AN IMPROVED YOLOV4 METHOD FOR RAPID DETECTION OF WHEAT EARS IN THE FIELD
/
一种改进型YOLOv4的田间麦穗快速检测方法

Zongwei JIA *[1)], Yi SHAO [2)], Yijie HOU [1)], ChenYu ZHAO [1)], ZhiChuan WANG [1)], Yiming HOU [3)], Jinpeng QIN [1)][1]
[1)] College of Information Science and Engineering, Shanxi Agricultural University, Taigu / China
[2)] School of Software, Shanxi Agricultural University, Taigu / China
[3)] School of hydraulic and Ecological Engineering, Nanchang Institute of Technology, Nanchang / China
Tel:13835441286; E-mail: jiazw@sxau.edu.cn
DOI: https://doi.org/10.35633/inmateh-69-17

Keywords: *Field wheat ears, MobileNet, Deep separable convolution, Alpha*

ABSTRACT
The automatic detection of wheat ears in the field has important scientific research value in yield estimation, gene character expression and seed screening. The manual counting method of wheat ears commonly used by breeding experts has some problems, such as low efficiency and high influence of subjective factors. In order to accurately detect the number of wheat ears in the field, based on MobileNet series network model, deep separable convolution module and alpha channel technology, the YOLOv4 model is reconstructed and successfully applied to the task of wheat ear yield estimation in the field. The model can adapt to the accurate recognition and counting of wheat ear images in different light, viewing angle and growth period. At the same time, the model volume with different alpha parameters is more suitable for mobile terminal deployment. The results show that the parameters of the improved YOLOv4 model are five times smaller than the original model, the average detection accuracy is 76.45%, and the detection speed FPS is two times higher than the original model, which provides accurate technical support for rapid yield estimation of wheat in the field.

摘要
田间麦穗的自动检测在产量估计、基因性状表达及种子筛选等方面都具有较为重要的科学研究价值，育种专家常用的麦穗人工计数方法存在效率低，主观因素影响较高等问题。为了精确检测田间麦穗数量，本文基于MobileNet系列网络模型，深度可分离卷积模块及Alpha通道等技术，重构了YOLOv4模型并成功应用到田间麦穗估产任务上，模型能够适应不同光照，视角，不同生长时期麦穗图像的准确识别和计数，同时采用不同Alpha参数的模型体积更加适应移动终端部署。结果表明，改进型YOLOv4模型参数量较原始模型体积缩小了5倍，平均检测精度达76.45%，检测速度FPS比原模型提升了2倍，为田间小麦快速估产提供准确的技术支撑。

INTRODUCTION
As one of the important food crops in the world, wheat has an annual output of 700 million tons *(http://faostat3.fao.org/faostat-gateway/go/to/browse/Q/QC/E)*. With the global climate change, natural disasters, regional economic fluctuations and other factors, the global food demand and price are increasing year by year, and the food crisis facing mankind is becoming more and more serious. This poses a higher challenge to agricultural breeding experts. Human beings need more cold resistant and high yield crop varieties to increase crop yield and improve crop tolerance to biological and environmental stresses. In order to find new varieties with high yield and stress tolerance, many breeding experts increasingly rely on high-throughput phenotypic technology to measure different crop traits, so as to understand the response and adaptability of crops to the surrounding environment, hoping to improve crop yield and quality. The phenotypic development technology of high-throughput images has been experienced for nearly a decade. In the early stage, it was mostly through special laboratories or greenhouses. *Li et al., (2017),* proposed a method to detect, count and measure the panicles of plants under the automatic control environment. *Bi et al. (2010)* and *Bi et al. (2011)* measured the different morphological characteristics of potted plants through the plant imaging device in the special laboratory, but the effect of such potted plants on water and nutrient absorption is greatly affected by the laboratory environment.

[1] *Zongwei Jia*, A.P. M.; Yi Shao, M. Stud.; Yijie Hou, M. Stud.; Chenyu Zhao, M. Stud.; Zhichuan Wang, M. Stud.; Yiming Hou, M. Stud; Jinpeng Qin, M. Stud.*

The results are quite different from the field natural environment data. In recent years, a series of large-scale plant phenotype image capture systems *(Perez-Sanz et al., 2017; Araus et al., 2014; Montes et al., 2007)*, UAVs *(Holman et al., 2016; Khan et al., 2018; Shi et al., 2016; Madec et al., 2017)* and satellite remote sensing images *(Azzari et al., 2015; Lobell et al., 2015)* have emerged. These methods can capture a wide range of crop phenotype information, but most of them can only carry out rough statistical information, such as average canopy coverage and average canopy color characteristics. It is worth noting that the field situation is often affected by the environment such as weather, light and wind speed, which brings great difficulties to image acquisition and image analysis algorithms. In order to adapt to the changing experimental conditions and improve the robustness of the algorithm, researchers usually need to manually analyze a large amount of data.

