RECOGNITION METHOD FOR SEED POTATO BUDS BASED ON IMPROVED YOLOv3-TINY
/
基于改进 YOLOv3-tiny 的马铃薯种薯芽眼检测方法

Wanzhi ZHANG[1,2,3], Yuelin HAN[1,2], Chen HUANG[1,2], Zhiwei CHEN [1,2] [1]
[1] College of Mechanical and Electronic Engineering, Shandong Agricultural University, /China;
[2] Shandong Provincial Key Laboratory of Horticultural Machinery and Equipment/China;
[3] Nanjing Research Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs/China
Tel: +86-0538-8249830; E-mail: zhangwanzhi@163.com
https://doi.org/10.35633/inmateh-67-37

Keywords: *improve YOLOv3-tiny, seed potatoes, buds recognition, automatic cutting, test*

ABSTRACT
This paper proposed a method of seed potato buds recognition based on improved YOLOv3-tiny. K-means clustering based on IoU is used to obtain the anchor box that meets the size of buds. Mosaic online data enhancement is used to increase image diversity and model generalization ability. The CIoU bounding box regression loss function is introduced to improve the regression effect of buds recognition. The results show that the precision (P), the recall (R), the average precision (AP), and the F_1 score of the model for seed potato buds recognition are 88.33%, 85.97%, 91.18% and 87.13% respectively. The real-time recognition speed of seed potato buds on the embedded platform NVIDIA Jetson Nano can reach 40FPS. The method proposed in this paper can meet the needs of real-time recognition of seed potato buds on the embedded platform.

摘要
本文提出一种基于 YOLOv3-tiny 网络的马铃薯种薯芽眼检测方法，以 YOLOv3-tiny 网络为基础，使用基于 IoU 的 K-means 聚类方法得到符合芽眼尺寸的先验框；使用 Mosaic 在线数据增强方式，以增加图像多样性和模型泛化能力；引入 CIoU 边框回归损失函数，以提高芽眼检测回归效果。结果表明，模型对马铃薯种薯芽眼检测的精准率 P、召回率 R、平均准确率 AP 和调和均值 F1 值分别为 88.33%、85.97%、91.18%和 87.13%；嵌入式平台 NVIDIA Jetson Nano 种薯芽眼实时检测速度可达 40FPS。本文方法能够满足在嵌入端马铃薯种薯芽眼实时检测的需求，可为后续马铃薯种薯自动化切块芽眼检测提供了技术支持。

INTRODUCTION
Potatoes are food crops with high nutritional value and good production efficiency. At present, the output of potatoes in China ranks first in the world, accounting for about 1/4 of the world yield (*Luo et al., 2018*). Potatoes are propagated by tubers, and it is necessary to ensure that each tuber has at least one bud when cutting seed potatoes. Manual cutting is mainly adopted in China, which has high labor costs. This method has seriously affected the large-scale planting and economic benefits of potatoes (*Li et al., 2019; Li et al., 2020*). Therefore, China urgently needs to realize automatic cutting of seed potatoes (*Lv et al., 2020; Wang et al., 2020; Lv et al., 2020*). Among them, the first thing to be realized is the automatic recognition of seed potato buds.

Some scholars have carried out related research on the automatic recognition of seed potato buds. Zhang Jinmin et al. proposed a seed potato buds recognition based on LBP and SVM, with a comprehensive recognition rate of 97.33% (*Zhang et al., 2020*). Li Yuhua et al. proposed a seed potato buds recognition method based on three-dimensional geometric features of color saturation, with a recognition rate of 91.48% and an average running time of 2.68 s (*Li et al., 2018*). Lv Zhaoqin et al. proposed a seed potato buds images recognition method based on Gabor features, with a recognition rate of 93.4%, a false recognition rate of 7.2%, and an average recognition speed of 0.28 s (*Lv et al., 2021*).

---
[1] *Wanzhi Zhang, Prof. Ph.D. Eng.; Yuelin Han, M.S. Stud. Eng.; Chen Huang, M.S. Stud. Eng.; Zhiwei Chen.*

Xi Rui et al. proposed a fast segmentation on potato buds with a chaos optimization-based K-means algorithm. The segmentation accuracy was 98.87%, and the average image consumption time per frame was 1.11 s (*Xi et al., 2020*).

The above research is based on the traditional visual method of extracting the color, contour, and texture characteristics of seed potato buds. The feature extraction has blindness and uncertainty, so it is difficult to meet the requirements for rapid recognition of the buds.

