IECE Transactions on Internet of Things
http://dx.doi.org/10.62762/TIOT.2023.539452

RESEARCH ARTICLE

An Improved YOLOv3-Based Method for Immature Apple Detection

Zhongqiang Huang[1], **Ping Zhang**[2,*], **Ruigang Liu**[1] **and Dongxu Li**[2]

[1] School of Informatics, Huazhong Agricultural University, Wuhan 430070, China
[2] School of Computer, BaoJi University of Arts and Sciences, Baoji 721016, China

Abstract

The identification of immature apples is a key technical link to realize automatic real-time monitoring of orchards, expert decision-making, and realization of orchard output prediction. In the orchard scene, the reflection caused by light and the color of immature apples are highly similar to the leaves, especially the obscuration and overlap of fruits by leaves and branches, which brings great challenges to the detection of immature apples. This paper proposes an improved YOLOv3 detection method for immature apples in the orchard scene. Use CSPDarknet53 as the backbone network of the model, introduce the CIOU target frame regression mechanism, and combine with the Mosaic algorithm to improve the detection accuracy. For the data set with severely occluded fruits, the F1 and mAP of the immature apple recognition model proposed in this article are 0.652 and 0.675, respectively. The inference speed for a single 416×416 picture is 12 ms, the detection speed can reach 83 frames/s on 1080ti, and the inference speed is 8.6 ms. Therefore, for the severely occluded immature apple data set, the method proposed in this article has a significant detection effect, and provides a feasible solution for the automation and mechanization of the apple industry.

Academic Editor:
Jinchao Chen

Submitted: 03 May 2023
Accepted: 09 June 2023
Published: 12 June 2023

Vol. 1, **No.** 1, 2023.
10.62762/TIOT.2023.539452

***Corresponding author:**
Ping Zhang
jcjyxy@163.com

Keywords: Orchard scene, Immature apple, Improved YOLOv3, Mosaic algorithm, CIOU target frame regression mechanism.

Citation
Huang, Z., Zhang, P, Liu R. & Li, D. (2023). An Improved YOLOv3-Based Method for Immature Apple Detection. IECE Transactions on Internet of Things, 1(1), 9–14.

1 Introduction

Apple is one of the common fruits and it occupies a large proportion in the domestic and foreign fruit market. Apple is the fourth most important fruit produced and eaten all around the world with a production of 84 million tonnes in 2014 [1]. Apple has high nutritional value and is one of the four major fruits in the world. Apple production plays an important role in the development of our country's economy and meeting the needs of the broad masses of people. However, the traditional apple planting, fertilization and pesticide application, picking, and a series of subsequent production and processing processes consume a lot of manpower, material resources and other resources. In the precise planting of apples, in order to carry out intelligent spraying and growth monitoring of apples, as soon as possible to estimate the production of apples and forecast the labor demand for picking, it is necessary to accurately detect immature apples [2]. Computer vision provides a very effective method for automatic fruit detection, but the detection of occluded or overlapping fruits in orchard scenes has always been a difficult problem for fruit detection.

In the orchard scene, the environment of the apple orchard is complex. There is not only the occlusion between the fruit and the fruit, the leaves and the

Figure 1. datasets

fruit, but also the shadow of the fruit is high due to the influence of light. Automated harvesting requires accurate detection and recognition of the fruit within a tree canopy in real-time in uncontrolled environments. However, occlusion, variable illumination, variable appearance and texture make this task a complex challenge [3–7].

Due to the poor robustness of traditional machine vision methods in complex scenes, it is difficult to meet the work requirements of picking robots in complex scenes. In recent years, the convolutional neural network [8] has been continuously improved in the field of target detection and has shown great advantages. It is mainly divided into two categories. One is that the algorithm generates a series of candidate frames as samples, and then convolution The neural network performs sample classification, represented by RCNN [9], Fast RCNN [10] and Faster RCNN [11]; a type of problem that directly converts the positioning of the target frame into a regression problem, does not need to generate candidate frames, and the iconic algorithms include SSD [12], YOLO [13], etc.

This paper improves the YOLOv3 network model [14], uses a new IoU bounding box regression loss function, and combines it with transfer learning [15] to transfer the knowledge learned by the model in the ImageNet dataset to the immature image recognition process , Proposes a method for identifying immature apples based on CIoU's [16] YOLOv3 neural network. The current deep learning model combined with wireless sensor network can more effectively realize agricultural automation [17].