In this paper, tripod and single RGB camera are used to obtain real field crop images, which can capture high-resolution crop phenotype images. At the same time, the imaging device is simple, easy to operate and cost-effective, which is also convenient for the rapid implementation of breeding experiments of various scientific research groups. The challenge of this paper is how to extract quantitative information from these high-resolution images, mainly the number and density of wheat ears. The field data acquisition perspective of this paper selects oblique top photography instead of the common head up angle photography. Under the oblique angle, more phenotypic information can be obtained, involving the texture, color, shape and other information of wheat ears. These traits can provide data support for the application of wheat phenotype, such as ear count, ear area, ear texture, disease monitoring, yield estimation and so on. Using a single RGB image to obtain data and study the phenotype of machine vision technology has been favored by most breeders.

Fernandez-Gallego et al. (2018) used RGB camera to obtain the image from the head up angle of the top of the plant canopy, and then processed it with Laplace operator and median filter algorithm to identify the local maximum as wheat ears, with an accuracy of 92%, but for wheat ears in different growth stages, the algorithm recognition rate is quite different. *Alharbi et al. (2018)* used Gabor filter, principal component analysis and K-means clustering technology, and the average recognition rate of wheat ear is 90.7%. However, this method is seriously limited by ear density, ear and straw color, texture and ear angle. *Zhou et al. (2018)* et al. used multispectral cameras to collect data and used the improved maximum entropy segmentation algorithm to identify wheat ears in the field, which achieved good results when the wheat ear density and wheat ear occlusion were small.

In recent years, deep learning is gradually surpassing the previous image analysis and machine learning methods, and has brought significant changes in the field of plant phenotype analysis, especially the rapid application of convolutional neural network in image analysis tasks, such as in agricultural crop organ detection tasks based on target detection models such as YOLO-v3 *(Gao et al., 2019)*, mask RCNN, Alex net *(Khan et al., 2018)*, YOLO-v4 *(Gong et al., 2020)*. Extensive research has been carried out to solve agricultural science problems such as plant leaf classification *(Wilf et al., 2016)*, root structure analysis *(Kumar et al., 2014; Kumar et al., 2015)*, plant stress degree measurement *(Singh et al., 2016)*, determination of crop growth period *(Sadeghi-Tehran et al., 2017)*. This paper presents a depth learning model designed to quickly detect and count the number of wheat ears in the field wheat ear dense image. This method relies on the wheat ear image training data set manually marked with rectangular box, and outputs a complete boundary box position and size information to detect the number of wheat ears in the unit area, so as to realize wheat yield estimation. Field wheat yield estimation is an important data in wheat breeding. The combination of deep learning theory and field wheat ear counting task has strong application value and practical significance.

MATERIAL AND METHODS

DATA ACQUISITION AND PREPROCESSING

Data sources

The experimental wheat ear data in this paper are from the global wheat target detection competition (gwhd)*(https://www.kaggle.com/c/global-wheat-detection.)* and the wheat experimental base of Shanxi Agricultural University (sxau_wheat). The images of the sxau_wheat dataset are all wheat from April to June. At this time, the wheat ears are relatively full, most of them are still in an upright state, and the color is more prominent. Finally, 11 wheat ear image subsets were obtained, a total of 3487, with a resolution of 1024 × 1024 field wheat ear image.

Data enhancement

In order to better improve the generalization ability and robustness of the model, data enhancement on the original data set is performed, which is divided into offline enhancement and online enhancement.

1) For offline enhancement, set the random chromaticity interval of wheat ear image as (0.4-2.6), contrast interval as (0.6-1.6) and sharpness change interval as (0.4-4.0) for enhancement, and randomly apply enhancement operations such as salt and pepper noise, Gaussian blur, rotation, horizontal flip, vertical flip and zoom. Fig. 1 shows the comparison of enhancement effect of wheat ear data.