In recent years, target recognition based on deep learning has good real-time and robustness, so it has been widely used in the agricultural field. Typical target recognition algorithms are generally divided into two categories. One is based on two-stage target recognition algorithms, such as Faster R-CNN (*Ren et al., 2017*). This type generally has a high accuracy rate, but it has a large amount of calculation and low real-time performance. For example, Xi Rui et al. adopted a seed potato buds recognition method based on improved Faster R-CNN. The recognition accuracy is 96.32%, and the average single-frame image recognition speed is 0.183 s (*Xi et al., 2020*). The other type is a single-stage target recognition algorithm represented by YOLO. This type adopts an end-to-end idea, with high accuracy and real-time performance. It maintains a good balance between recognition speed and recognition accuracy. At present, many scholars have applied the YOLO algorithm to specific target recognition and proposed different improvement schemes for specific problems. Li Wentao et al. proposed an improved YOLOv3-tiny algorithm, which is applied to the recognition of pedestrians and agricultural machinery obstacles in the field. The average precision (*AP*) is 86.5%. On the embedded platform Jetson TX2, it takes an average of 122 ms per frame to detect the target (*Li et al., 2020*). Lv Shilei et al. proposed the orange recognition method based on an improved YOLOv3-LITE lightweight neural network, using MobileNet-v2 as the backbone network. For the lightly occluded orange dataset, the average precision (*AP*) is 91.13%, the F_1 score is 95.27%, and the recognition speed on GPU can reach 246 frames per second (*Lv et al., 2019*). Liang C et al. proposed a method based on YOLOv3 to detect litchi fruits and fruit stems in a night environment. The average precision (*AP*) under high brightness, normal brightness, and low brightness are 96.8%, 99.6%, and 89.3%, respectively (*Liang et al., 2020*). Liu Fang et al. proposed an improved multi-scale YOLO algorithm to quickly identify tomato fruits in a complex environment. The average precision (*AP*) is 97.13%, the precision (*P*) is 96.36%, and the recall (*R*) is 96.03% (*Liu et al., 2020*).

The development of the above research provides a reference and feasibility basis for the application of a convolutional neural network to the recognition of seed potato buds. And it can avoid the shortcomings of the feature extraction process in traditional machine vision methods. Therefore, this paper proposes a method of seed potato buds recognition based on improved YOLOv3-tiny. Firstly, use *imgaug* data enhancement of the collected samples to prevent overfitting. Then the K-means clustering based on IoU is used to obtain anchor boxes that meet the size of the buds. Finally, based on the YOLOv3-tiny network framework, transfer learning, Mosaic data enhancement, and CIoU regression loss function are introduced in the training process to realize real-time recognition of seed potato buds.

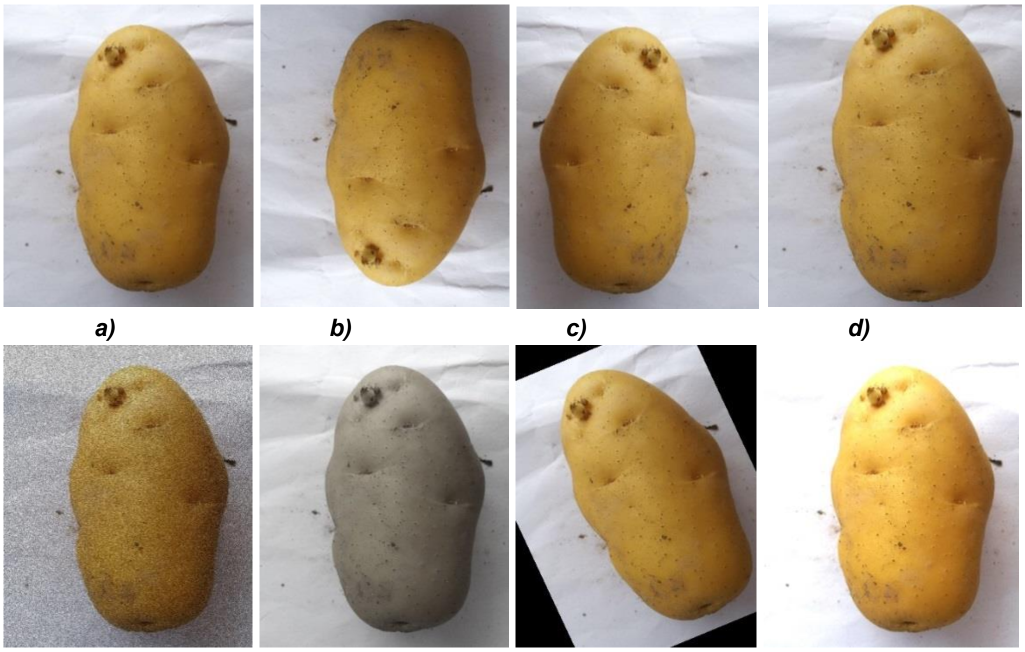
MATERIALS AND METHODS

Data collection

The collection site of seed potatoes images data is the potato planting base in Tengzhou City, Shandong Province. The variety is “Favorita”. Selected seed potatoes are disease-free, non-injured, smooth-skinned, and well-stored for image collection. The quality of each seed potato is 300-500 g. 1,350 seed potatoes have been selected for the experiment. Since the bud information contained in the front and back sides of the seed potato is different, 2 images are collected for each seed potato, for a total of 2700 images.

