 2).

By applying the YOLOv8-based watermelon object detection model to each sub-image, XML files containing information of detected watermelon targets are generated. The XML file contains the pixel locations and confidence information associated with each detected watermelon target. To perform a comprehensive count of the targets within the panoramic image, it is imperative to stitch together the annotated files of the sub-images. This stitching process ensures a unified view, enabling accurate target counting.

(2) Sub-images stitching

The XML file of the panoramic image, which contains all the bounding boxes of the watermelon targets, is generated by stitching together the XML files of the sub-images. First, a coordinate system is established with the same resolution as the panoramic image. Then, according to the partitioning information, the coordinates of the sub-images are converted to the coordinates of the panoramic image. The procedure starts at the upper left corner of the panoramic image and incorporates the XML files of each sub-image into the process.

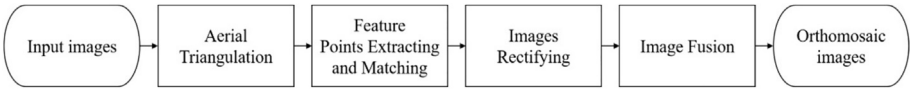


Fig. 6. Orthomosaic image generation steps.

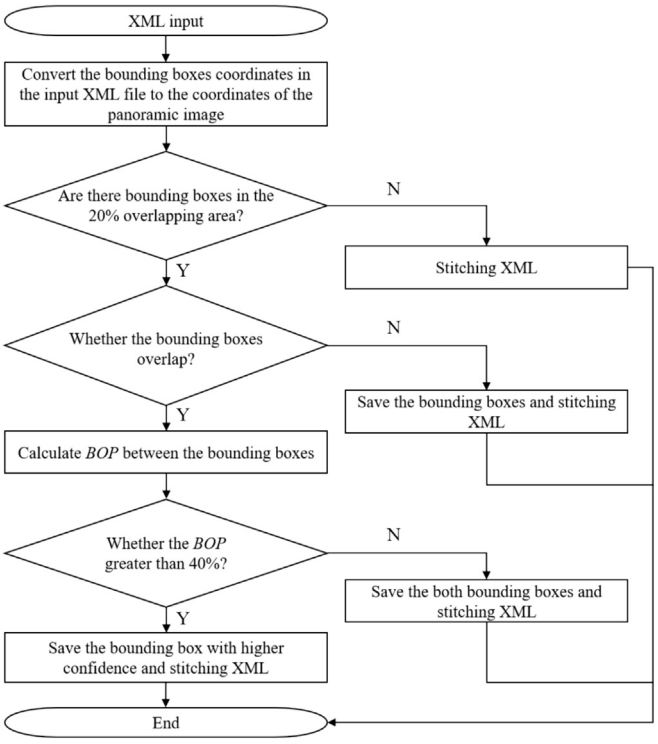


Fig. 7. The process of XML file stitching. The rectangular boxes represent the processing steps, and the diamonds represent the judgment steps.

The process requires the filtering of overlapping bounding boxes, the steps of which are illustrated in Fig. 7. In the 20 % overlapping area, there may be cases where the same target appears in two sub-images. To address this issue, the *Bounding box Overlap Percentage* (*BOP*) is proposed for filtering bounding boxes. The *BOP* is calculated from Eq. (1), and the *Bounding Box A* and the *Bounding Box B* represent the coordinate areas of the two input bounding boxes.

$$BOP = \frac{Bounding\ Box\ A \cap Bounding\ Box\ B}{Bounding\ Box\ A \cup Bounding\ Box\ B} \tag{1}$$

When a new XML file is input, the method first checks whether there are any bounding boxes in the 20 % overlapping area. If no such bounding boxes are found, the information of other bounding boxes in the input XML file is retained in the panoramic XML file. When overlapping area has bounding boxes, it will be checked whether there is an overlap between these bounding boxes. If there is no overlap between the bounding boxes, the XML files are stitched directly. If there is overlap, the *BOP* between the two bounding boxes is calculated. If the *BOP* is greater than 40 %, the two bounding boxes are recognized as the same target. In this case, the bounding box with higher confidence is retained by filtering. If the *BOP* is less than 40 %, both bounding boxes are saved in the panoramic XML file.

2.7. Performance evaluation

The performance of the watermelon detection models was evaluated using *IoU*, *Precision (P)*, *Recall (R)*, *Average Precision (AP)*, and *mAP*.

