Research article

Efficient defective cocoon recognition based on vision data for intelligent picking

Jun Chen[1], Xueqiang Guo[2], Taohong Zhang[2,3] and Han Zheng[1,*]

[1] Key Laboratory of AI and Information Processing (Hechi University), Education Department of Guangxi Zhuang Autonomous Region, Hechi 546300, China
[2] Department of Computer, School of Computer and Communication Engineering, University of Science and Technology Beijing (USTB), Beijing, China
[3] Beijing Key Laboratory of Knowledge Engineering for Materials Science, Beijing, China

* **Correspondence:** Email: 05057@hcnu.edu.cn.

Abstract: Cocoons have a direct impact on the quality of raw silk. Mulberry cocoons must be screened before silk reeling can begin in order to improve the quality of raw silk. For the silk product sector, the cocoons' level of categorization and sorting is crucial. Nonetheless, the majority of mulberry cocoon production facilities in use today choose the cocoons by hand. The accuracy and efficiency of mulberry cocoon plucking can be significantly improved by automatic methods. To increase efficiency, mulberry cocoons must be sorted automatically and intelligently using machine vision. We proposed an effective detection technique based on vision and terahertz spectrum characteristics data for distinguishing defective cocoons, including common and thin shelled defective cocoons. Each mulberry cocoon with a defect had its spatial coordinate and deflection angle computed so that grippers could grasp it. With 3762 photos in our dataset, our approach could detect mAP values up to 99.25% of the time. Furthermore, the GFLOPS of our suggested model was only 8.9 G, and its Parameters were only 5.3 M, making it appropriate for use in real-world application scenarios.

Keywords: deep learning; intelligent picking; terahertz detection; defective cocoons

1. Introduction

People want to use silk products of a higher caliber than they did in the past, and as society

advances and living standards rise, stringent control over the reeling process's production technology is necessary. Approximately 50,000 tons of raw silk are produced worldwide each year. Manually identifying the different sorts of cocoons is the primary method of cocoon picking. The hand picking process involves not only sluggish inspection and low efficiency, but also subjective consciousness among cocoon workers when it comes to picking quality. Young workers are unwilling to perform this tedious work. With the development of data-driven artificial intelligence and machine vision, intelligent equipment for cocoon recognition and automatic picking is objective and stable, and it can reduce the labor intensity of workers, save labor costs, and greatly improve labor productivity.

The defective cocoons should be filtered out first in the silk reeling production process once the cocoons have been collected. because it has a significant impact on the saw silk quality. Defective cocoons come in various forms, such as cotton, pointed, thin-shelled, deformed, and so on. Employees manually select mulberry cocoons in facilities that produce them. It is predicated on the capacity of the human eye to perceive and assess the exterior traits of cocoons. The skill of those who collect cocoons has a significant impact on how efficiently it can be picked. Furthermore, it is impossible to consistently evaluate the type of cocoons because of the underlying causes of subjective consciousness and emotions as well as the variations in each person's physique. In addition, selecting cocoons by hand requires a lot of labor, and workers must always maintain a fixed posture and a high level of mental focus. We propose a data-driven intelligent method for robot grippers to identify and pick mulberry cocoons, based on deep learning and image processing technologies. It significantly increases labor productivity while lowering worker labor intensity, saving labor costs, and producing objective, stable, and highly accurate picking outcomes.

Artificial intelligence technology has been slowly applied to the silkworm industry, for example, gender recognition of cocoons [1–4]. Automatic picking reduces manual operations and improves the efficiency and quality of picking operations. Automatic picking has been successfully applied to food picking [5–7], industrial waste picking and recycling [8,9], and e-commerce warehouses [10]. In [11], the use of machine learning techniques on silk seeds, including random forests, support vector machines, and deep learning, is examined. It focuses mostly on intelligent decision-making and automation of production processes. An image-based grading system was proposed in [12]. To distinguish different types of cocoons, form and color attributes are retrieved from photos of silkworm pupae based on three colors: RGB, HSV, and L*a*b. In [13], a convolutional neural network-based approach to identify the sex and species of silkworm cocoons is described. This method can also be used for other purposes, such as breeding studies and the classification of cocoon quality. Training on a large number of near-infrared spectral pictures allows for excellent identification of the gender and species of cocoons. In [14], it is suggested to employ deep learning methods to extract useful information from terahertz imaging. It has a high degree of prediction and detection accuracy for classifying cocoons.

This paper's major contributions are as follows:

(1) The network structure of EV-YOLO X is created for recognition in the cocoon picking job, and a deep learning technique based on computer vision is suggested;

(2) The deflection angle of the cocoon by an external rectangle was calculated using the Gaussian fuzzy processing method in this paper, which was based on the OpenCV image processing method;

(3) Optical penetration technology and terahertz detection of spectral properties were used to distinguish thin-shelled cocoons from normal cocoons, further improving the overall quality of cocoon picking.

