The Field Wheat Count Based on the Efficientdet Algorithm

Liangben Cao[a], Xixin Zhang[b], Jingyu Pu[c], Siyuan Xu[d], Xinlin Cai[e], Zhiyong Li[*]
Sichuan Agricultural University, Ya'an, China
*Corresponding author email:lzy@sicau.edu.cn, [a]caoliangben@163.com,[b]zhangxixin@stu.sicau.edu.cn,
[c]pujingyu@stu.sicau.edu.cn, [d]xsy6669@163.com, [e]CDDaff31415@163.com

***Abstract*—It is of great significance to realize the low cost and fast statistics of wheat ears to predict wheat yield. At present, a variety of methods have been applied to the measurement of wheat planting density, although these methods can count ears of wheat, they are expensive and difficult to be put into actual production. In order to further improve the accuracy of wheat ear identification and detection counting under field environment, based on the image processing and deep learning technology, EfficientDet algorithm is proposed to detect the wheat ear image. The idea is to frame out the wheat in the wheat ear image and then count the detected target number to realize the automatic counting of wheat ears. EfficientDet-D3 model is adopted in this paper, and adamw algorithm is used to train the model, with a learning rate of 1e5. The experimental results show that the model can quickly and accurately recognize wheat ear images with different densities under various lighting conditions. The final accuracy rate reaches 92.92%, and the test time of the single sheet is 0.2s. We carry out ear counting test on 20 wheat ear images with different densities, and the average accuracy rate reaches 95.30%. This method can detect wheat ear image in a very short time and the detection effect is good. To some extent, the accuracy of wheat ear counting has been increased, and has the potential to be put into actual production.**

Keywords—EfficientDet, BiFPN, wheat ear count

I. INTRODUCTION

The number of ears per unit wheat planting area is one of the important parameters for wheat yield prediction and planting density assessment in local area. At the moment, some scholars have applied image processing to wheat ear counting and have done some meaningful work on wheat yield prediction. Nowadays, many scholars have used traditional image processing methods to count ears of wheat, but these methods all require high cost, and the application scene has great limitations. The counting accuracy of the wheat and wheat ear counting method based on improved K-means proposed by Zhe Liu [6] reaches about 94%. These methods need a lot of time in the preliminary preparation work, and the detection speed is slow, so it is difficult to put them into actual production. In this paper, EfficientDet, a deep learning model, is used to detect the target of wheat ear image. With the result to be output, the number of wheat ears is output at the same time, which greatly shortens the detection time, and the detection accuracy reaches 92.92% while counting accuracy reaches 95.30%, which are improved to some extent.

II. ALGORITHM ANALYSIS

A. EffificientDet Architecture

The EfficientDet algorithm was proposed by the Google brain team. By means of improving the multiple dimensioned feature fusion structure of FPN and borrow ideas from the EfficientNet model scaling method for reference, it is a model scalable and efficient target detection algorithm.

EffificientDet consists of three parts. As shown in Figure 1, the first part is the pre-trained EfficientNet as the backbone network by ImageNet. The second part is BiFPN, it does the top-down and bottom-up feature fusion multiple times for the output characteristic of Level 3-7 in EfficientNet. The third part is the classification and detection box prediction network, to regress and classify the wheat ear frame. Modules in Part two and Part three can be repeated multiple times, depending on hardware conditions. EfficientNet-B3 is taken as the BackBone in this paper, featuring extraction of wheat ear image input into the network, with a small number of featuremap parameters, rich information can be extracted, which ensures the detection speed and accuracy to a certain extent. Then input P3-P7 to BiFPN for feature fusion. BiFPN adopts weighted feature fusion to obtain semantic information of different sizes. From the detection results, very small wheat ear information can also be extracted. Since EffificientDet is the target detection of anchor-based, the initial value of anchor should be adjusted appropriately to achieve better results. The highly accurate version of The Efficientdet-D7 achieves the highest accuracy in the published paper: 51.0map on the COCO data set. Compared to the previous best algorithm, the number of arguments was reduced fourfold, the flop was 9.3 times smaller and the accuracy was higher (+0.3%mAP).