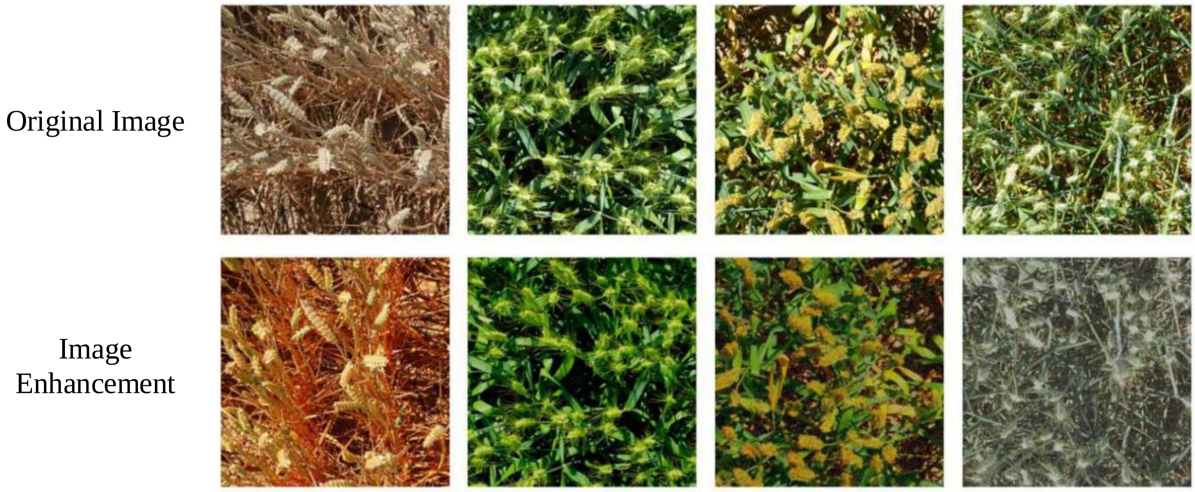

Fig. 1 - Contrast Map of Wheat Spike Image Enhancement Effect

2) Online image enhancement is carried out by turning, brightness adjustment, contrast, random clipping and scale transformation. In order to take into account the corresponding changes of image scale transformation and image file and annotation file after random clipping, a random online preprocessing function get for real-time data enhancement is designed in this get_random_data, through the online enhancement before each epoch to further improve the robustness of the model. Fig. 2 shows an example of online enhancement of wheat ear image.

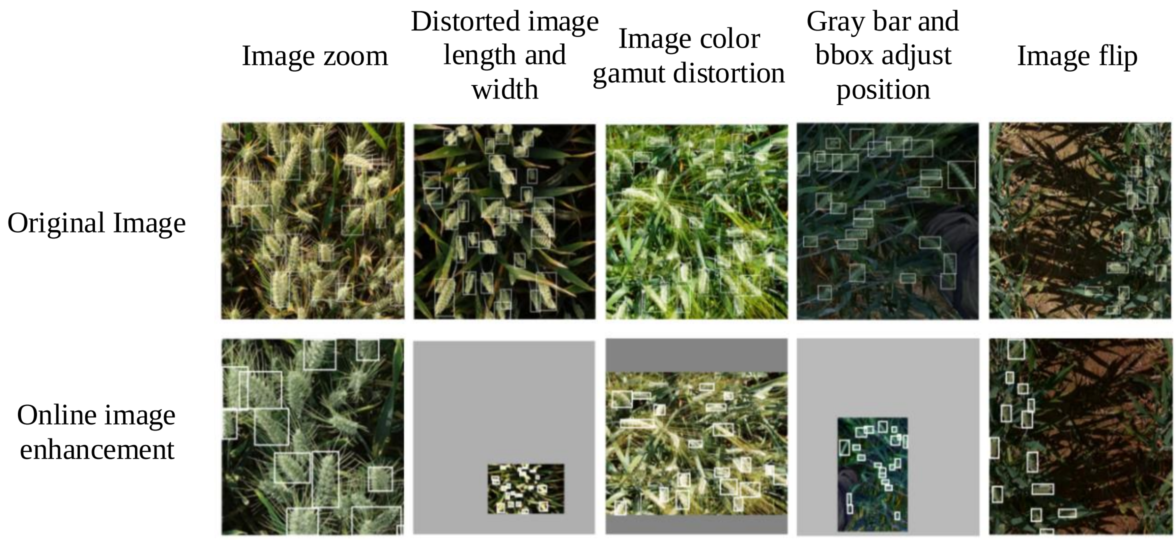

Fig. 2 - Example of Online Enhancement of Wheat Spike Image

OPTIMIZATION MODEL CONSTRUCTION

Backbone feature extraction and network replacement

YOLOv4 is one of the current YOLO series models with high accuracy, but its backbone feature extraction network CSPdarknet53 has the characteristics of large amount of parameters and slow model convergence. In this paper, the MobileNet series network is selected to replace the original CSPdarknet53 network *(Howard et al., 2017)*, and the specific selection of Mobilenetv1, Mobilenetv2 Mobilenetv3 is one of the three models to obtain effective features. The shape sizes of the three effective feature layers feat1, feat2 and feat3 corresponding to the three models are shown in Table 1.

Table 1

Effective Feature Sizes Corresponding to Different Replacement Models

Model	Feat1 shape	Feat2 shape	Feat3 shape
MobileNetV1	(52, 52, 256)	(26, 26, 512)	(13, 13, 1024)
MobileNetV2	(52, 52, 32)	(26, 26, 92)	(13, 13, 320)
MobileNetV3	(52, 52, 40)	(26, 26, 112)	(13, 13, 160)

MobileNet series models are selected as the backbone feature extraction network to replace CSP-Darknet53 network. The reconstructed models are named mnv1-YOLOv4, mnv2-YOLOv4 and mnv3-YOLOv4 respectively.

At this time, the replacement of the YOLOv4 backbone feature extraction network is completed, as shown in Table 2. After analyzing the size of the replaced model parameters, it can be found that the number of model parameters is reduced by nearly 30% compared with the original YOLOv4 model, effectively improving the detection speed of the model.

Table 2

Number of Model Parameters after Backbone Network Replacement

Model name	Parameter quantity
YOLOv4	64,429,405
MNV1-YOLOv4	41,005,757
MNV2-YOLOv4	39,124,541
MNV3-YOLOv4	40,043,389

PANet enhances feature extraction and network optimization

For the PANet enhanced feature extraction network, its parameters are mainly concentrated on many ordinary 3 × 3 convolution. If these ordinary 3 × 3 convolution can be modified, the parameter quantity and volume of the whole model can be greatly reduced.