2 Materials and Methods

2.1 Datasets

The Apple Orchard is located in Baoji City, Shaanxi Province. During October 2020, using the Huawei honor X10 mobile phone device, at a distance of 1m from each apple tree, more than 500 images of immature apple fruits that were blocked to varying degrees were collected from different viewing angles and different directions. In the end, 400 pictures were screened out and divided into 300 training sets and 100 test sets with a ratio of 0.75:0.25. The data mainly covers the pictures where the fruit is blocked when the light is reflected, backlit, or forward on sunny and cloudy days.

In order to standardize the data set, the original image resolution is fixed to 416x416 using the center cropping method.

2.2 Data Augmentation

Use the opencv tool library to perform data amplification on the original data set, rotate the original image, and the rotation angle is randomly selected $+-90°$; the original image is randomly mirrored, horizontally flipped, and vertically flipped;

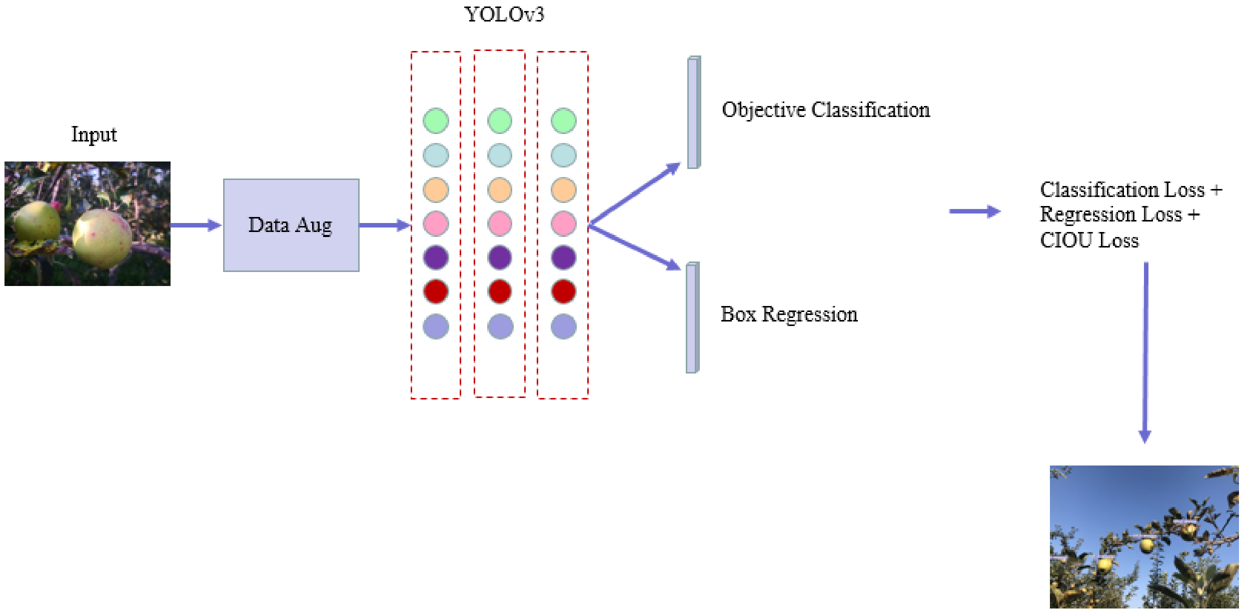

Figure 2. Model Structure

expanded by cropping and zooming data set. Considering that data enhancement will cause more serious changes in the shape and quality of the images in the pictures, each picture is randomly amplified by the above method.