To measure the degree of overlap between the ground truth box and the predicted box, this paper uses *IoU*, which is calculated as shown in Eq. (2). All samples are classified into four categories according to the combinations of the true and predicted class: *True Positive (TP)*, *False Positive (FP)*, *True Negative (TN)* and *False Negative (FN)*. The *P* and *R* of the detection network are defined in Eq. (3) and Eq. (4) according to the classification of the samples. *AP* is an indicator that reflects the global performance of the network, which is calculated by *P* and *R* in Eq. (5).

$$IoU = \frac{Prediction \cap Ground\ truth}{Prediction \cup Ground\ truth} \tag{2}$$

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

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

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

The *mAP* is an average of *AP* scores across multiple categories. Different *mAPs* are obtained when different *IoU* thresholds are employed. The *mAP*@0.5 refers to the *mAP* value obtained by calculating all images within each category when the threshold of *IoU* is 0.5. The *mAP*@0.5:0.95 refers to the average *mAP* value calculated at 10 thresholds ranging from 0.5 to 0.95 with a step size of 0.05 for *IoU*. For watermelon object detection, the *mAP* is the *AP* of the watermelon category as there is only one category of targets.

In order to evaluate the counting accuracy of the methods reasonably, the performance of watermelon counting methods was assessed by *Counting Accuracy (CA)*. The *CA* of the counting methods is defined

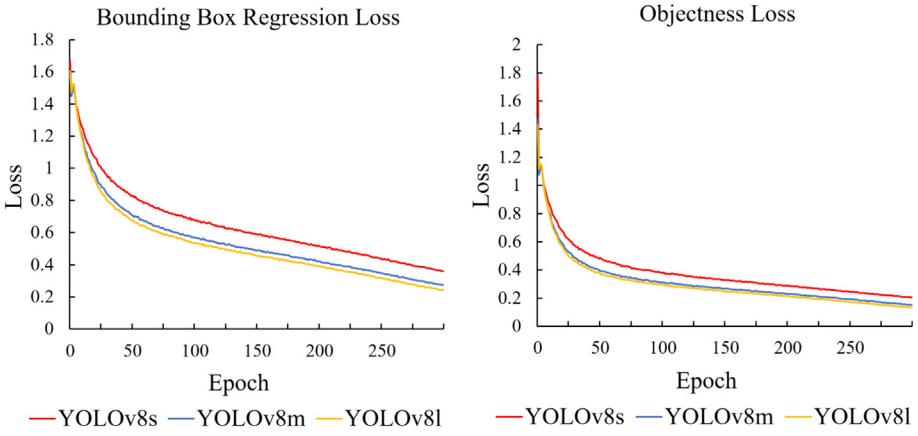


Fig. 8. Comparison of loss curves of watermelon object detection models. Bounding Box Regression loss represents the error between the predicted box and the ground truth box. Objectness loss indicates the confidence of models that a target actually exists in a given region.