2. Intelligent recognition of defective cocoons

Defective cocoons refer to cocoons that cannot be reeled or are difficult to reel, which are mostly malformed cocoons, thin-shelled cocoons, cotton cocoons, pointed cocoons, macular cocoons, etc., as shown in Figure 1. These defective cocoons downgrade the overall raw silk quality and need to be picked out from good cocoons. Most defective cocoons (malformed cocoons, cotton cocoons, pointed cocoons, and macular cocoons) can be identified by their appearances, which can be intelligently recognized by the machine vision method after establishing a deep learning-based object detection model. However, this object detection algorithm cannot achieve high accuracy and effective differentiation for thin-shelled cocoons. It is because there is no obvious feature difference between thin-shelled cocoons and normal good cocoons by visual imaging. To deal with this identification, a light penetration imaging method is proposed for distinguishing thin-shelled cocoons from good cocoons. In addition, silk, a natural organic polymer material, exhibits strong absorption and dispersion properties in the terahertz band wave, which makes it possible to apply terahertz detecting technology to distinguish thin-shelled cocoons and double palace cocoons from flawless cocoons to screen high-quality cocoons, which further improves the identification and classification accuracy.
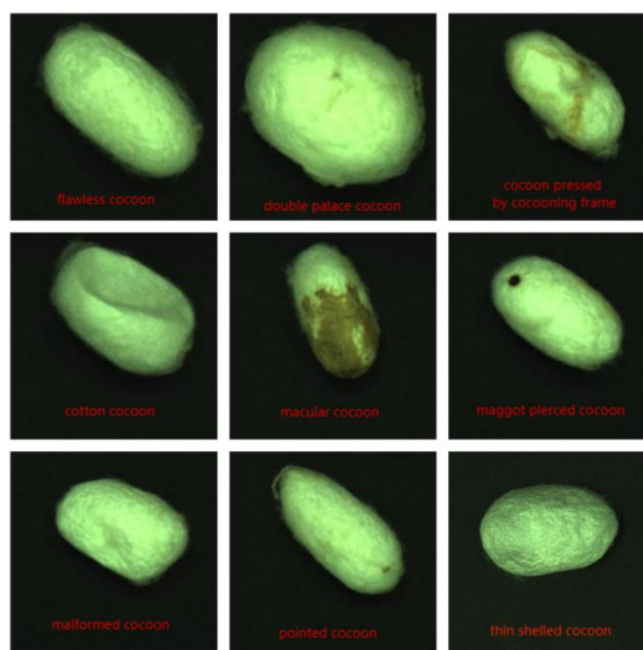

Figure 1. Classification of mulberry defective cocoons.

2.1. Intelligent recognition of common defective cocoons

For defective cocoons, except for thin-shelled cocoons, which are denoted as common defective cocoons, there is an obvious optical distinction between flawless and double palace cocoons. A deep learning object detection model based on the cocoon data set is established here for identifying and classifying cocoons. The detection technology roadmap is shown in Figure 2. The deep learning network for the identification and classification of different cocoons is designed as shown in Figure 3.
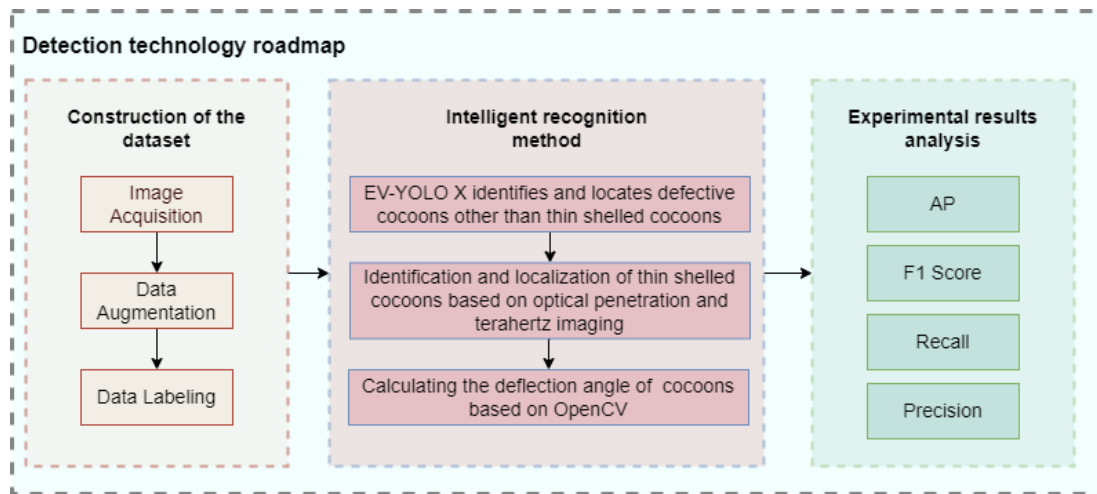
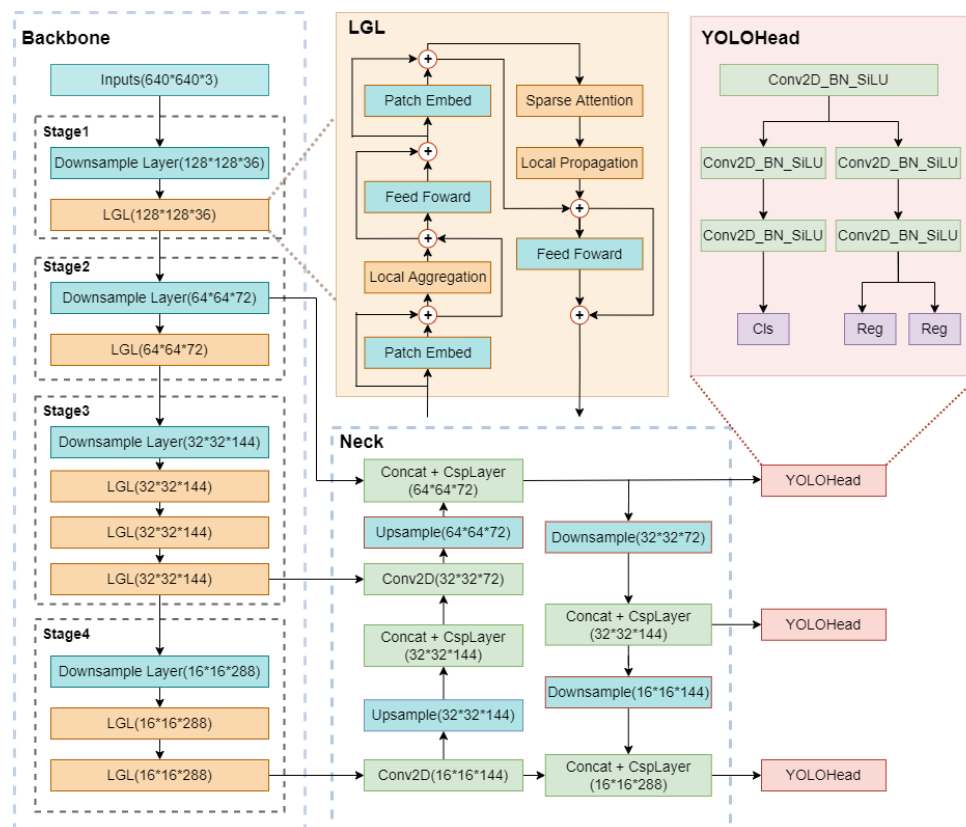**Figure 2.** Detection technology roadmap.