2.3 Border regression loss function based on CIoU

IoU is the degree of overlap between the predicted frame and the marked frame in the original picture. In the field of target detection, the IoU value of the frame regression is often used as an evaluation indicator. However, most detection frameworks do not combine this value to optimize the loss function. IoU can be backpropagated, and it can be directly used as the objective function to optimize. Considering the choice between the optimization metric itself and the use of alternative loss functions, the best choice is to optimize the metric itself. As a loss function, traditional IoU has two shortcomings: if two objects do not overlap, the IoU value will be zero, and its gradient will be zero, which cannot be optimized; two objects overlap in multiple different directions, and the intersection level is the same, Its IoU will be exactly the same, IoU cannot accurately reflect the degree of overlap between the two. A good target frame regression loss should consider three important geometric factors: overlap area, center point distance, and aspect ratio. Therefore, CIOU not only considers the overlap area and the center point distance, but also considers the aspect ratio, and its convergence accuracy is higher. Therefore, the value of the IoU function does not reflect how the two objects overlap. In the fruit recognition of the apple picking robot, the accuracy of the position of the return frame directly determines the success rate of the robot hand picking. Therefore, this article proposes to introduce CIoU to solve the shortcomings of IoU. CIoU is a very good distance metric, which can replace the loss function of border regression in most target detection algorithms, as shown in equations (1)-(3).

$$\mathrm{IOU} = \left|\frac{A \cap B}{A \cup B}\right| \tag{1}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right) \tag{2}$$

$$L_{\mathrm{CIOU}} = 1 - \mathrm{IOU} + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + av \tag{3}$$

Where a is the parameter used for trade-off, and v is the parameter used to measure the consistency of the aspect ratio.

2.4 Model structure

YOLO is an end-to-end target detection model. The basic idea of the YOLO algorithm is: first extract features from the input features through a feature extraction network to obtain a feature map output of a specific size. The input image is divided into 13×13 grid cells, and then if the center coordinates of an object in the real frame fall in a certain grid cell, then the grid cell will predict the object. Each object has a fixed number of bounding boxes. There are three bounding boxes in YOLO v3. Logistic regression is used to determine the regression box used for prediction.

The following is the loss function formula:

$$\lambda_{coord} \sum_{i=0}^{s^2} \sum_{j=0}^{B} 1_{i,j}^{obj} \left[(b_x-\hat{b}_x)^2+(b_y-\hat{b}_y)^2+(b_w-\hat{b}_w)^2+(b_h-\hat{b}_h)^2 \right] + \sum_{i=0}^{s^2} \sum_{j=0}^{B} 1_{i,j}^{obj} \left[-\log(p_c) \right] + \sum_{i=1}^{n} BCE(\hat{c}_i, c_j) + \lambda_{noobj} \sum_{i=0}^{s^2} \sum_{j=0}^{B} 1_{i,j}^{noobj} \left[-\log(1-p_c) \right] \tag{4}$$

Among them, s is the number of grids, which s^2 is 13*13, 26*26, 52*52. B is box. If the box has a target, the value of $1_{i,j}^{obj}$ is 1, otherwise it is 0. If the box has no target, the value of $1_{i,j}^{\text{noobj}}$ is 1, otherwise it is 0.

3 Experiments

The experiment is based on the ICOU algorithm and data Augmentation, and after inputting the data into the model, we got the accuracy, recall rate, F1, and mAP values are 0.616, 0.692, 0.652, 0.675 respectively.

Table 1. Detection results using CIoU loss function

Dataset	Model	Pre	Recall	F1	mAP
mmature apple (Test)	I YOLOv3+CIOU	0.616	0.692	0.652	0.675

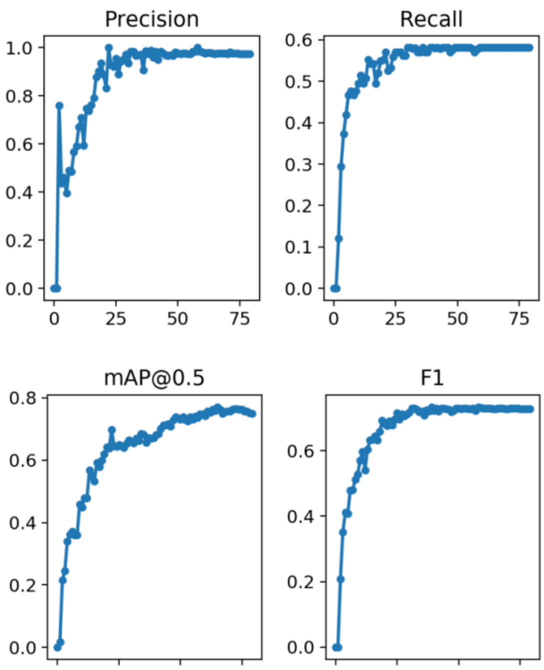

Figure 3. Data index during training

4 Conclusion

1) This article proposes an improved YOLOv3 neural network to identify immature apples. Experimental results show that in general natural scenes, affected by conditions such as illumination, occlusion, and overlap, the F1 and mAP values detected by the model reach 65.2

2) Subsequent research should focus on how to improve the detection performance of the model under the influence of illumination, occlusion, overlap, etc., and provide good technical support for the positioning of the picking robot.