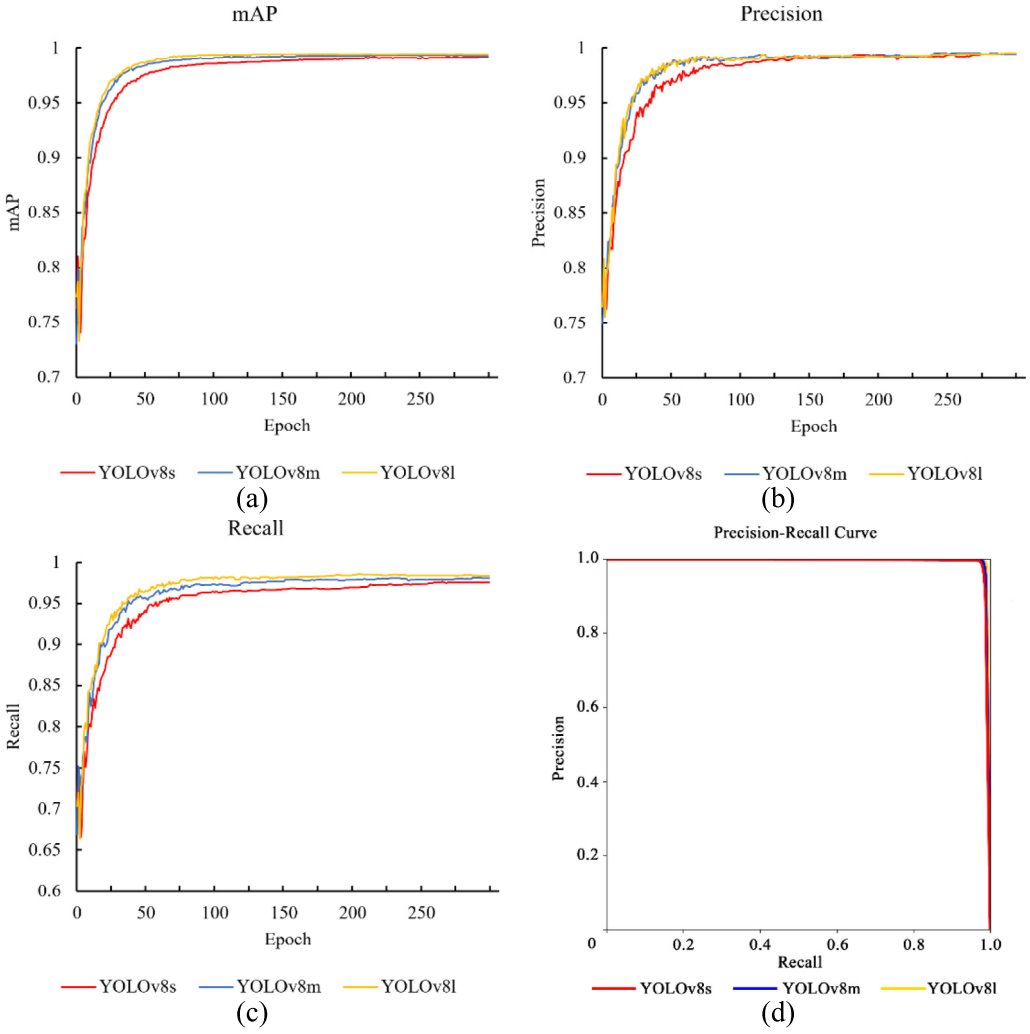


Fig. 9. Comparison of mAP curves, Precision, Recall and Precision-Recall curves of watermelon object detection models. (a) mAP curves, (b) Precision curves, (c) Recall curves, (d) Precision-Recall curves.

in Eq. (6) based on *Real Quantity* (*RQ*) and *Counting Number* (*CN*). The *RQ* represents the true quantity in videos, while the *CN* reflects the quantity gained through the counting methods.

$$CA = 1 - \left|\frac{CN - RQ}{RQ}\right| \tag{6}$$

3. Results and discussion

In this section, the *P*, *R*, *AP* and *mAP* of YOLOv8s, YOLOv8m and YOLOv8l were calculated for evaluation of watermelon detection. Then, the PSOPDCM and the VTDCM were evaluated.

3.1. Performance of object detection models

3.1.1. Training evaluation

Training of the watermelon detection models was completed using predefined parameters based on the watermelon dataset. Throughout the training process, the loss curves demonstrated a gradual and consistent decrease, eventually converging to low values, signifying stable training processes (Fig. 8). From the training results of the YOLOv8s, YOLOv8m and YOLOv8l models (Fig. 9), the accuracy of all three models is at a high level with little difference.

3.1.2. Model evaluation

The three watermelon detection models based on YOLOv8s, YOLOv8m and YOLOv8l were evaluated on the test dataset (Table 1). The watermelon detection model based on YOLOv8s performs well in watermelon detection, with *P* values exceeding 0.99 and *R* values exceeding 0.97, *mAP*@0.5 reaching 0.992 and *mAP*@0.5:0.95 reaching 0.947. Although the performance indicators of the models based on YOLOv8m and YOLOv8l are slightly higher, the model based on YOLOv8s has only 28.6 FLOPs and 11.2 M Params. Due to the focus on large-scale object detection in this task, a large number of images need to be inputted at once. The model size and detection speed are chosen as the final criteria. YOLOv8s stands out as a watermelon object detection model with its small size and fast speed.

With the watermelon detection model based on YOLOv8s, each sub-image is detected and corresponding annotation files are generated (Fig. 10a). The annotation files of all sub-images are then stitched together to generate the panoramic image detecting result (Fig. 10b). The watermelon detection model based on YOLOv8s has achieved promising results in the detection of panoramic image.

Table 1
Comparison of testing effects of watermelon object detection models.

Models	*P*	*R*	*mAP* @0.5	*mAP* @0.5:0.95	FLOPs	Params (M)	Model size (MB)	Speed (ms)
YOLOv8s	0.994	0.974	0.992	0.947	28.6	11.2	21.5	137.0
YOLOv8m	0.994	0.980	0.992	0.972	78.9	25.9	49.7	253.2
YOLOv8l	0.994	0.982	0.994	0.980	165.2	43.7	83.7	405.8