Figure 3. EV-YOLO X network structure.

Algorithm 1 summarizes the main processing of our proposed cocoon detection, which includes the steps of normalization, local aggregation, feed-forward network processing, etc., and outputs the cocoon features after comprehensive processing.

2.1.1. Deep learning based detection model

Unlike image classification, which focuses on the overall information of the whole image, object detection focuses on a specific object, and the category information and location information of this object are obtained through the object detection algorithm. Compared with image classification, object detection gives an understanding of the foreground and of the image, separates the object of interest from the background, and determines the category and location of this object. Deep learning-based object detection methods include two major categories: Two-stage methods and single-stage methods. Two-stage refers to the way to achieve detection with two main processes: the first step is to extract object regions, and the second step is to classify and recognize the regions based on a convolutional neural network. Therefore, the two-stage detection algorithms are based on candidate regions. The two-stage detection algorithms are the pioneers of deep learning detection algorithms. Common two-stage algorithms include R-CNN [15], Fast R-CNN [16], Faster R-CNN [17], SPPNet [18], etc. Their advantages are a low recognition error rate and a low missed recognition rate, but the speed is so slow that they cannot meet real-time detection scenarios. The single-stage object detection algorithms do not need to generate candidate frames and directly convert the object localization task to regression task processing, i.e., they directly generate the class probabilities and the corresponding position coordinate values for each object. Single-stage based algorithms require only a single detection to obtain the final detection result. Common single-stage object detection algorithms include SSD [19], RetinaNet [20], YOLO series [21–25], etc. Compared with the two-stage object detection algorithms, the single-stage object detection algorithm has faster detection speed, realizes real-time recognition with a high frame rate, and is suitable for industrial applications.

In the cocoon picking task, a network architecture of EV-YOLO X is designed for object detection, as shown in Figure 3. The network architecture is divided into four parts, namely the input, backbone, neck, and head. After inputting the mulberry cocoon image into the model, the modified EdgeViT [26] network structure is used as the backbone to extract the image features. The neck part is based on the structure of PAFPN [27] to efficiently fuse feature maps at different layers. The head part is for result prediction, and the main highlight of the head is the use of a decoupled detection head. Classification and regression conflicts are a common problem in object detection tasks, so decoupling the classification and localization heads has been widely used in single-stage and two-stage object detection tasks. However, although the backbone and feature pyramid networks of the Yolo series have evolved, their detection heads are in a coupled manner. Here, the EV-YOLO X decouples the detection heads with a lightweight decoupling head, which significantly improves the convergence speed of the model. In addition, EV-YOLO X does not use anchor boxes. The benefits are as follows: First, no IoU calculation is involved, and the number of prediction frames generated is also greatly reduced compared to the anchor-based approach, which will reduce the computational effort of the model. Then, the anchor-free method produces only 1/3 of the prediction frames of the anchor-based method. Since most of the prediction frames are negative samples, the anchor-free method can reduce the number of negative samples, which further alleviates the problem of positive and negative sample imbalance. In addition, the anchor-free method avoids the adjustment of anchors. The scale of the anchor box of the anchor-based method is a hyperparameter. Setting different hyperparameters can affect the performance of the model. The anchor-free method avoids this drawback.