978-1-7281-8304-6/20/$31.00 ©2020 IEEE

Dalian, China
September 27-29, 2020

Authorized licensed use limited to: Oestfold University College. Downloaded on February 20,2023 at 18:15:34 UTC from IEEE Xplore. Restrictions apply.

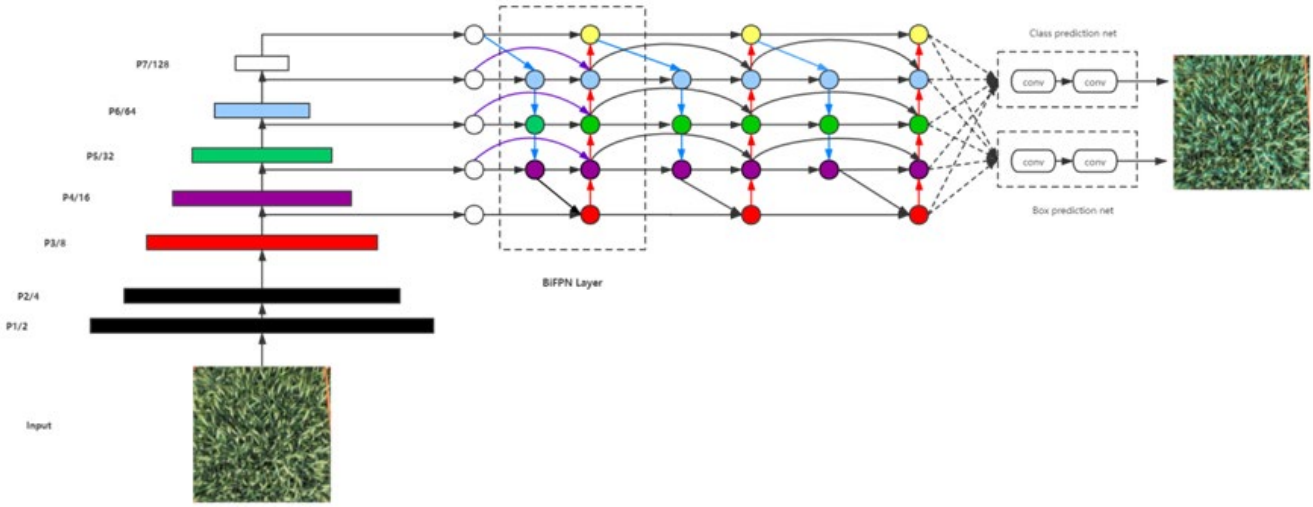

Fig. 1. EfficientDet architecture

B. BiFPN

Traditional FPN gathers multi-scale characteristics from top to bottom, as shown in Figure 2 (a), the output feature map of layer 7 is obtained after a convolution of the input feature map of layer 7. The output feature map of layer 6 can be obtained by convolving the fusion feature map obtained by up-sampling the output feature map of layer 7 and adding the input feature map of layer 6. And so on, the output feature map of layer 3 can be obtained by convolving the fusion feature map obtained by up-sampling the output feature map of layer 4 and adding the input feature map of layer 3. The conventional FPN aggregates multi-scale features in a top-down manner:

P out 7 = Conv (P in 7)

P out 6 = Conv (P in 6 + Resize (P out 7))

...

P out 3 = Conv (P in 3 + Resize (P out 4))

Where Resize is usually a upsampling or downsampling op for resolution matching, and Conv is usually a convolutional op for feature processing.

As shown in Figure 2(b), PANet improves FPN's feature fusion method, adopting not only the top-down method, but also the bottom-up method. Figure 2(c) shows that an irregular characteristic network topology, the NAS-FPN, is found by applying neural architecture search (NAS) method. To obtain this result, a large amount of GPU computing time is needed.

The BiFPN shown in Figure 2 (d) is optimized in three ways: (1) If a node (feature map) has only one input, its contribution to feature fusion is small and can be deleted. (2) Jump connections are established at each level (level) to incorporate more features at little cost. (3) It is proposed that each BiFPN can be splicing as a module, and the output of the previous BiFPN can be used as the input of the next BiFPN. The specific number of such structures required depends on the situation.