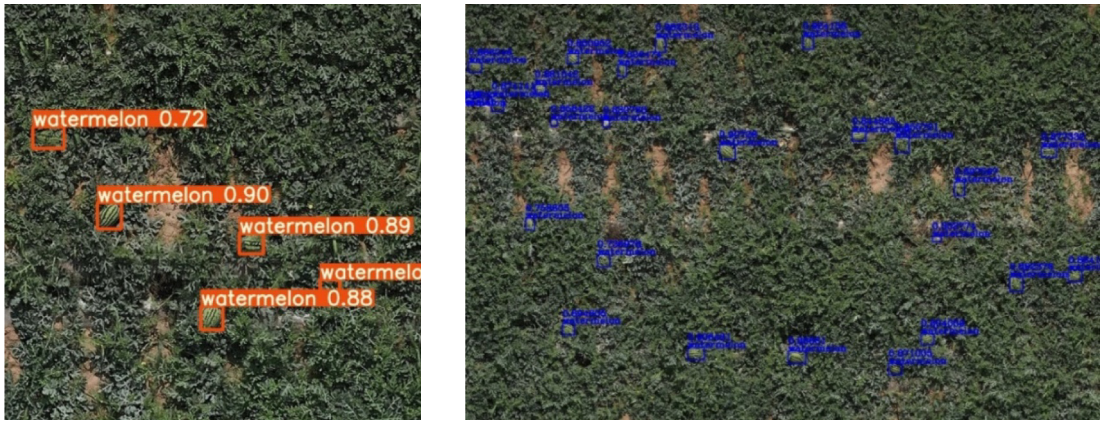


Fig. 10. The detection results of sub-images and panoramic image. The detection results were obtained by drawing boxes on the image based on the XML files. Each target watermelon is enclosed by a bounding box, and associated text above shows the recognized category and its respective confidence rate.

3.2. Performance of the counting algorithms

The accuracy of the two counting methods were tested on videos B and C using the *CA* as the indicator. The *RQ* of watermelons in video B and C are 637 and 502, respectively. For the VTDCM, the *CN* in video B is 1478, and the *CN* in video C is 1232 (Table 2). The count results obtained are highly biased due to false and missed detections in the VTDCM (Fig. 11). For the PSOPDCM, the *CN* in video B is 623, with the *CA* of 97.80 %. The *CN* in video C is 485, with the *CA* of 96.61 % (Table 2). The PSOPDCM demonstrated an average counting accuracy exceeding 96 % in the two watermelon test videos.

Although the VTDCM directly detects and tracks watermelon targets in UAV videos, which is simpler compared to the PSOPDCM, the results are not satisfactory. This is mainly due to the high resolution of the input videos, which shifts the watermelon object detection challenge to detecting small target objects against a large background. Consequently, this leads to changes in watermelon ID and the loss of target tracking (Fig. 11a, b, and c). At the same time, the VTDCM relies on the detection accuracy of watermelon targets between consecutive frames, which leads to unsatisfactory results due to the similar characteristics of watermelons. The watermelons grow in open field with watermelon vines arranged in rows, and the watermelons are easy to be blocked by vines and leaves (Fig. 11d). Furthermore, the similarity in colour and texture between watermelon targets and backgrounds leads to errors in tracking.

In the PSOPDCM, panoramic images are used as input, which allows an overall evaluation of the field. The overlap partitioning detection effectively improves the detection success rate in large scenes, increases the accuracy of watermelon counting, and achieves ideal results in experimental data videos (Figs. 11 and 12).

Although the PSOPDCM achieves higher accuracy than the VTDCM, field watermelon detection inevitably encounters issues such as missed and false detections. These problems may be caused by occlusion of vines or leaves in the image, or image distortion or blurring during data collection and image stitching. A multi-view data acquisition approach can be used to reduce the interference. At the same time, this method can achieve higher efficiency by further optimize the stitching strategy of the annotated files. The current stitching method being used can result in multiple overlapping areas inputs. When stitching, it is necessary to filter the bounding boxes in the four overlapping areas A, B, C and D, as shown in Fig. 13. The method proposed in this paper needs to verify each region separately, which can be simplified by inputting the four overlapping areas as a whole for verification.

Table 2
Performance evaluation of counting algorithms.

Method	Test video	*RQ*	*CN*	*CA* (%)
VTDCM	B	637	1478	–
	C	502	1232	–
PSOPDCM	B	637	623	97.80
	C	502	485	96.61

3.3. Discussion

An algorithmic pipeline of UAV videos was proposed for the detection and counting of watermelons. To achieve target counting in the field, it is necessary to distinguish each target. Existing field target detection methods mainly focus on a set of images, using traditional image processing techniques or deep learning-based methods (Ho et al., 2019; Zhao et al., 2017). These methods use a set of small scene images as input, resulting in multiple detections of the same target, which is not conducive to counting work. Some studies of counting for large agricultural scenes used the VTDCM (Khokher et al., 2023). However, due to the complex background of watermelon targets, the