Data preprocessing

To enrich the dataset, effectively extract the characteristics of seed potato buds, and enhance the robustness of the model, data enhancement is used to enhance the seed potato dataset. Denoising, enhancing, cropping, flipping, and resampling are performed through imgaug data enhancement methods to increase the diversity of the image and reduce the lack of model generalization ability due to the small sample size. The enhanced images are 5400, and some of the seed potato images are enhanced as shown in Fig.1. Then the dataset is randomly divided into the training set, verification set, and test set according to the ratio of 7:2:1, so 3780 images are used for training, 1080 images are used for verification, and 540 images are used for testing.



a) b) c) d)

e) f) g) h)

Fig. 1 – Image enhancement

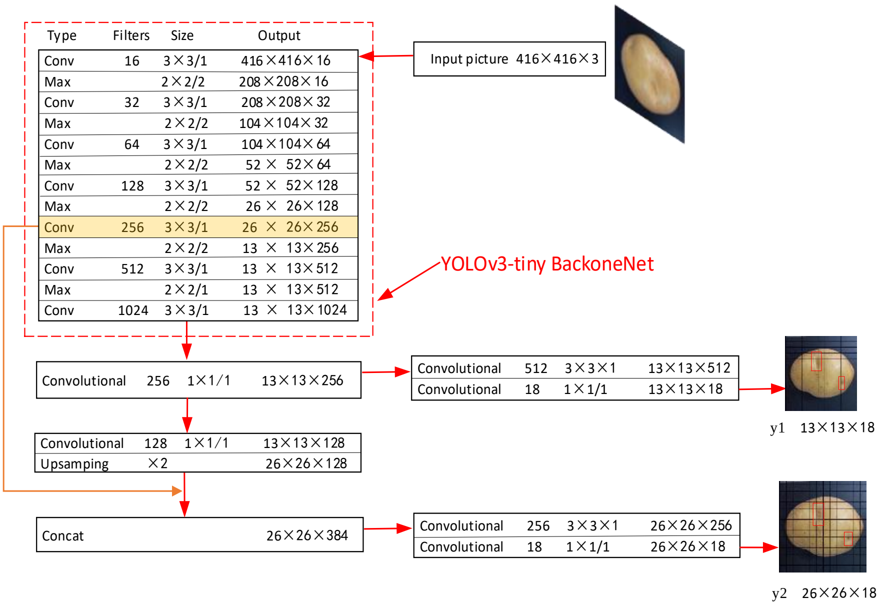
a) Original image; b) Vertical mirror; c) Horizontal mirroring; d) Random crop; e) Add Gaussian noise; f) Rotation; g) Grayscale; h) Change brightness

Data annotation

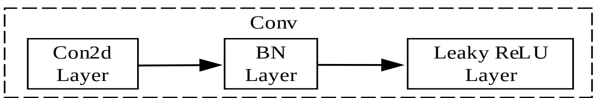
After completing the enhancement of the sample data, the LabelImg image annotation tool is used to mark the seed potato buds area with the smallest enclosing rectangle. The label type is marked with “buds”. The sample images are all in the PASCAL VOC2007 datasets formats to generate a label file in the “.xml” format to facilitate subsequent processing.

YOLOv3-tiny network

The YOLOv3-tiny network is mainly used for the rapid recognition of seed potato buds. The YOLOv3-tiny network is a simplified version of the YOLOv3 network, and its structure is shown in Fig.2.



a)



b)

Fig. 2 – YOLOv3-tiny network

a) YOLOv3-tiny network structure; b) Convolutional layer structure

The YOLOv3-tiny network structure has 24 layers. The backbone network consists of 7 convolution layers and 6 pooling layers. The convolution layer uses a 3×3 convolution kernel to extract the characteristic information of the seed potato. The 1×1 convolution kernel is used to increase the non-linear features while keeping the size of the feature map unchanged and reduce the network training parameters. The convolutional layer structure (Conv) is shown in Fig.2b, which is mainly composed of a 3×3 convolutional layer, BN (batch normalization) layer, and Leaky ReLU in sequence.

The pooling layer is used to extract bud features and adjust the size of the feature map, while removing other redundant information, reducing the number of parameters required for network calculations. The first five pooling layers of the backbone network have a step size of 2, and each time the maximum pooling is performed, the size of the feature map becomes 1/2 of the original. The step size of the last pooling layer is 1, and the size of the feature map remains unchanged after pooling. It can be seen that when the seed potato image is input, it can get feature maps of different scales by down-sampling through the backbone network. Here, the backbone network mainly outputs two feature maps (13×13×1024 and 26×26×256). Among them, the feature map (13×13×1024) is subjected to the convolution (1×1×256) for feature extraction to obtain a feature map (13×13×256). Then, after two convolution operations, a feature map y1 (13×13×18) is output. In addition, the feature map (13×13×1024) is also upsampling after one convolution and then spliced with the feature map (26×26×256) output by the backbone network. Finally, after 2 convolution operations, feature map y2 (26×26×18) is output.