In this paper, the deep separable convolution block *(Sandler et al., 2018)* in the MobileNet network is used to replace the ordinary 3 used in the enhanced feature extraction network PANet in the original YOLOv4 × 3, the overall parameters of the model are further reduced, and the enhanced extraction of image edge features by the model is also improved. At this time, the YOLOv4 network structure after two-step optimization is shown in Figure 3.

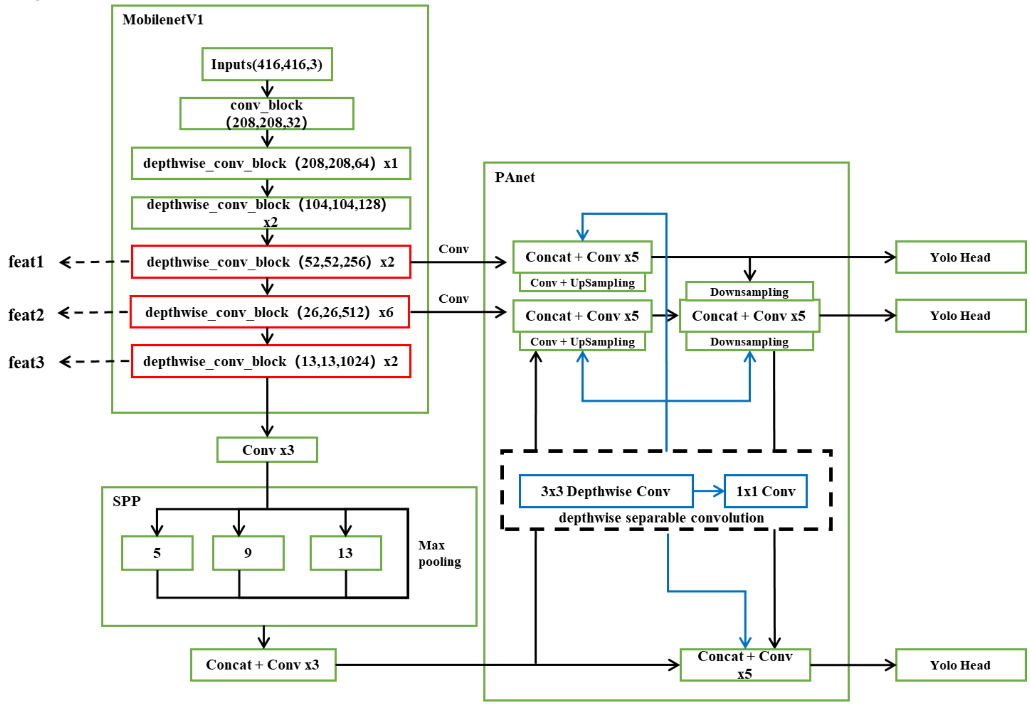

Fig.3 - Optimized YOLOV4 Network Structure

As shown in Table 3, by comparing and analyzing the parameters of the optimized PANet model, it can be seen intuitively that the parameters of the optimized PANet model are nearly 70% less than the total parameters of the model after the backbone feature extraction network is replaced, which basically meets the deployment requirements of mobile devices.

Table 3

Parameter Quantities after PANet Model Optimization

Model name	Model parameters after backbone network replacement	Model parameters after optimizing PANet
MNV1-YOLOv4	41,005,757	12,754,109
MNV2-YOLOv4	39,124,541	10,872,893
MNV3-YOLOv4	40,043,389	11,791,741

Alpha parameter channel adjustment

In order to further reduce the volume of the model, this paper modifies the number of channels by adjusting the alpha parameter value, and sets the alpha parameters for the number of channels such as 128, 256, 512 and 1024 in the model, replacing (128 * alpha) with 128 channels, replacing (256 * alpha) with 256 channels, replacing (512 * alpha) with 512 channels and replacing (1024 * alpha) with 1024 channels to complete the replacement operation. Next, you can directly set and change the alpha parameter according to your needs. The smaller the alpha parameter setting, the fewer channels of the whole model will be.

Through the replacement of backbone feature extraction network and PANet, strengthen the modification of feature extraction network and the application and setting of alpha parameters. Select the backbone feature extraction network to be used in mobilenetv1, mobilenetv2 and mobilenetv3, which is represented by backbone parameters. The alpha parameter is 1 by default (setting the alpha parameter to 1 means that the number of channels of the whole model has not been adjusted). The optional range is set to 0.25, 0.5, 0.75 and 1.0. Continuous comparison is made within this range. The comparative analysis of the parameters of the optimized and improved model is shown in Table 3. The total parameters of the original YOLOv4 model are 64429405. Through the reconstruction and optimization of the YOLOv4 model, the changes of the parameters of the whole model are shown in Table 4.

Table 4

Comparative Analysis of the Number of Model Parameters after Improvement and Optimization

Model name	Model parameters after backbone network replacement	Model parameter quantity after optimizing PANet (Alpha parameter is 1.0 by default)	Model parameter quantity with Alpha parameter = 0.75	Model parameter quantity with Alpha parameter = 0.5	Model parameter quantity with Alpha parameter = 0.25
MNV1-YOLOv4	41005757	12754109	7299789	3356125	680381
MNV2-YOLOv4	39124541	10872893	6257773	2370541	476568
MNV3-YOLOv4	40043389	11791741	6315309	2823754	568156

Evaluating indicator

For the wheat ear detection model, the most important is the AP value, FPS and the size of model parameters. The AP value mainly reflects the relationship between the accuracy and recall of the model. The higher the AP value, the better the overall detection effect of the model. FPS represents the image detection speed, and the model parameter refers to the size of the parameters in the internal network of the model.