C. Three feature fusion methods

When the features of different scales are fused, the common practice is to unify the scales first, and then add the corresponding features. This assumes that the weight of different features to the final fused feature is the same. In fact, different input features should contribute differently to the final fusion feature due to their different resolution. Therefore, three weighted feature fusion methods are proposed:

Unbounded fusion: $O = \sum_i w_i \cdot I_i$, where w_i is a learnable weight that can be a scalar (per-feature), a vector (per-channel), or a multi-dimensional tensor (per-pixel). The scalar weight is the least expensive to calculate without obvious loss of precision, so the scalar weight is adopted. The disadvantage of the weighting method is that there is no constraint on the weight, so the training of the model may be unstable.

Softmax-based fusion: $O = \sum_i \frac{e^{w_i}}{\sum_j e^{w_j}} \cdot I_i$, this formula normalizes the weight of the previous formula, but this method has the disadvantage of increasing the amount of computation.To minimize the extra latency cost, we further propose a fast fusion approach.

Fast normalized fusion: $O = \sum_i \frac{w_i}{\varepsilon + \sum_j w_j} \cdot I_i$, where $w_i \geq 0$ is ensured by applying a Relu after each w_i, and $\varepsilon = 0.0001$ is a small value to avoid numerical instability. Ablation study shows that the results of this weighting method are similar to those of softmax feature fusion, the computing speed on GPU is improved by 30%.

Authorized licensed use limited to: Oestfold University College. Downloaded on February 20,2023 at 18:15:34 UTC from IEEE Xplore. Restrictions apply.

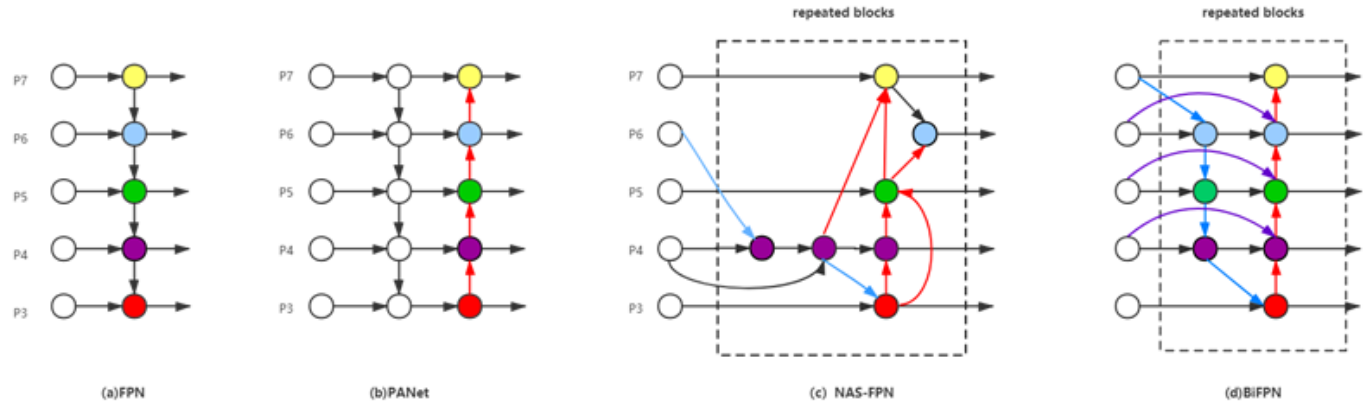

Fig. 2. Feature network design

D. Compound Scaling

Aiming at optimizing both accuracy and efficiency, we would like to develop a family of models that can meet a wide spectrum of resource constraints. A key challenge here is how to scale up a baseline EfficientDet model.

Backbone network – We use EfficientNet-b0 to EfficientNet B6, and the width and depth are the same as the seven networks.

BiFPN network – The width of BiFPN network increases exponentially while the depth increases linearly, which satisfy the following formula:

$$W_{bifpn} = 64 \cdot (1.35^{\phi}) \ , \ D_{\text{bifpn}} = 3 + \phi \qquad (1)$$