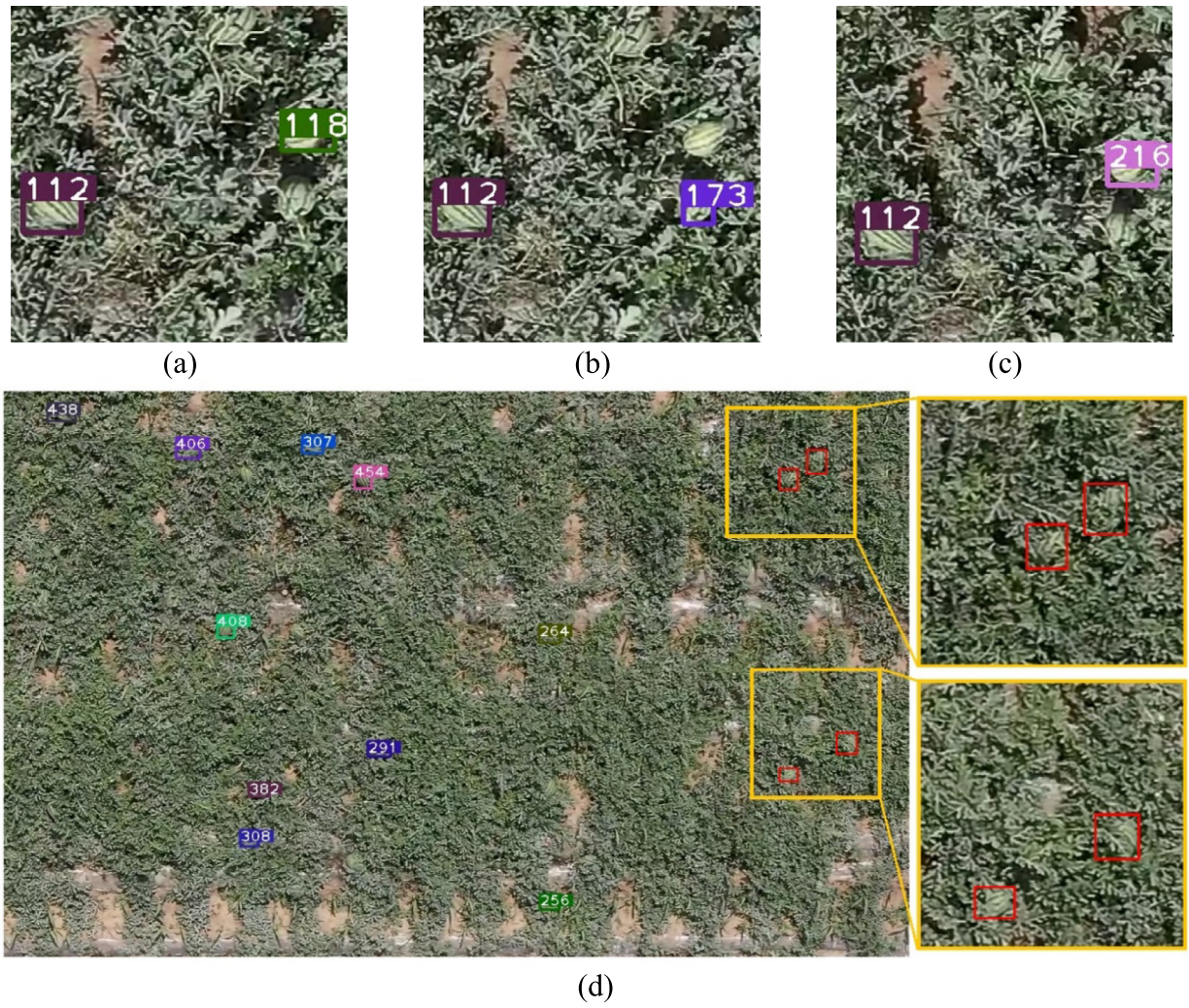


Fig. 11. False and missed detections in the VTDCM. (a), (b) and (c) show the false detection such as the same target object is assigned different IDs. And (d) shows the overall scene of watermelon field during tracking process, with a magnified view on the right showing the missed detection of the watermelon target caused by vines occlusion (red box).

applicability of tracking technique to watermelon is limited. The panorama stitching method can convert videos into panoramic images that contain the entire scene, allowing for an overall assessment of the field (Mekhalfi et al., 2020). Using panoramic images as input can avoid the issue of redundant detection when using a set of images as input. And an overlap partition detection method is proposed to improve the detection accuracy in panoramic images. By utilizing the PSOPDCM with YOLOv8, we achieved accurate detection and counting results in watermelon (Table 3).