(1) AP (average precision) and PR curve

Because the object category detected in the whole experiment is only wheat, mAP is not used as an index in the model test effect, but AP value is used to measure the performance of target detection algorithm. AP value refers to the average accuracy rate of a certain category. AP value is the combination of different precision and recall points. The larger the area under the PR curve, the larger the AP value, and the higher the average accuracy rate of the model, the better the performance of the model and the better the effect of the model in detecting wheat ears.

AP value is defined as equation (1).

$$AP = \int_0^1 P(R)dR = \sum_{k=0}^{n} P(k)\Delta R(k) \tag{1}$$

(2) TP and FP

TP is used to represent the number of positive samples predicted by the model and actually positive samples, and FP is used to represent the number of positive samples predicted by the model and actually negative samples.

(3) FPS (frames per second) image detection speed

FPS simply means that it can detect the number of pictures in one second. FPS can well evaluate the real-time performance of model detection, in seconds. Generally, the speed is measured by the number of frames of input video processed in one second. The higher the FPS, the better the real-time performance of the trained model.

(4) Model size and parameter quantity

The unit of model size is usually expressed in megabytes (MB), which is also one of the important factors of wheat ear target detection and evaluation indicators. The target detection model includes backbone network architecture information, optimization information and parameter information. If the model is too large to affect the portability of transplantation, the unit of parameter is usually expressed in "pieces".

TRAINING AND EVALUATION

The field wheat ear enhancement data set selected in this paper has a total of 5632 wheat ear images, all of which are labeled with labelimg to generate a VOC format data set *(Everingham et al., 2010)* , in which the number of bounding box wheat ears is 242309.

The data set is divided into training set, test set and verification set according to 8:1:1, which are 4505, 563 and 564 respectively. The wheat ear detection model was trained and tested on NVIDIA geforce RTX 2080 Ti GPU.

Model training

Set the number of iterative training to 100 rounds, of which the first 50 rounds are frozen training. Freeze some features to extract the weight of the network, so that the network weight of the frozen part is not updated, and other parts of the network are trained first. After 50 rounds of thawing, the network layer to be trained shall be trained. Freeze the number of pictures entered in each training batch_size is set to 8 and the learning rate is 1e-3. Unfreeze training number of pictures input each time batch_size is set to 6 and the learning rate is 5e-5. The training learning rate adopts the adjustment strategy of linear attenuation, that is, if the loss value does not decrease after 5 epochs, the learning rate becomes one-half of the original.

During the training process, the model is constantly updated and adjusted, and the parameters are set to save the best model every 5 epochs. Finally, after continuous iterative training, the model structure and parameters are saved to obtain the wheat ear real-time detection model. If the loss value of the model does not decrease under 10 consecutive epochs, it indicates that the model has fitted the data, and the model will automatically stop training.

RESULTS AND ANALYSIS

It can be seen from Figure 4 that the five models can detect wheat ears in different growth periods, varieties, light and dark, and the confidence values are not different. When the alpha coefficient is adjusted to 0.25, the effect of the prediction model is the worst, and there are many missed errors, which is directly due to the lack of model parameters.