Box/class prediction network –The width of the network, is the same as the output of the previous part of BiFPN. But the depth satisfies the following formula:

$$D_{\text{box}} = D_{class} = 3 + \lfloor \phi / 3 \rfloor \qquad (2)$$

Input image resolution – Because the input of BiFPN is the P3 to P7 layer of the backbone network adopted,the resolution of the input image should be divisible by 2^7 . The resolution of the image should meet the following formula:

$$R_{\text{input}} = 512 + \phi \cdot 128 \qquad (3)$$

III. Image Acquisition and Experimental Results

A. Image acquisition and processings

Wheat ear images were taken in the provincial key Laboratory of Plant Genetics and Breeding of Sichuan Agricultural University in mid-to-late May 2018 and mid-to-late May 2019, and the grown trend of wheat is good. The camera adopts Shaanxi MV-E industrial camera, which possesses 10 megapixels and features clear image quality, good color reduction and stable operation. It supports IO signal to input and output. Wheat is mainly winter wheat, which is shot vertically under sunny and backlight conditions with a height of about 1.3m from the ground. A total of 608 images were taken, and labelImg was used to label the images, of which 540 were used as training sets and 68 as test sets. Prior to training, we're doing data augmentation for the training set picture which includes: SMOTE, SamplePairing and mixup. In the case of not substantially increasing the amount of data, the limited data can generate value equivalent to more data, thus improving the training effect.

B. Experimental results and analysis

In this paper, EffificientDet-D0 and EffificientDet-D3 were trained respectively. We use Pytorch frame of 1.4.0 version, graghics card drive of CUDA-10.0 version and Python3.7. computerese, and the epoch was set as 200, the experimental system in this paper is Linux Ubuntu 16.04, The graphics card is GeForce RTX 2060Ti.Finally, the accuracy of wheat ear identification in the training set and the test set was shown in Table I:

TABLE I. Accuracy and Speed

EffificientDet-	D0	D1	D2	D3
Training set accuracy	87.66%	90.67%	93.49%	95.36%
Test set accuracy	82.48%	89.74%	91.08%	92.92%
Time/s	0.09	0.11	0.13	0.24
FPS	11.8	9.57	7.6	4.2

With the same epoch, EfficientDet-D0 to EfficientDet-D3 the detection accuracy gradually increases, the final Efficientdet-D3 had a training accuracy of 95.36% and the test accuracy of 92.92%. We also tested the four models, the results of detection of the same wheat ear image with high density using EffificientDet-D0 to EffificientDet-D3 models that have been trained are shown in Figure 3 to Figure 6:

Figure 3(1) and Figure 3(2) are the detection results of EfficientDet-D0 and EfficientDet-D1 respectively. In these figures, most ears of wheat are not recognized, and some frames contain multiple ears of wheat. Figure 3(3) shows the detection result of Efficientdet-D2. Only a small number of ears of wheat have not been identified, and they are all wheat ears with small individual size. Figure 3(4) is the detection result of EfficientDet-D3. Almost all ears of wheat can be identified, and there is no case of containing multiple ears of wheat. Through experiment and comparison, we can see that EffificientDet-D3 can achieve very good results in wheat ear detection.

Efficientdet-d3 model was used to detect the experimental samples and compare the number of ears detected from the image with the number of ears counted by hand, then the error and accuracy of wheat ears counting were calculated. as shown in Table II. The average measurement accuracy of the final statistics was 95.30%.

Authorized licensed use limited to: Oestfold University College. Downloaded on February 20,2023 at 18:15:34 UTC from IEEE Xplore. Restrictions apply.