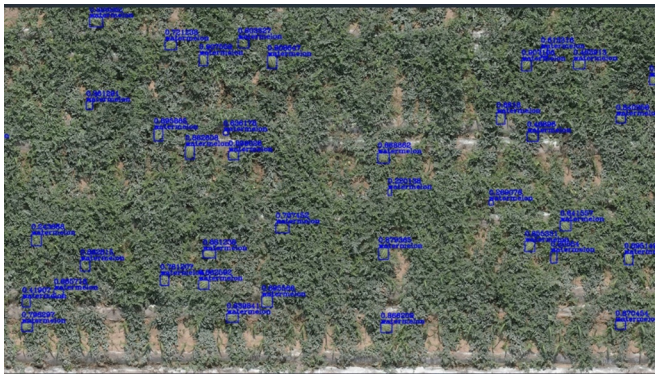


Fig. 12. Detection result of the PSOPDCM.

4. Conclusions

By combining a deep learning-based object detection model with the PSOPDCM, an algorithmic pipeline of UAV videos for detection and counting of watermelons was developed. The algorithmic pipeline composed of three main stages: panoramic image generation, watermelon detection and watermelon counting. Input of panoramic images enables a holistic assessment of field, allowing for overall yield estimation of the watermelon field. Watermelon detection model was trained based on YOLOv8s by transfer learning, which reached a detection accuracy of 99.20 %. The PSOPDCM has achieved an accuracy exceeding 96.61 %, reducing duplications compared to the VTDCM. Despite the achievements of accuracy in detecting and counting, there is still potential for improving the efficiency of the method. Future refinements could involve optimizing the annotated files stitching strategy for speed enhancement. And watermelon field data of different growth stages could be collected to train a multi-class object detection model, gaining a more comprehensive understanding of watermelon cultivation dynamics.

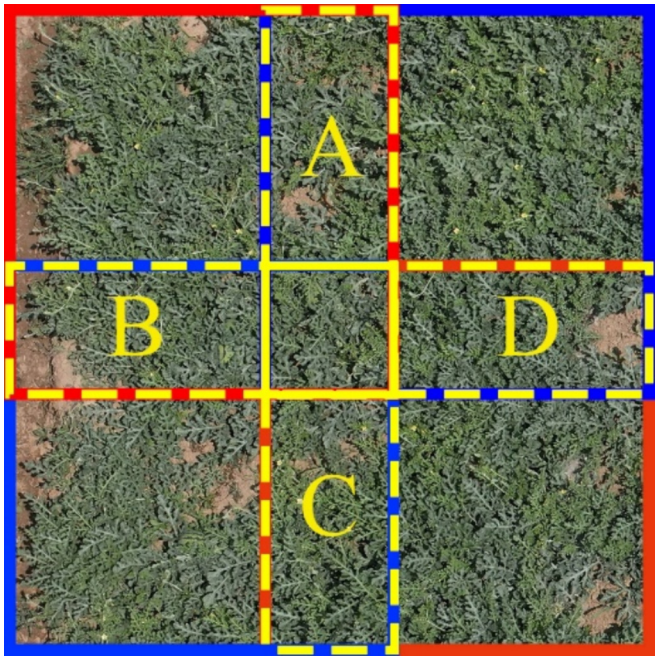


Fig. 13. Stitching strategy of annotated files. The red and blue solid line boxes in the figure represent the input sub-images, and the yellow dashed line boxes A, B, C and D represent the overlapping area.

Table 3
Comparison with other researches.

Author	Target	Detecting accuracy (%)	Counting accuracy (%)
Ours	Watermelon	99.20	96.61
Zhao et al., 2017	Melon	79.29	–
Ho et al., 2019	Watermelon	99.00	–
Khokher et al., 2023	Inflorescence	80.00	88.97
Mekhalfi et al., 2020	Kiwifruit	85.00	85.00

CRediT authorship contribution statement

Liguo Jiang: Conceptualization, Data curation, Methodology, Writing – original draft. **Hanhui Jiang:** Methodology, Validation, Writing – original draft. **Xudong Jing:** Conceptualization, Methodology, Software. **Haojie Dang:** Methodology, Software. **Rui Li:** Data curation, Project administration, Writing – review & editing. **Jinyong Chen:** Funding acquisition, Resources. **Yaqoob Majeed:** Conceptualization, Writing – review & editing. **Ramesh Sahni:** Investigation. **Longsheng Fu:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China (32371999); Science and Technology Program of Yulin City, China (2023-CXY-183); Open Project of Key Laboratory of Agricultural Equipment for Hilly and Mountainous Areas in Southeastern China (Co-construction by Ministry and Province), Ministry of Agriculture and Rural Affairs, China (QSKF2023002); National Foreign Expert Project, Ministry of Science and Technology, China (QN2022172006L, DL2022172003L).

References

Bewley, A., Ge, Z., Ott, L., Ramos, F., Upcroft, B., 2016. Simple online and realtime tracking. 2016 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 3464–3468. https://doi.org/10.1109/ICIP.2016.7533003.