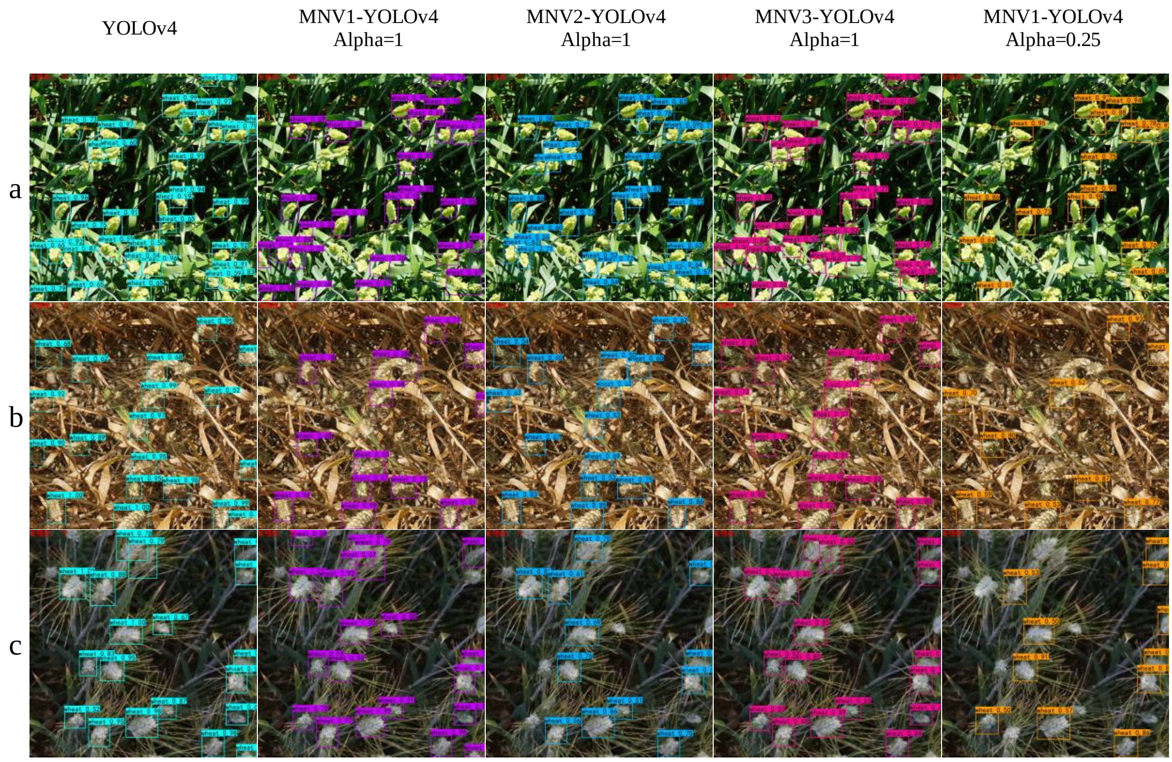

Fig.4 - Some Forecast Results

Figure a shows the test results of wheat ears under insufficient light; Figure b shows the test results of wheat ears under sufficient light; Figure c shows the detection results of dark light and only part of wheat ears in the image

Table 5

Comparative Analysis of Wheat Spike Detection Models

Target detection Model	Average accuracy AP（%）	Image detection speed FPS（Frame/s）	Model size（MB）	Model parameter quantity（PCs.）
YOLOv4	76.74	13.03	751.2	64,429,405
MNV1-YOLOv4-Alpha=1	74.83	22.52	145.1	12,754,109
MNV2-YOLOv4-Alpha=1	73.65	23.64	110.9	10,872,893
MNV3-YOLOv4-Alpha=1	76.45	24.82	124.2	11,791,741
MNV1-YOLOv4-Alpha=0.25	62.8	44.69	10.4	680,381

It can be seen from table 5 that the AP value of YOLOv4 detection model is higher than that of the improved model, which is 1.91, 3.09 and 0.29 percentage points higher than mnv1-YOLOv4-alpha = 1, mnv2-YOLOv4-alpha = 1, mnv3-YOLOv4-alpha = 1 and mnv1-YOLOv4-alpha = 0.25 respectively.

On the premise of sacrificing part of the accuracy, the detection speed and model parameters of the model have been highly optimized and improved. Mnv1-YOLOv4-alpha = 0.25 model pursues the ultimate detection speed and minimal volume, resulting in low AP of the model, but it provides a reference for the transplantation of the model to embedded device terminals in the future.

The optimized model has been greatly optimized in terms of detection speed and model volume, and the accuracy sacrifice is small.

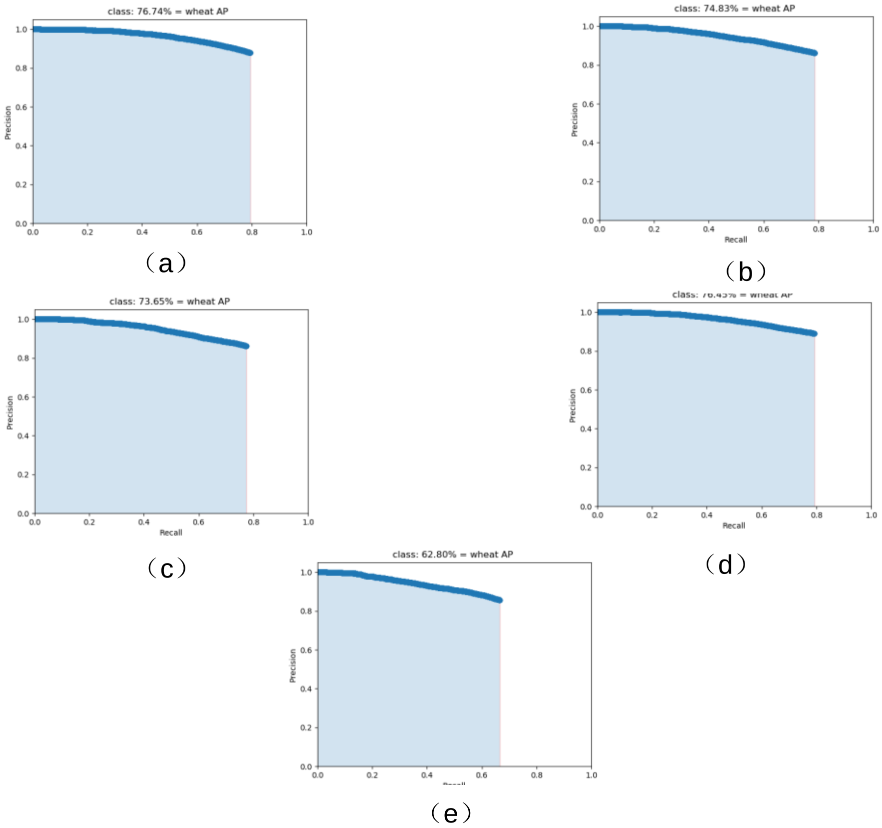

Fig. 5 - P-R Curve

Fig. 5 (a-e) shows the P-R curves of YOLOv4, MNV1-YOLOv4-Alpha = 1, MNV2-YOLOv4-Alpha = 1, MNV3-YOLOv4-Alpha = 1 and MNV1-YOLOv4-Alpha = 0.25. The confidence of the model is set to 0.5 and the IOU is set to 0.3 to obtain the AP values of five wheat ear detection models.