To reduce the training time and model hyperparameters, the proposed EV-YOLO X takes SimOTA to dynamically match positive samples, calculates pairwise matching, i.e., the cost relationship between each ground truth and each feature point, and selects the top k predictions with

the lowest cost in the fixed central region as positive samples. Finally, the corresponding grids of these positive predictions are divided into positive grids, and the remaining grids are divided into negative grids. The calculation of the loss function is a comparison of the predicted results of the grids with the real results. Consistent with the prediction results of the network, the loss of the network also has three components, where two use the binary cross-entropy loss (BCE Loss), while one uses IoU Loss [28]. In addition, they calculate only the loss of positive samples, while calculates both positive and negative samples loss. The loss function of EV-YOLO X is shown in Eq (1):

$$\text{Loss} = \frac{L_{cls} + \lambda L_{reg} + L_{obj}}{N_{pos}} \tag{1}$$

where L_{cls} stands for classification loss, L_{reg} stands for localization loss, L_{obj} stands for IoU loss, λ stands for the balance coefficient of localization loss, λ is taken as 5.0 in this paper. λ is an adjustable parameter used to balance the two losses. and stands for the number of anchor points classified as positive samples.

2.1.2. Deflection angle recognition

The mulberry cocoon picking operation requires the use of industrial robots equipped with tooling fixtures to pick up the defective cocoons, and the tooling fixtures need to know the deflection angle of the cocoons for cocoon picking. In this paper, the calculation of the deflection angle of mulberry cocoons is based on the OpenCV image processing method. The mulberry cocoon image is first binarized, and then edge detection is performed. After that, the external rectangle and the minimum area of the external rectangle are calculated. For dealing with the surface of defective cocoons, an easily mis-detected problem, this paper uses Gaussian blur processing. In order to obtain the real external rectangle of the mulberry cocoon, it is necessary to screen out the tiny defective external rectangle according to the area, and this paper takes 15,000 as the boundary to keep the rectangle larger than this boundary. As shown in Figure 4(a), the blue box is the positive external rectangle of the mulberry cocoon, and the green box is the minimum external rectangle of the mulberry cocoon. Finally, the deflection angle can be calculated according to the external rectangle and the minimum external rectangle, as shown in Figure 4(b), where α is the requested deflection angle.
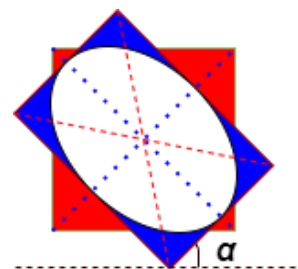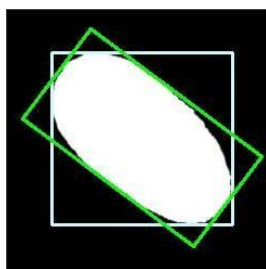
(a)External rectangle and minimum external rectangle   (b) Calculation of the deflection angle

Figure 4. Calculation of deflection angle of mulberry cocoon.

2.1.3. Dataset

Data acquisition is a key step in the construction of the intelligent recognition model for defective cocoons. The quality of data collection determines the accuracy of the model and its robustness. As shown in Figure 1, according to the quality of raw silk mulberry cocoons, they can be divided into double palace cocoons, good cocoons, and defective cocoons. The defective cocoons are divided into pointed cocoons, thin shelled cocoons, maggot-pierced cocoons, cotton cocoons, macular cocoons, cocoons pressed by a cocooning frame, and malformed cocoons. The dataset is constructed by high-definition industrial cameras to take images, as shown in Figure 5.
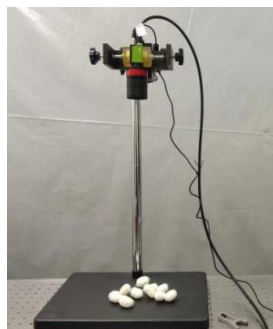

Figure5. Image acquisition device for mulberry cocoons.

A total of 553 images of mulberry cocoons were obtained. To ensure the recognition efficiency and accuracy of the intelligent picking model, data augmentation is done on these 553 images using horizontal flip, vertical flip, rotation of 90 degrees, rotation of 180 degrees, and rotation of 270 degrees. The augmented dataset of 3762 images was obtained, as shown in Table 1. The annotation of the dataset is done under the guidance of professional mulberry cocoon picking workers. The ratio of the total number of training, validation, and test sets is 8:1:1.

Table 1. Details of the data set.