The seed potato image is processed by the YOLOv3-tiny network, which can realize the fusion of low-level features and deep-level features, and extract more comprehensive bud feature information. The feature map (13×13×1024) has a large receptive field, which is suitable for detecting large seed potato buds. The feature map (26×26×256) has a small receptive field, which is suitable for detecting small seed potato buds.

Loss function

Convolutional neural networks usually calculate the loss function based on Intersection over Union (IoU). IoU is the ratio of the intersection and union between the target prediction box and the ground truth box, reflecting the degree of overlap between the two. The higher the value, the closer the prediction box and the ground truth box are. Early experiments found that the Intersection over Union (IoU) was not sensitive to the size of the bud, and cannot optimize its overlapping part. Therefore, the position loss function of the YOLOv3-tiny is calculated based on the global intersection and union ratio (Complete Intersection over Union, CIoU). CIoU can take into account the overlap area of the target and the prediction frame, the distance between the center point, the aspect ratio, and the penalty items, making the target frame regression more stable, and avoiding problems such as divergence of IoU during the training process, thereby ensuring the convergence speed and accuracy of the prediction box regression (*Zheng J et al, 2020*). The convergence speeds and accuracy of the prediction box regression are guaranteed, as shown in Fig. 3.

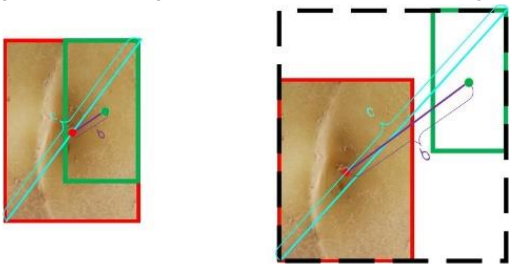


Fig. 3 – CIoU bounding box regression

The CIoU calculation formula can be defined as:

$$\text{CIoU} = \text{IoU} - \frac{\rho^2(b,b^{gt})}{c^2} - \varepsilon\upsilon \tag{1}$$

Where:

b represents the center point of the prediction box, b^{gt} represents the center point of the ground truth box, $\rho 2(b, b^{gt})$ represents the Euclidean distance between the prediction box and the center point of the ground truth box, c represents the diagonal distance that can contain the minimum closure area of the prediction box and the ground truth box at the same time. ε is the trade-off coefficient, and υ is the consistency coefficient of the aspect ratio.

The calculation formulas are defined respectively as:

$$\begin{cases} \varepsilon = \frac{\upsilon}{(1-\mathrm{IoU})+\upsilon} \\ \upsilon = \frac{4}{\pi^2}\left(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h}\right)^2 \end{cases} \tag{2}$$

Where:

w^{gt} and h^{gt} represent the width and height of the ground truth box respectively, *w* and *h* respectively represent the width and height of the prediction box. The loss function L_{CIoU} of the YOLOv3-tiny model in this paper. The calculation formula is defined as:

$$L_{\mathrm{CIoU}} = 1 - \mathrm{IoU} + \frac{\rho^2(b,b^{gt})}{c^2} + \varepsilon\upsilon \tag{3}$$

According to formula (3), the location-based training set Loss value and the validation set Val_Loss value can be calculated separately.

Mosaic data enhancement

To solve the problem of overfitting and insufficient generalization ability in the small dataset, Mosaic data enhancement is used to enhance the training set during the training process. After acquiring a batch of images, transformation operations are randomly performed such as hue, saturation, brightness, rotation, scaling, translation, and shearing on the batch of images with a certain probability to improve the robustness and accuracy of the model for bud recognition. The implementation process of the Mosaic data enhancement method is as follows. First, read 4 seed potato pictures at a time. Then perform the above-mentioned random transformation operation on the 4 pictures respectively, and place them according to the positions in the 4 directions. Finally, complete the combination of seed potato images and buds label frame, as shown in Fig. 4.

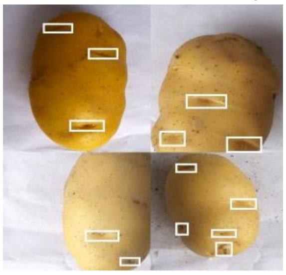


Fig. 4 – Mosaic data enhancement results

Anchor boxes clustering

The anchor boxes in the original YOLOv3-tiny network were generated using a clustering algorithm on the PASCAL VOC datasets. The PASCAL VOC datasets are mostly based on natural scenes, with many target categories. In seed potato bud recognition, there is only one type of recognition target, so the original anchor boxes will affect the network training speed and recognition accuracy. To speed up the network convergence, the K-means clustering algorithm is used to re-cluster the seed potato dataset to obtain anchor boxes suitable for the seed potato buds recognition scene (*Coates et al., 2012*). At the same time, the average intersection ratio is used instead of the Euclidean distance as the clustering analysis standard to reduce the influence of the anchor box size on the recognition error. The anchor box clustering calculation result is shown in Fig. 5. Since the seed potato buds recognition network proposed is based on 2 scale feature maps on the YOLOv3-tiny network, 6 anchor boxes of different sizes are obtained through clustering. The widths and heights of the 6 anchor boxes are: (68×17), (40×112), (16×49), (16×33), (26×72), (44×16), where the first 3 anchor boxes are used for output prediction of feature map y1, and the last 3 anchor boxes are used for output prediction of feature map y2.