Figure 6 shows the TP and FP results of five wheat ear detection models, which are obtained from 563 pictures in the test set. The number of frames manually labeled in 563 pictures is 13399, (a-e) respectively represents the TP / FP results of five models: YOLOv4, MNV1-YOLOv4-Alpha = 1, MNV2-YOLO4-Alpha = 1, MNV3-YOLOv4-Alpha = 1 and MNV1-YOLOv4-Alpha = 0.25.

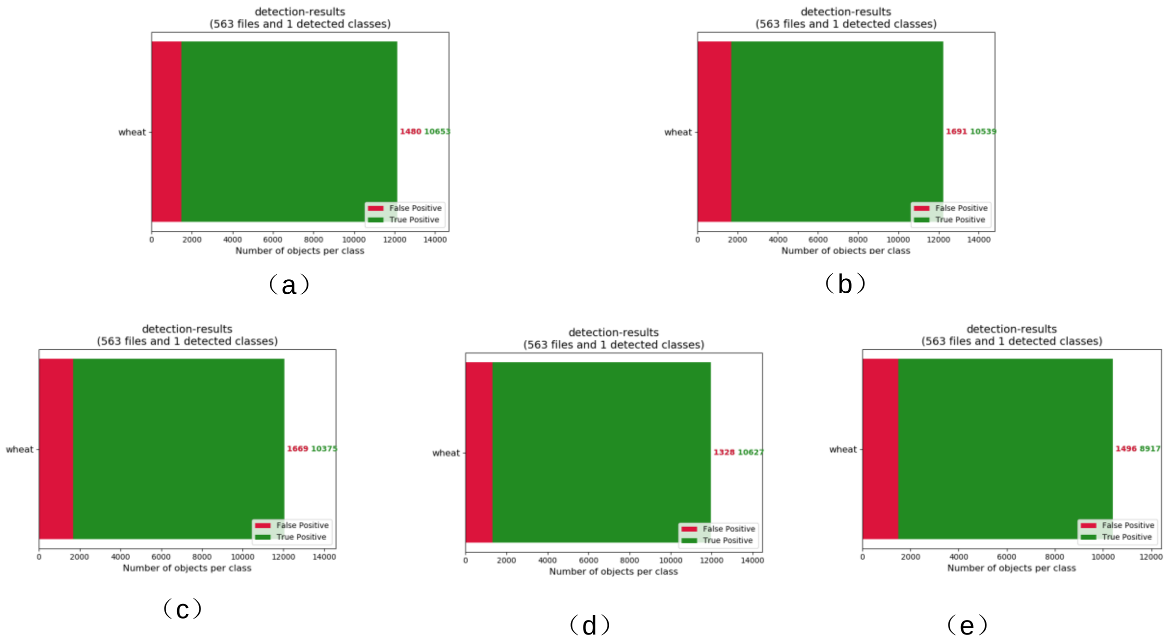

Fig. 6 - TP/FP of Wheat Spike Detection Model

It can be concluded from Fig.6 that except for figure (e), the TP values of other optimized models are similar, but the number of false detections of the model changes greatly. Among them, the number of false detections of mnv3-YOLOv4-alpha = 1 model is 152 less than that of the original YOLOv4 model, indicating that the performance of the optimized model has been effectively improved compared with the original YOLOv4 model.

CONCLUSIONS

In this paper, MobileNet series backbone network, deep separable convolution module and alpha channel number dynamic adjustment technology are used to improve and optimize the YOLOv4 model, and it is applied to the field wheat ear detection and counting task. When the detection accuracy is very close, the detection speed of the optimized detection model is increased by two times, and the volume of the model is reduced to one fifth. At the expense of small accuracy, the model can adapt to the rapid detection and counting tasks of wheat ears in the field under different illumination, angles and growth periods, provide support for agricultural breeding experts to estimate yield quickly and accurately, and provide possibility for the transplantation and deployment of mobile terminals.

ACKNOWLEDGEMENT

This research, titled ‘AN IMPROVED YOLOV4 METHOD FOR RAPID DETECTION OF WHEAT EARS IN THE FIELD’, was funded by Shanxi Province Basic Research Program (Free Exploration) Project(NO:202103021224149), Shanxi Province Postgraduate Education Teaching Reform Project (NO:2021YJJG087), Shanxi Province Educational Science "14th Five-Year Plan" Education Evaluation Special Project (NO:PJ-21001). The authors are grateful and honored to receive this support.