	Amount of data	Total number of cocoons	Defective cocoons	Good cocoons
Before augmentation	553	1548	864	684
After augmentation	3762	9288	5184	4104

2.1.4. Results

In order to measure the quality of the neural network for the detection and recognition of mulberry cocoons, we use precision (P), recall (R), F1 score (F1), and mean average precision (mAP) as evaluation criteria. The calculation formulas are as follows:

$$P = Tp/(Tp + Fp) \qquad (2)$$

$$R = Tp/(Tp + F_N) \qquad (3)$$

$$F1 = 2PR/(P + R) \tag{4}$$

$$mAP = \frac{1}{C}\sum_{K=1}^{C} J(P, R, K) \tag{5}$$

where Tp is the number of correct judgments in the prediction, Fp is the number of incorrect judgments, FN is the number of missed detections, C is the number of cocoon categories, K is the category serial number, and J is the area function of the P-R curve and the coordinate axis.

We put the EV-YOLO X model in the Pytorch framework with a host conFigureuration of an Intel Core i9-9900K CPU @ 3.60 GHz * 16, 64 GB RAM, NVIDIA GeForce RTX 2080Ti * 2, and Ubuntu 18.04 OS. We use Python as the interaction language and CUDA 10.2.89 and cuDNN 7.6.4 for accelerated computing. The preprocessed dataset is then fed into the EV-YOLO X network for training with 100 epochs, weights saved every 10 iterations, a momentum of 0.9, an initial learning rate of 0.001, a decay coefficient of 0.0005, and a batch size of 64.
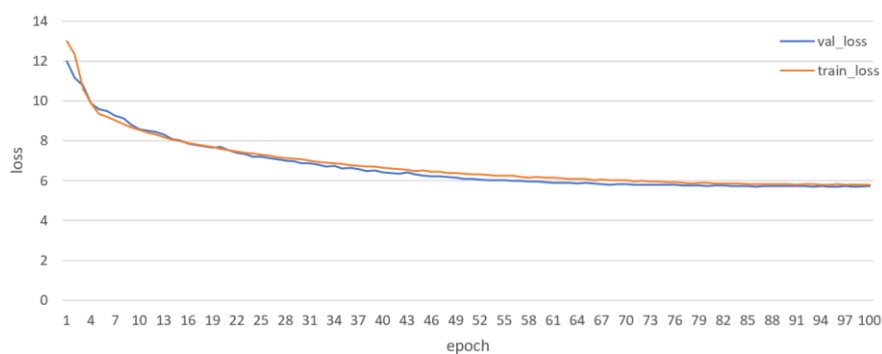

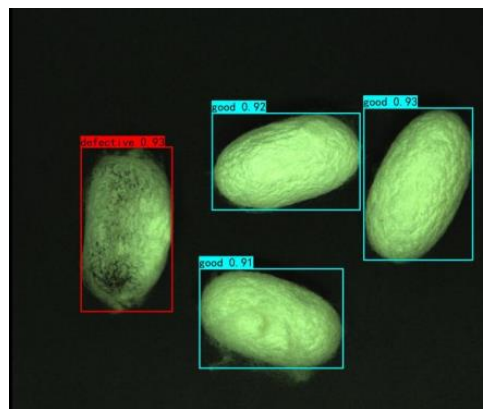**Figure 6.** Loss function curve.

Figure 7. Model detection result.

With enough training for the created mulberry cocoon dataset, EV-YOLO X is able to converge rapidly, as evidenced by the loss function's decrease with increasing epoch and its convergence to a reduced interval range after roughly 10 epochs. The test set's mulberry cocoon recognition result is shown in Figure 7.

The detection results of EV-YOLO X and YOLO X-S [21] for mulberry cocoons are shown in Table 2. It can be seen that EV-YOLO X has good recognition on the constructed mulberry cocoon dataset, where the mAP reaches 99.25% and the F1 score for both defective and flawless cocoons

reaches 98%. In contrast, the model YOLO X-S has an mAP of 99.77% and an F1 score of 98% and 97% for defective and flawless cocoons, respectively. The target detection speed and accuracy of several networks for cocoon detection are shown in Table 3, together with the computational and parametric parameters. Table 3 illustrates that the number of parameters of EV-YOLO X is only 59.05% of YOLO X-S, 14.5% of RetinaNet, 14.2% of YOLOv7-l, 11.3% of YOLOv5-l, and 8.6% of YOLOv3. Based on these comparisons, our suggested technique has the least amount of computational and parametric materials. Of YOLO X-S, only 33.42% of surgeries are performed. The overall performance of the YOLO X-S and the EV-YOLO X proposed in this paper is shown in Figure 8.

Table 2. Detection results on the mulberry cocoon dataset.

Model	Category	Precision	Recall	F1 Score	AP	mAP
YOLO X-S	Defective (0)	98.67%	98.11%	98%	99.81%	99.77%
	Flawless (1)	95.77%	98.79%	97%	99.73%	
EV-YOLO X	Defective (0)	97.54%	97.54%	98%	99.25%	99.25%
	Flawless (1)	97.14%	98.79%	98%	99.25%	

Table 3. Results of Params, GFLOPs, mAP and FPS experiments with different models.