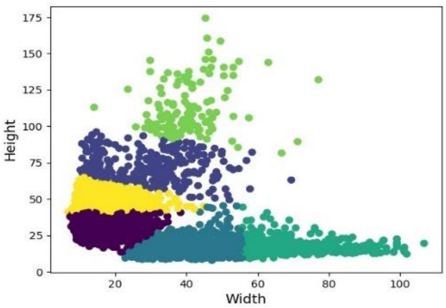


Fig. 5 – Anchor box clustering calculation

RESULTS
Test evaluation index

To evaluate the performance of YOLOv3-tiny for detecting seed potato buds, the precision (*P*), recall (*R*), average precision (*AP*), and F_1 score are used as evaluation indicators. The calculation formula of the evaluation index is as follows.

The calculation formulas for precision (*P*) and recall (*R*) are defined as:

$$P = \frac{TP}{TP+FP} \times 100\% \tag{4}$$

$$R = \frac{TP}{TP+FN} \times 100\% \tag{5}$$

Where:

TP, FP, and FN mean the number of correctly detected seed potato buds objects (true positives), the number of falsely detected seed potato buds objects (false positives), and the number of missed seed potato buds objects (false negatives), respectively.

The average precision (AP) represents the comprehensive effect of precision (P) and recall (R). The reflection in the graph is the size of the area included under the P and R curve, which is used to evaluate the performance of the model on a single recognition category of the bud, which reflects the overall performance of algorithm recognition, its calculation formula is defined as:

$$AP = \int_0^1 P(R)dR \tag{6}$$

The F_1 score represents the balance parameter of the precision (P) and the recall (R). When the F_1 score is closer to 1, it means that the recognition effect is better. The calculation formula is defined as:

$$F_1 = \frac{2 \times P \times R}{P+R} \tag{7}$$

Test Platform

The desktop computer system is Windows 10 (64-bit), equipped with an Intel Core i7-9700F CPU, clocked at 3.0GHz, equipped with NVIDIA GeForce RTX 2070 GPU, and 16G of memory. The test running environment is Tensorflow-gpu 1.13.2, Keras 2.3.1 artificial neural network library, CUDA version 10.0 parallel computing architecture, and cuDNN version 9.0 deep neural network library. The programming language is the Python 3.6 version. Image morphology processing uses OpenCV 4.2.0 vision library.

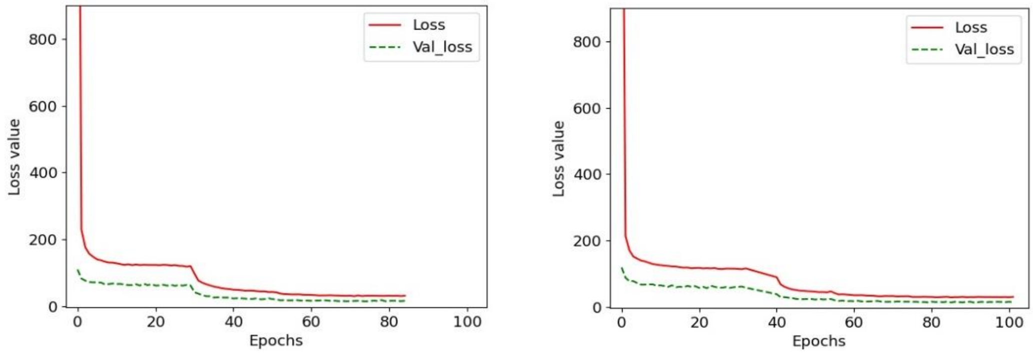
Model performance testing under different iteration times

The proposed model is trained using a seed potato dataset. To prevent the model from underfitting due to insufficient iteration times of Epochs, a better recognition model is obtained by adjusting the Epochs times. According to the comparison in Table 1, when the number of Epochs in the first stage increases from 30 to 40, the Loss/Val_Loss loss value drops by 4.719/3.262. When the number of Epochs in the first stage is greater than 40, the drop rate of the loss value remains unchanged, and the training loss value remains at about 12. Therefore, when a larger number of Epochs is set, the reduction of the model loss value is too small and the training time is increased. So, it is more reasonable to set the number of Epochs in the first stage to 40. In the second stage of training, to prevent underfitting due to insufficient training iterations, gradually increase the number of iterations. And in the training process, the early stopping mechanism is added to prevent the model from overfitting after training. From the above experiments, it can be obtained that by adjusting the training Epochs, the seed potato dataset can be fully trained. At the same time, the increase of the early stopping mechanism saves training time, accelerates the learning speed, and prevents the model from overfitting. Therefore, the number of Epochs in the first stage is 40, and the number of Epochs in the second stage is 120.