Cui, Z., Zhou, P., Wang, X., Zhang, Z., Li, Y., Li, H., Zhang, Y., 2023. A novel geo-localization method for UAV and satellite images using cross-view consistent attention. Remote Sens. 15, 4667. https://doi.org/10.3390/rs15194667.

Dhanya, V.G., Subeesh, A., Kushwaha, N.L., Vishwakarma, D.K., Nagesh Kumar, T., Ritika, G., Singh, A.N., 2022. Deep learning based computer vision approaches for smart agricultural applications. Artif. Intell. Agric. 6, 211–229. https://doi.org/10.1016/j.aiia.2022.09.007.

Feng, H., Tao, H., Fan, Y., Liu, Y., Li, Z., Yang, G., Zhao, C., 2022. Comparison of winter wheat yield estimation based on near-surface hyperspectral and UAV hyperspectral remote sensing data. Remote Sens. 14, 4158. https://doi.org/10.3390/rs14174158.

Fu, L., Feng, Y., Wu, J., Liu, Z., Gao, F., Majeed, Y., Al-Mallahi, A., Zhang, Q., Li, R., Cui, Y., 2021. Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model. Precis. Agric. 22, 754–776. https://doi.org/10.1007/s11119-020-09754-y.

Gao, F., Fang, W., Sun, X., Wu, Z., Zhao, G., Li, G., Li, R., Fu, L., Zhang, Q., 2022. A novel apple fruit detection and counting methodology based on deep learning and trunk tracking in modern orchard. Comput. Electron. Agric. 197, 107000. https://doi.org/10.1016/j.compag.2022.107000.

Gao, J., Liao, W., Nuyttens, D., Lootens, P., Alexandersson, E., Pieters, J., 2023. Cross-domain transfer learning for weed segmentation and mapping in precision farming using ground and UAV images. Expert Syst. Appl. 246, 122980. https://doi.org/10.1016/j.eswa.2023.122980.

Guo, Y., Aggrey, S.E., Yang, X., Oladeinde, A., Qiao, Y., Chai, L., 2023. Detecting broiler chickens on litter floor with the YOLOv5-CBAM deep learning model. Artif. Intell. Agric. 9, 36–45. https://doi.org/10.1016/j.aiia.2023.08.002.

Ho, M., Lin, Y., Hsu, H., Sun, T., 2019. An efficient recognition method for watermelon using faster R-CNN with post-processing. 2019 8th International Conference on Innovation, Communication and Engineering (ICICE). IEEE, pp. 86–89. https://doi.org/10.1109/ICICE49024.2019.9117374.

Hsu, H., Ho, M., Sun, T., 2019. Watermelon recognition and yield estimation using mathematical morphology and naïve bayesian classifier from air borne image of watermelon field. 2019 8th International Conference on Innovation, Communication and Engineering (ICICE). IEEE, pp. 82–85. https://doi.org/10.1109/ICICE49024.2019.9117524.

Jiang, H., Sun, X., Fang, W., Fu, L., Li, R., Cheein, F.A., Majeed, Y., 2023. Thin wire segmentation and reconstruction based on a novel image overlap-partitioning and stitching algorithm in apple fruiting wall architecture for robotic picking. Comput. Electron. Agric. 209, 107840. https://doi.org/10.1016/j.compag.2023.107840.

Jiao, Y., Luo, R., Li, Q., Deng, X., Yin, X., Ruan, C., Jia, W., 2020. Detection and localization of overlapped fruits application in an apple harvesting robot. Electron 9, 1–14. https://doi.org/10.3390/electronics9061023.

Kalantar, A., Edan, Y., Gur, A., Klapp, I., 2020. A deep learning system for single and overall weight estimation of melons using unmanned aerial vehicle images. Comput. Electron. Agric. 178, 105748. https://doi.org/10.1016/j.compag.2020.105748.

Khokher, M.R., Liao, Q., Smith, A.L., Sun, C., Mackenzie, D., Thomas, M.R., Wang, D., Edwards, E.J., 2023. Early yield estimation in viticulture based on grapevine inflorescence detection and counting in videos. IEEE Acc. 11. https://doi.org/10.1109/ACCESS.2023.3263238 37790–37808.

Li, G., Fu, L., Gao, C., Fang, W., Zhao, G., Shi, F., Dhupia, J., Zhao, K., Li, R., Cui, Y., 2022. Multi-class detection of kiwifruit flower and its distribution identification in orchard based on YOLOv5l and euclidean distance. Comput. Electron. Agric. 201, 107342. https://doi.org/10.1016/j.compag.2022.107342.

Liao, Z., Dai, Y., Wang, H., Ketterings, Q.M., Lu, J., Zhang, F., Li, Z., Fan, J., 2023. A double-layer model for improving the estimation of wheat canopy nitrogen content from unmanned aerial vehicle multispectral imagery. J. Integr. Agric. 22, 2248–2270. https://doi.org/10.1016/j.jia.2023.02.022.