Model	Params(MB)	GFLOPs	mAP	FPS
YOLOv3	61.63	156.62	98.79	30
YOLOv5-l	46.64	115.92	99.08	26
YOLOv7-l	37.3	106.47	97.98	34
RetinaNet	36.35	191.42	98.75	18
YOLOX-S	8.94	26.76	99.77	32
EV-YOLOX (ours)	5.28	8.94	99.25	33

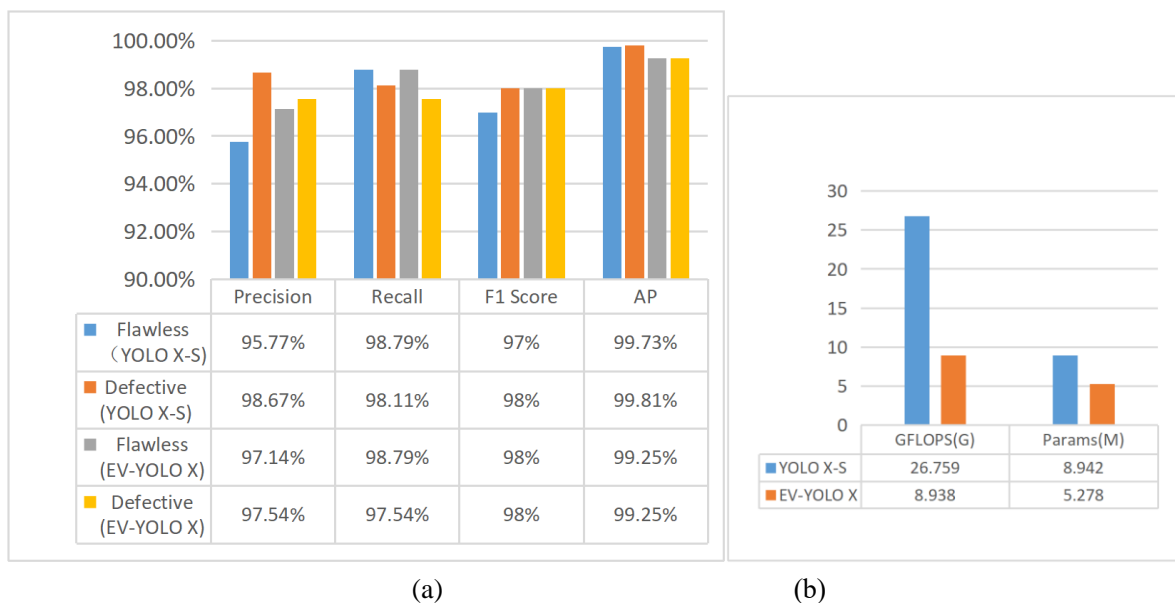
(a)   (b)

Figure 8. Overall performance evaluation of EV-YOLO X.

2.2. Recognition of thin shelled cocoons

2.2.1. Recognition based on bottom light source

For thin-shelled cocoons, the cocoon layer is thin and inelastic, and the quality is much worse compared to high-quality cocoons. Therefore, in order to obtain high-quality cocoons, the thin-shelled cocoons need to be further selected. However, there is no obvious distinguishability between thin-shelled cocoons and flawless cocoons, and thus no high-precision distinction can be obtained based on optical images fed directly into a deep learning-based object detection algorithm. We distinguish thin-shelled and flawless cocoons based on the finding that they have different penetrations of light. The specific principle is that thin-shelled cocoons, due to their thinner shell, can penetrate more light compared to flawless cocoons when illuminated by light, and this distinguishing feature can be used to distinguish between flawless and thin-shelled cocoons. As shown in Figure 9, Figure 9(a) is a thin-shelled cocoon, and Figure 9(b) is a flawless cocoon. This distinctive feature can guide industrial robots equipped with tooling fixtures to complete grasping and picking operations.
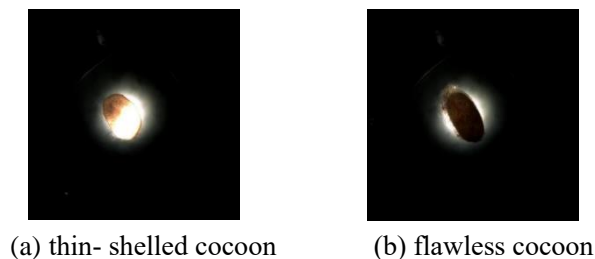

(a) thin- shelled cocoon   (b) flawless cocoon

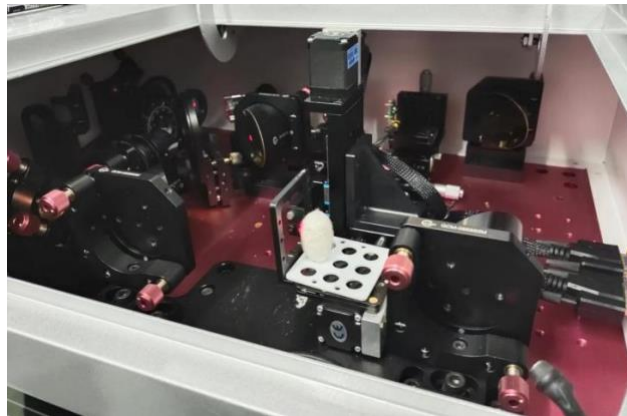**Figure 9.** Light penetration images of thin shelled and flawless cocoons.