Table 1

Model-checking analysis under different Epochs

Test group number	The first stage of freezing training		The second stage of thawing training		Training time /h	*AP*/%
	Epochs	Loss/Val_Loss	Epochs	Loss/Val_Loss		
1	30	17.177/ 16.229	90	7.064/ 6.798	10.29	88.07
2	40	12.458/ 12.967	100	3.363/ 3.395	11.43	88.76
3	50	12.438/ 12.849	110	3.273/ 3.486	13.10	90.19
4	40	12.127/ 12.645	120	3.191/ 3.321	12.36	91.18



a) *b)*

Fig. 6 – Loss curve during training

a) Test group number 1; b) Test group number 4

Recognition effect after using transfer learning

Transfer learning can reduce the amount of calculation required for deep learning models, and better solve the problem of the small dataset that is easy to overfitting on complex network structures (*Pan et al., 2010*). Therefore, this paper adopts transfer learning and fine-tuning on the seed potato data collection to initialize the training parameters. At the same time, the parameters of all layers are trained from scratch as a control experiment. The recognition results on the seed potato test set are shown in Table 2. It can be seen that compared with the original training method, the precision (*P*) and the recall (*R*) of the improved model are increased by 3.42% and 15.18%, respectively; the average precision (*AP*) is increased by 8.15%, the F_1 score is increased by 10.29%, and the training time reduced by 1.14 h. Therefore, the transfer learning method can improve the recognition effect and reduce the training time.

Table 2

Test results of different training methods

Transfer learning	*P*/%	*R*/%	*AP*/%	F_1/%	Training time /h
Yes	88.33	85.97	91.18	87.13	10.29
No	84.91	70.16	83.65	76.84	11.43

The recognition effect of using different loss functions

The YOLOv3-tiny original loss function and the CIoU regression loss function based on the improved formula (3) are used to train the seed potato dataset. The result is shown in Fig.7. It can be seen that the loss value decreases rapidly in the early stage of training, and with the same downward trend. This is because in the early stage of training, using transfer learning and a larger learning rate for calculation, the model can jump out of the local optimal solution and make the loss value drop to a lower level. In the second stage of training, a smaller learning rate helps to ensure the stability of the model, and the change of loss value tends to be stable. Through the early stopping mechanism, the loss value stabilizes after 97 Epochs based on the original loss function model, and the final loss value is about 4.5. The loss value based on the CIoU regression loss function tends to be stable after 103 Epochs. The final loss value is 3.1, which is lower. It shows that the improved YOLOv3-tiny loss function improves the convergence effect compared to the original loss function in the training process.

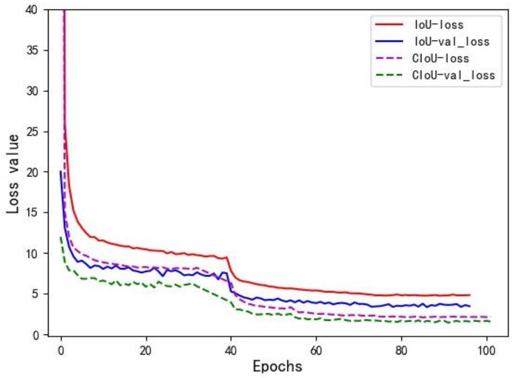


Fig. 7 – Loss curves of two YOLO models

The original loss function and CIoU regression loss function are used for recognition on the seed potato test set. The results are shown in Table 3. In test group A, the average precision *(AP)* and the F_1 score of using the CIoU regression loss function compared with using the original loss function for recognition are improved by 1.44% and 4.82% respectively. In test group B, the average precision *(AP)* and the F_1 score of using the CIoU regression loss function are improved by 3.11% and 4.65% respectively. Therefore, the model can be trained with the CIoU regression loss function. In addition, Mosaic data enhancement is added to the training process, and group B uses the Mosaic data enhancement method for training. According to the analysis in Table 3, based on the same loss function, the average precision *(AP)* and the F_1 score of the B-1 group enhanced with Mosaic data compared with the A-1 group without this method are respectively increased by 0.42% and 3.86%. Compared with the A-2 group, the average precision *(AP)* and the F_1 score of the B-2 group increased by 2.09% and 3.69%, respectively.

Through the above experiments, it can be obtained that the improved model with the Mosaic data enhancement method still maintains a high average precision (*AP*) and F_1 score for the recognition of seed potato buds.