Figure 4. Generate graph during training

Figure 5. General scene rendering

Figure 6. Severe occlusion effect picture

Acknowledgement

This research was funded by the Special Project of Education Department of Shaanxi Provincial Government of china, grant number 16JK1048.

References

[1] Musacchi, S., & Serra, S. (2018). Apple fruit quality: Overview on pre-harvest factors. *Scientia Horticulturae*, 234 , 409–430. [CrossRef]

[2] Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., & Liang, Z. (2019). Apple detection during different growth stages in orchards using the improved yolo-v3 model. *Computers and Electronics in Agriculture*, 157, 417–426. [CrossRef]

[3] Liu, Q., Cheng, L., Jia, A. L., & Liu, C. (2021). Deep reinforcement learning for communication flow control in wireless mesh networks. *IEEE Network*, 35(2), 112-119. [CrossRef]

[4] Huang, Y., Cheng, L., Xue, L., Liu, C., Li, Y., Li, J., & Ward, T. (2021). Deep adversarial imitation reinforcement learning for QoS-aware cloud job scheduling. *IEEE Systems Journal*, 16(3), 4232-4242. [CrossRef]

[5] Cheng, L., Wang, Y., Liu, Q., Epema, D. H., Liu, C., Mao, Y., & Murphy, J. (2021). Network-aware locality scheduling for distributed data operators in data centers. *IEEE Transactions on Parallel and Distributed Systems*, 32(6), 1494-1510. [CrossRef]

[6] Cheng, F., Huang, Y., Tanpure, B., Sawalani, P., Cheng, L., & Liu, C. (2022). Cost-aware job scheduling for cloud instances using deep reinforcement learning. *Cluster Computing*, 1-13. [CrossRef]

[7] Li, J., Tong, X., Liu, J., & Cheng, L. (2023). An Efficient Federated Learning System for Network Intrusion Detection. *IEEE Systems Journal*. [CrossRef]

[8] Huang, H., Xue, X., Liu, C., Wang, Y., Luo, T., Cheng, L., ... & Li, X. (2023). Statistical Modeling of Soft Error Influence on Neural Networks. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*. [CrossRef]

[9] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 580–587). Columbus, OH, USA. [CrossRef]

[10] Girshick, R. B. (2015). Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1440–1448). Santiago, Chile: IEEE Computer Society. [CrossRef]

[11] Ren, S., He, K., Girshick, R. B., & Sun, J. (2015). Faster R-CNN: towards real-time object detection with region proposal networks. *arXiv*. [CrossRef]

[12] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In *Proceedings of the 14th European Conference* (Vol. 9905, pp. 21–37). Amsterdam, The Netherlands. [CrossRef]

[13] Redmon, J., Divvala, S. K., Girshick, R. B., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 779–788). Las Vegas, NV, USA: IEEE Computer Society. [CrossRef]

[14] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv*. [CrossRef]

[15] Torrey, L., & Shavlik, J. (2010). Transfer learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques* (pp. 242–264). IGI Global

[16] Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020). Distance-iou loss: Faster and better learning for bounding box regression. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence* (pp. 12993–13000). New York, NY, USA: AAAI Press. [CrossRef]

[17] Li, Y., & Cao, J. (2021). Wsn node optimal deployment algorithm based on adaptive binary particle swarm optimization. *ASP Transactions on Internet of Things*, 1 (1), 1–8. [CrossRef]

Zhongqiang Huang Currently studying in the first year of PhD in Huazhong Agricultural University. The research direction is computer vision and natural language processing.

Ping Zhang Graduated with M.S at the Department of Computer and Information Science of Liaoning Normal University from 2008 to 2011. He is a lecturer in Baoji University of Arts and Sciences. His current research interests include bioinformatics machine learning and graph neural network.

Ruigang Liu He is a master student in the School of information at Huazhong agriculture University, Wu'han, China. He expertise is link prediction and deep learning.

Dongxu Li Graduated with B.S at the Department of Computer of Baoji University of Arts and Sciences from 2018 to 2022. His current research interests include machine learning and computer vision.