Figure 10. Terahertz spectral measurement equipment.

2.2.2. Recognition based on terahertz spectrum

Between infrared and microwave, terahertz is an electromagnetic wave with a frequency in the range of 0.1 THz to 10 THz and a wavelength of about 0.03 mm to 3 mm, which has many features and advantages that other electromagnetic waves do not have. Terahertz has excellent spectral discrimination ability, and its band contains rich spectral information. Silk exhibits strong absorption

and dispersion properties in this band. This makes it possible to apply terahertz detection technology to further analysis of flawless cocoons and thin-shelled cocoons to improve the identification accuracy of thin-shelled cocoons and obtain higher-quality cocoons. The terahertz spectroscopy measurement equipment used for this experiment is shown in Figure 10. The results of the data measured by the terahertz technique for the flawless and thin-shelled cocoons are shown in Table 4, and the visualization results are shown in Figure 11. It can be seen that for flawless cocoons, the amplitude is less than 3×10^{-3}, and for thin-shelled cocoons, the amplitude is greater than 3×10^{-3}, indicating more energy absorption for flawless cocoons and less for thin-shelled cocoons. Based on this observation, we can distinguish between thin-shelled cocoons and flawless cocoons. Through the analysis of amplitude, it can be obtained that 3/6.34 = 47%; the distinguishing threshold is 47%. When the amplitude ratio between the placed cocoon and the not-placed cocoon is greater than the threshold value of 47%, it can be judged as a thin-shelled cocoon; otherwise, it is a flawless cocoon.
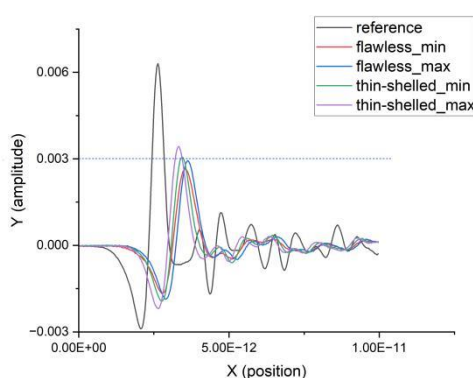

Figure 11. Terahertz detection of cocoons.

Table 4. Terahertz detection results.

	Position (X)	Amplitude (Y)
Flawless cocoon	$(20.45–20.47) \times 10^{-13}$	$(2.63–2.93) \times 10^{-3}$
Defective cocoon	$(20.48–20.50) \times 10^{-13}$	$3.09–3.44) \times 10^{-3}$
Nothing (reference)	20.61×10^{-13}	6.34×10^{-3}

3. Conclusions

In this work, an efficient and lightweight object detection network is constructed for recognizing normal defective cocoons. It achieves 99.25% mAP on the constructed mulberry cocoon dataset, and it has only 33.42% of the GFLOPS of YOLO X-S and 59.05% of the parameters of YOLO X-S. Light penetrating and combined with terahertz spectrum data are utilized for thin-shelled cocoon picking and classification. Intelligent detection of multiple types of defects in mulberry cocoons can be achieved, and multiple characteristics of defects can be expressed quantitatively through a non-contact approach. This can greatly improve the production efficiency of silk reeling enterprises and the quality of raw silk. Although the network is designed to be lightweight, the network structure can be further optimized to adapt to resource-constrained environments in subsequent, more complex application scenarios. Moreover, because the dataset used in this study is a self-constructed dataset, which lacks a certain

degree of diversity and performance in different scenarios, the dataset can be expanded to diversify the dataset in subsequent studies.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgment

Conflict of interest

The authors declare that there are no conflicts of interest.

References

1. A. N. J. Raj, R. Sundaram, V. G. V. Mahesh, Z. Zhuang, A. Simeone, A multi-sensor system for silkworm cocoon gender classification via image processing and support vector machine, *Sensors*, **19** (2019), 2656. https://doi.org/10.3390/s19122656

2. J. Cai, L. Yuan, B. Liu, L. Sun, Nondestructive gender identification of silkworm cocoons using X-ray imaging with multivariate data analysis, *Anal. Methods*, **18** (2014), 67224–7233. https://doi.org/10.1039/C4AY00940A

3. F. Guo, F. He, D. Tao, G. Li, Automatic exposure correction algorithm for online silkworm pupae (Bombyx mori) sex classification, *Comput. Electron. Agric.*, **198** (2022), 107108. https://doi.org/10.1016/j.compag.2022.107108

4. Y. Ma, Y. Xu, H. Yan, G. Zhang, On-line identification of silkworm pupae gender by short-wavelength near infrared spectroscopy and pattern recognition technology, *J. Near Infrared Spectrosc.*, **29** (2021), 207–215. https://doi.org/10.1177/0967033521999745