Table 3

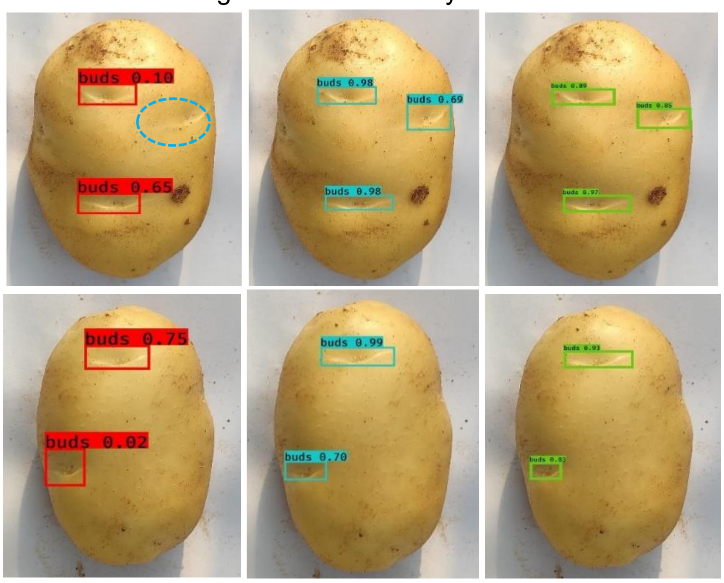
Test results of different test methods of the model

Test group number	Intersection over Union	Loss/Val_Loss	*P* [%]	*R* [%]	*AP* [%]	F_1 [%]
A-1	IOU	4.866/5.072	84.99	73.14	87.65	78.62
A-2	CIOU	3.251/3.168	91.27	76.84	89.09	83.44
B-1	IOU	5.212/4.027	86.73	78.62	88.07	82.48
B-2	CIOU	3.191/3.328	88.33	85.97	91.18	87.13

Buds recognition generalization test

To establish the optimal recognition model of YOLOv3-tiny seed potato buds, K-means clustering anchor boxes, CIoU regression loss function, and Mosaic data enhancement method are used. The number of Epochs in the first and second stage is selected to be 40 and 120, respectively, and batch size is 32. The IoU threshold is 0.45, and the Score threshold is 0.01. The recognition result shows that the precision (*P*) is 88.33%, the recall (*R*) is 85.97%, the average precision (*AP*) is 91.18%, and the F_1 score is 87.13%.

To further test the generalization of the proposed model, Fig. 8 shows the recognition and visualization results of the proposed improved YOLOv3-tiny model in different forms of the seed potato dataset. It can be seen that in different samples, the improved model can accurately recognize the seed potato buds. The YOLOv3-tiny has low accuracy and missed recognition in the recognition process (corresponding to the blue dotted ellipse in the figure). The improved YOLOv3-tiny model is slightly inferior to the YOLOv3 model in the confidence score of the recognition target of seed potato buds, but it can achieve more complete recognition. Therefore, the improved model has better generalization ability and robustness.

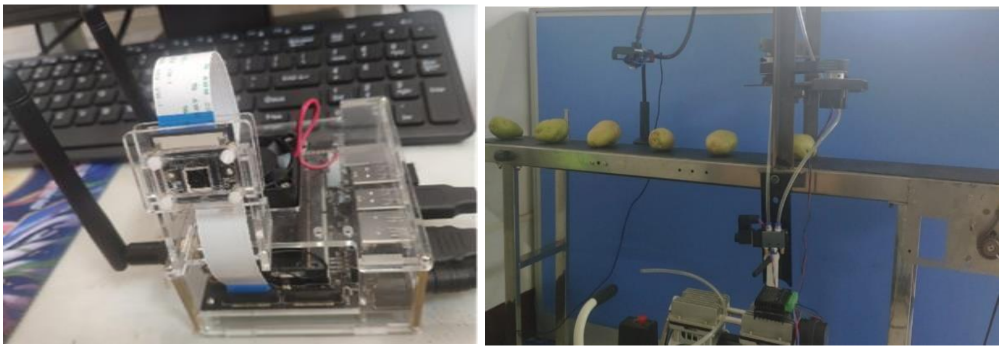


a) b) c)

Fig. 8 – Generalization test results of different models

a) YOLOv3-tiny; b) YOLOv3; c) Improved YOLOv3-tiny

To verify the recognition speed of the method proposed in this paper on the embedded platform, the trained model has been transplanted to the NVIDIA Jetson Nano, as shown in Fig. 9. Results show that the recognition speeds can reach 40FPS, so the method is also suitable for seed potato buds recognition at the embedded platform.



a) *b)*

Fig. 9 – Real-time recognition of seed potato buds on the embedded platform
a) NVIDIA Jetson Nano; b) Test bench

CONCLUSIONS

This paper proposes a method of seed potato buds recognition based on improved YOLOv3-tiny. Firstly, use imgaug data enhancement of the collected samples to prevent overfitting. Then the K-means clustering based on IoU is used to obtain anchor boxes that meet the size of the buds. Finally, based on the YOLOv3-tiny network framework, transfer learning, Mosaic data enhancement, and CIoU regression loss function are introduced in the training process to realize real-time recognition of seed potato buds.

Results show that the improved model has an average precision (*AP*) of 91.18%, and a recognition speed of 40FPS for the recognition of seed potato buds. Therefore, it can meet the real-time recognition requirements of seed potato buds, which lays the foundation for the subsequent intelligent cutting.