REFERENCES

[1] Alharbi N., Zhou J., Wang W., (2018), Automatic counting of wheat spikes from wheat growth images. In: *7th International Conference on Pattern Recognition Applications and Methods*, p. 346–355.

[2] Araus J.L., Cairns JE., (2014), Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci*, 19(1):52–61.

[3] Azzari G., Lobell DB., (2015), Satellite estimates of crop area and maize yield in Zambia’s agricultural districts. In: *Proceedings of the AGU fall meeting*.

[4] Bi K., Jiang P., Li L., Shi B., Wang C., (2010), Non-destructive measurement of wheat spike characteristics based on morphological image processing. *Trans Chin Soc Agric Eng*, 26(12):212–6.

[5] Bi K, Jiang P., Wei T., Huang F., Wang C., (2011), The design of wheat variety BP classifier based on wheat ear feature. *Chin Agric Sci Bull*, 28(6):464–8.

[6] Everingham M., Van Gool L., Williams C K I., et al. (2010), The Pascal Visual Object Classes (VOC) challenge[J]. *International journal of computer vision*, 88(2): 303-338.

[7] Fernandez-Gallego JA., Kefauver SC., Gutiérrez NA., Nieto-Taladriz MT., Araus JL., (2018), Wheat ear counting in-field conditions: high throughput and low-cost approach using RGB images. *Plant Methods*, 14(1):22.

[8] Gong B., Ergu D., Cai Y., et al. (2020), Real-Time Detection for Wheat Head Applying Deep Neural Network[J]. *Sensors*, 21(1):191.

[9] Gao Y.P., (2019), Study on field wheat ear detection method based on depth neural network[D]. Beijing: *Master Thesis of Beijing Forestry University*.

[10] Holman FH., Riche AB., Michalski A., Castle M., Wooster MJ., Hawkesford MJ., (2016), High throughput field phenotyping of wheat plant height and growth rate in field plot trials using UAV based remote sensing. *Remote Sens*, 8(12):1031.

[11] Howard A.G., Zhu M., Chen B, et al. (2017), Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861.

[12] Khan Z., Rahimi-Eichi V., Haefele S., et al. (2018), Estimation of vegetation indices for high-throughput phenotyping of wheat using aerial imaging[J]. *Plant Methods*, 14(1): 20.

[13] Kumar P., Cai J., Miklavcic SJ., (2015), A complete system for 3D reconstruction of roots for phenotypic analysis. *Adv Exp Med Biol*, 823:249–70.

[14] Kumar P., Huang C., Cai J., Miklavcic SJ., (2014), Root phenotyping by root tip detection and classification through statistical learning. *Plant Soil*, 380(1):193–209.

[15] Li Q, Cai J., Berger B., Okamoto M., Miklavcic SJ., (2017), Detecting spikes of wheat plants using neural networks with Laws texture energy. *Plant Methods*, 13(29046709):1–13.

[16] Lobell DB., Thau D., Seifert C., Engle E., Little B., (2015), A scalable satellite-based crop yield mapper. *Remote Sens Environ*, 164:324–33.

[17] Madec S., Baret F., deSolan B., Thomas S., Dutartre D., Jezequel S., Hemmerle M., Colombeau G., Comar A., (2017), High-throughput Phenotyping of Plant Height: Comparing Unmanned Aerial Vehicles and Ground LiDAR Estimates. *Front Plant Sci*, 8:2002.

[18] Montes JM., Melchinger AE., Reif JC., (2007), Novel throughput phenotyping platforms in plant genetic studies. *Trends Plant Sci*, 12(10):433–6.

[19] Perez-Sanz F., Navarro PJ., Egea-Cortines M., (2017), Plant phenomics: an overview of image acquisition technologies and image data analysis algorithms. *GigaScience*, 6(11):1–18.

[20] Sadeghi-Tehran P., Sabermanesh K., Virlet N., Hawkesford MJ., (2017), Automated method to determine two critical growth stages of wheat: heading and flowering. *Front Plant Sci*, 8(252):1–14.

[21] Sandler M., Howard A., Zhu M., et al. (2018), MobileNetV2: Inverted Residuals and Linear Bottlenecks[J]. *The IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 4510-4520.

[22] Shi Y., Thomasson JA., Murray SC., Pugh NA, Rooney WL., Shafian S., Rajan N., Rouze G., Morgan CLS., Neely HL., (2016), Others: unmanned aerial vehicles for high-throughput phenotyping and agronomic research. *PLoS ONE*, 11(7):1–26.

[23] Singh A., Ganapathysubramanian B., Singh AK., Sarkar S., (2016), Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci*, 21(2):110–24.

[24] Wilf P., Zhang S., Chikkerur S., Little SA., Wing SL., Terre T., (2016), Computer vision cracks the leaf code. *Proc Natl Acad Sci*, 113:3305–10.

[25] Zhou C., Liang D., Yang X., Xu B., Yang G., (2018), Recognition of wheat spike from field based phenotype platform using multi-sensor fusion and improved maximum entropy segmentation algorithms. *Remote Sens*, 10(2):246.

[26] *http://faostat3.fao.org/faostat-gateway/go/to/browse/Q/QC/E*

[27] *http://www.kaggle.com/c/global-wheat-detection*.