Lin, T., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. IEEE Conf. Comput. Vis. Pattern Recognit., 936–944. https://doi.org/10.1109/CVPR.2017.106.

Liu, S., Qi, L., Qin, H., Shi, J., Jia, J., 2018. Path aggregation network for instance segmentation. IEEE Conf. Comput. Vis. Pattern Recognit. 8759–8768.

Liu, X., Chen, S., Liu, C., Shivakumar, S.S., Das, J., Taylor, C.J., Underwood, J., Kumar, V., 2019. Monocular camera based fruit counting and mapping with semantic data association. IEEE Robot. Autom. Lett. 4, 2296–2303. https://doi.org/10.1109/LRA.2019.2901987.

Liu, L., Li, P., Wang, D., Zhu, S., 2024. A wind turbine damage detection algorithm designed based on YOLOv8. Appl. Soft Comput. 154, 111364. https://doi.org/10.1016/j.asoc.2024.111364.

Luna, I., Lobo, A., 2016. Mapping crop planting quality in sugarcane from UAV imagery: a pilot study in Nicaragua. Remote Sens. 8, 500. https://doi.org/10.3390/rs8060500.

Mekhalfi, M.L., Nicolò, C., Ianniello, I., Calamita, F., Goller, R., Barazzuol, M., Melgani, F., 2020. Vision system for automatic on-tree kiwifruit counting and yield estimation. Sensors 20, 1–18. https://doi.org/10.3390/s20154214.

Milioto, A., Lottes, P., Stachniss, C., 2017. Real-time blob-wise sugar beets VS weeds classification for monitoring fields using convolutional neural networks. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci., 41–48. https://doi.org/10.5194/isprs-annals-IV-2-W3-41-2017.

Miranda, J., Ponce, P., Molina, A., Wright, P., 2019. Sensing, smart and sustainable technologies for Agri-food 4.0. Comput. Ind. 108, 21–36. https://doi.org/10.1016/j.compind.2019.02.002.

Nan, Y., Zhang, H., Zeng, Y., Zheng, J., Ge, Y., 2023. Intelligent detection of multi-class pitaya fruits in target picking row based on WGB-YOLO network. Comput. Electron. Agric. 208, 107780. https://doi.org/10.1016/j.compag.2023.107780.

Nepal, U., Eslamiat, H., 2022. Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs. Sensors 22. https://doi.org/10.3390/s22020464.

Song, C., Zhang, F., Li, J., Xie, J., Yang, C., Zhou, H., Zhang, J., 2023. Detection of maize tassels for UAV remote sensing image with an improved YOLOX model. J. Integr. Agric. 22, 1671–1683. https://doi.org/10.1016/j.jia.2022.09.021.

Tang, Y., Zhou, H., Wang, H., Zhang, Y., 2023. Fruit detection and positioning technology for a Camellia oleifera C. Abel orchard based on improved YOLOv4-tiny model and binocular stereo vision. Expert Syst. Appl. 211, 118573. https://doi.org/10.1016/j.eswa.2022.118573.

Tripathi, A., Gupta, M.K., Srivastava, C., Dixit, P., Pandey, S.K., 2022. Object detection using YOLO: a survey. 2022 5th International Conference on Contemporary Computing and Informatics (IC3I). IEEE, pp. 747–752. https://doi.org/10.1109/IC3I56241.2022.10073281.

Van Etten, A., 2018. You Only Look Twice: Rapid Multi-Scale Object Detection in Satellite Imagery. Comput. Vis. Pattern Recognit, IEEE Conf. https://doi.org/10.48550/arXiv.1805.09512.

Velusamy, P., Rajendran, S., Mahendran, R.K., Naseer, S., Shafiq, M., Choi, J.-G., 2021. Unmanned aerial vehicles (UAV) in precision agriculture: applications and challenges. Energies 15, 217. https://doi.org/10.3390/en15010217.

Wojke, N., Bewley, A., Paulus, D., 2017. Simple online and realtime tracking with a deep association metric. 2017 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 3645–3649. https://doi.org/10.1109/ICIP.2017.8296962.

Xiao, D., Pan, Y., Feng, J., Yin, J., Liu, Y., He, L., 2022. Remote sensing detection algorithm for apple fire blight based on UAV multispectral image. Comput. Electron. Agric. 199, 107137. https://doi.org/10.1016/j.compag.2022.107137.

Zhao, T., Wang, Z., Yang, Q., Chen, Y., 2017. Melon yield prediction using small unmanned aerial vehicles. Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping II. https://doi.org/10.1117/12.2262412 1021808.