ACKNOWLEDGEMENT

This work was supported by Shandong Province Potato Industry Technology System Agricultural Machinery Post Expert Project (SDAIT-16-10), China Postdoctoral Science Foundation Funded Project (2020M681690), and Shandong Provincial Natural Science Foundation (ZR2019BC018).

REFERENCES

[1] Coates, A., Ng, A. Y., (2012), Learning Feature Representations with K-means, *Lecture Notes in Computer Science*, Vol. 7700, pp. 561-580, Berlin/Germany.

[2] Li, T., Yang, M., Zhang, X., et al., (2020), Potato Production Technology and Difference of Mechanization Level in Northwest China (西北地区马铃薯机械化生产技术与差异性研究), *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 51, Issue s1, pp. 307-313, Beijing/China.

[3] Li, W., Zhang, Y., Mo, J., et al., (2020), Detection of Pedestrian and Agricultural Vehicles in Field Based on Improved YOLOv3-tiny (基于改进 YOLOv3-tiny 的田间行人与农机障碍物检测), *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 51, Issue s1, pp.1-8, Shanghai/China.

[4] Li, Y., Li, T., Niu, Z., et al., (2018), Potato bud eyes recognition based on three-dimensional geometric features of color saturation(基于色饱和度三维几何特征的马铃薯芽眼识别), Transactions of the Chinese Society of Agricultural Engineering *(Transactions of the CSAE)*, Vol. 34, Issue 24, pp. 158 -164, Shandong/China.

[5] Li, Z., Wen, X., Lv, J., et al., (2019), Analysis and prospect of research progress on key technologies and equipment of mechanization of potato planting (马铃薯种植机械化关键技术与装备研究进展分析与展望), *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 50, Issue3, pp. 1-16, Hebei/China.

[6] Liang, C., Xiong, J., Zheng, Z., et al., (2020), A visual detection method for nighttime litchi fruits and fruiting stems, *Computers and Electronics in Agriculture*, Vol. 169, Issue 6, pp. 105192-105192, Guangdong/China.

[7] Liu, F., Liu, Y., Lin, S., et al., (2020), Fast Recognition Method for Tomatoes under Complex Environments Based on Improved YOLO (基于改进型 YOLO 的复杂环境下番茄果实快速识别方法), *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 51, Issue 6, pp. 229-237, Beijing/China.

[8] Lv, J., Yang, X., Li, Z., et al., (2020), Design and Test of Seed Potato Cutting Device with Vertical and Horizontal Knife Group (纵横刀组协同式马铃薯种薯切块装置设计与试验), *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 51, Issue 8, pp. 89-97, Heilongjiang/China.

[9] Lv, S., Lu, S., Li, Z., et al., (2019), Orange recognition method using improved YOLOv3-LITE lightweight neural network (基于改进 YOLOv3-LITE 轻量级神经网络的柑橘识别方法), *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, Vol. 35, Issue 17, pp. 205-214, Guangdong/China.

[10] Lv, Z., Qi, X., Zhang, W., et al., (2021), Buds Recognition of Potato Images Based on Gabor Feature (基于 Gabor 特征的马铃薯图像芽眼识别), *Agricultural Mechanization Research*, Vol. 43, Issue 2, pp. 203-207, Shandong/China.

[11] Luo, Y., Wang, X., (2018), Analysis of China's potato production efficiency based on DEA model, *China Agricultural Abstracts-Agricultural Engineering*, Vol. 30, Issue 4, pp. 13-18, Hebei/China.

[12] Pan, S., Yang, Q., (2010), A Survey on Transfer Learning, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 22, Issue 10, pp. 1345-1359, Hong Kong/China.

[13] Ren, S., He, K., Girshick, R., et al., (2017), Faster R-CNN: Towards real-time object detection with region proposal networks, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, Vol. 39, Issue 6, pp. 1137-1149, Anhui/China.

[14] Wang, X., Zhu, S., Li, X., et al., (2020), Transactions of the Chinese Society for Agricultural Machinery (定向排列纵横切分马铃薯种薯切块机设计与试验), *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 51, Issue 6, pp. 334-345, Shandong/China.

[15] Xi, R., Jiang, K., Zhang, W., et al., (2020), Recognition Method for Potato Buds Based on Improved Faster R-CNN (基于改进 Faster R-CNN 的马铃薯芽眼识别方法), *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 51, Issue 4, pp. 216-223, Shandong/China.

[16] Xu, N., Zhang, H., Zhang, R., et al., (2021), Current Situation and Prospect of Potato Planting in China (中国马铃薯种植业现状与展望), *Chinese Potato Journal*, Vol. 35, Issue 1, pp. 81-96, Heilongjiang/China.

[17] Zhang, J., Yang, T., (2020), Potato sprout eye recognition based on LBP and SVM (基于 LBP 与 SVM 的马铃薯芽眼识别), *Journal of Shandong Agricultural University (Natural Science Edition)*, Vol. 51, Issue 4, pp. 744-748, Gansu/China.