5. A. Nasiri, M. Omid, A. Taheri-Garavand, An automatic sorting system for unwashed eggs using deep learning, *J. Food Eng.*, **283** (2020), 110036. https://doi.org/10.1016/j.jfoodeng.2020.110036

6. V. Pavithra, R. Pounroja, B. S. Bama, Machine vision based automatic sorting of cherry tomatoes, in *2015 2nd International Conference on Electronics and Communication Systems (ICECS)*, (2015), 271–275. https://doi.org/10.1109/ECS.2015.7124907

7. F. Wang, J. Zheng, X. Tian, J. Wang, L. Niu, W. Feng, An automatic sorting system for fresh white button mushrooms based on image processing, *Comput. Electron. Agric.*, **151** (2018), 416–425. https://doi.org/10.1016/j.compag.2018.06.022

8. W. Xiao, J. Yang, H. Fang, J. Zhuang, Y. Ku, X. Zhang, Development of an automatic sorting robot for construction and demolition waste, *Clean Technol. Environ. Policy*, **22** (2020), 1829–1841. https://doi.org/10.1007/s10098-020-01922-y

9.   W. Du, J. Zheng, W. Li, Z. Liu, H. Wang, X. Han, Efficient recognition and automatic sorting technology of waste textiles based on online near infrared spectroscopy and convolutional neural network, *Resour., Conserv. Recycl.*, **180** (2022), 106157. https://doi.org/10.1016/j.resconrec.2022.106157

10. Z. Tan, H. Li, X. He, Optimizing parcel sorting process of vertical sorting system in ecommerce warehouse, *Adv. Eng. Inf.*, **48** (2021), 101279. https://doi.org/10.1016/j.aei.2021.101279

11. H. Nadaf, G. V. Vishaka, M. Chandrashekharaiah, M. S. Rathore, Scope and potential applications of artificial intelligence in tropical tasar silkworm Antheraea mylitta D. seed production, *Entomol. Zool.*, **9** (2021), 899–903.

12. K. Kanjanawanishkul, An image-based eri silkworm pupa grading method using shape, color, and size, *Int. J. Autom. Smart Technol.* **12** (2022), 2331–2331. https://doi.org/10.5875/ausmt.v12i1.2331

13. F. Dai, X. Wang, Y. Zhong, S. Zhong, C. Chen, Convolution neural network application in the simultaneous detection of gender and variety of silkworm (bombyx mori) cocoons, in *5th International Conference on Computer Science and Information Engineering (ICCSIE 2020)*, **1769** (2021), 012017. https://doi.org/10.1088/1742-6596/1769/1/012017

14. H. Xiong, J. Cai, W. Zhang, J. Hu, Y. Deng, J. Miao, et al., Deep learning enhanced terahertz imaging of silkworm eggs development, *Iscience*, **24** (2021). https://doi.org/10.1016/j.isci.2021.103316

15. R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2022), 580–587. https://doi.org/10.1109/CVPR.2014.81

16. R. Girshick, Fast R-CNN, in *2015 IEEE International Conference on Computer Vision (ICCV)*, (2015), 1440–1448. https://doi.org/10.1109/ICCV.2015.169

17. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (2017), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

18. K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, *IEEE Trans. Pattern Anal. Mach. Intell.*, **37** (2015), 1904–1916. https://doi.org/10.1109/TPAMI.2015.2389824

19. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, et al., SSD: Single shot multibox detector, in *European Conference on Computer Vision*, (2016), 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

20. T. Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in *2017 IEEE International Conference on Computer Vision (ICCV)*, (2017), 2999–3007. https://doi.org/10.1109/ICCV.2017.324

21. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 779–788. https://doi.org/10.1109/ICCV.2017.324

22. J. Redmon, A. Farhadi, YOLO9000: Better, faster, stronger, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), 6517–6525. https://doi.org/10.1109/CVPR.2017.690

23. J. Redmon, A. Farhadi, YOLOv3: An incremental improvement, preprint, arXiv: 1804.02767.

24. A. Bochkovskiy, C. Y. Wang, H. Liao, YOLOv4: Optimal speed and accuracy of object detection, preprint, arXiv: 2004.10934.

25. Z. Ge, S. Liu, F. Wang, Z. Li, J. Sun, YOLOX: Exceeding YOLO Series in 2021, preprint, arXiv: 2107.08430.

26. J. Pan, A. Bulat, F. Tan, X. Zhu, L. Dudziak, H. Li, et al., EdgeViTs: Competing light-weight cnns on mobile devices with vision transformers, in *Computer Vision – ECCV 2022*, (2022), 294–311. https://doi.org/10.1007/978-3-031-20083-0_18

27. S. Liu, L. Qi, H. Qin, J. Shi, J. Jia, Path aggregation network for instance segmentation, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2018), 8759–8768. https://doi.org/10.1109/CVPR.2018.00913

28. J. Yu, Y. Jiang, Z. Wang, Z. Cao, T. Huang, UnitBox: An advanced object detection network, in *Proceedings of the 24th ACM international conference on Multimedia*, (2016), 516–520. https://doi.org/10.1145/2964284.2967274