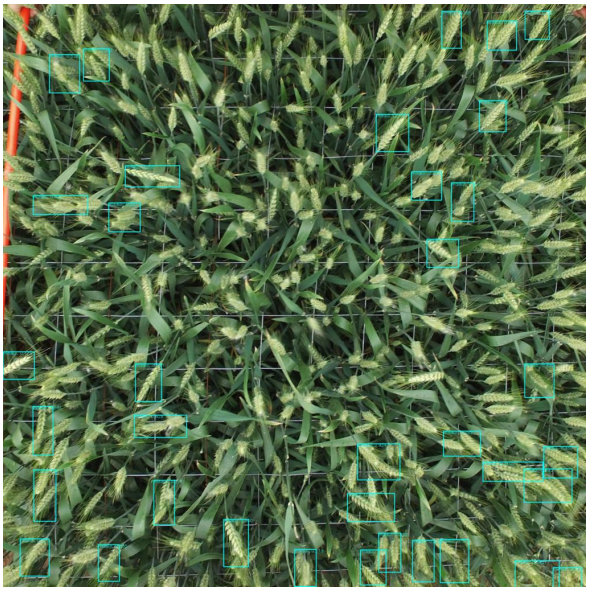

Fig. 3. Output of EfficientDet-D0

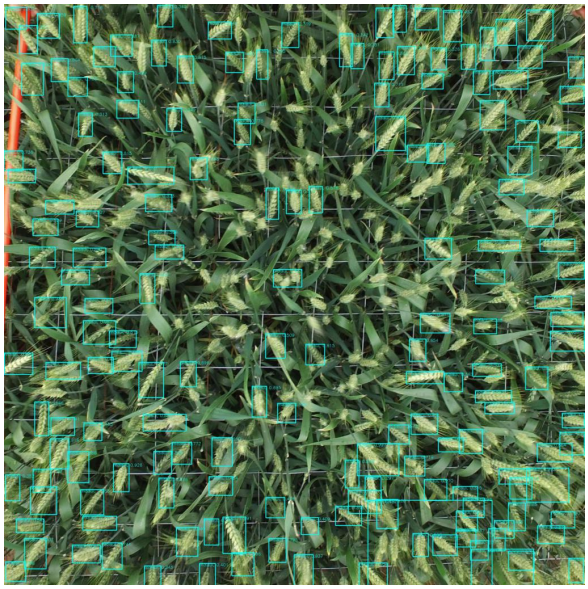

Fig. 5. Output of EfficientDet-D2

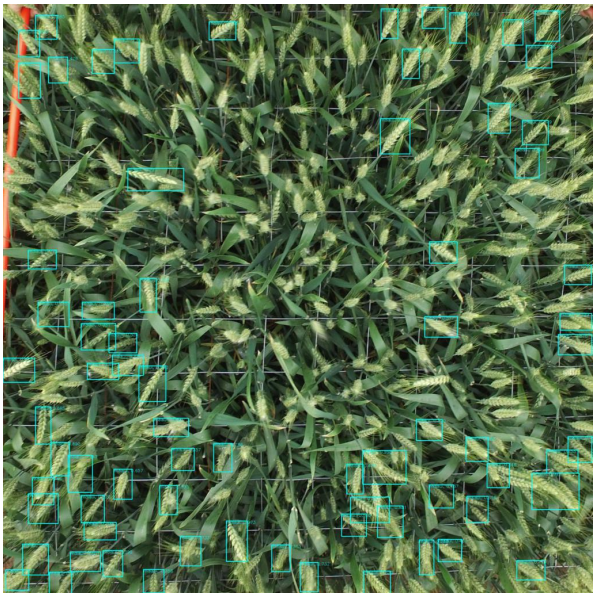

Fig. 4. Output of EfficientDet-D1

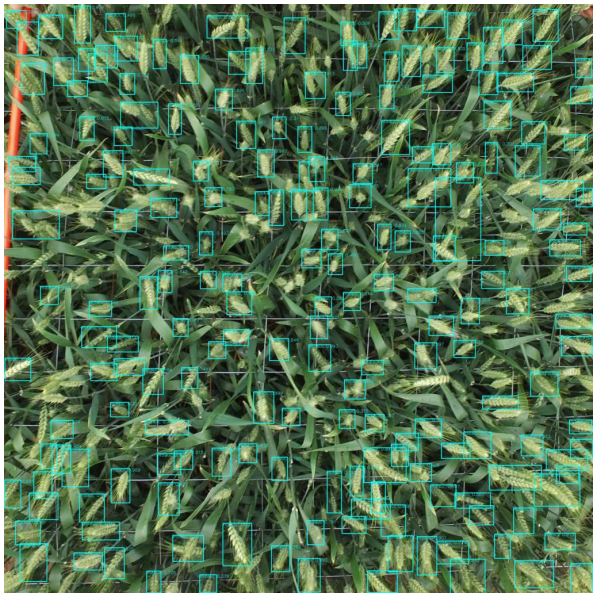

Fig. 6. Output of EfficientDet-D3

TABLE II. ALGORITHM COUNTING ACCURACY

No.	Manual statistics	Algorithm statistics	Statistical Error/ %	Statistical Accuracy/ %
1	51	49	3.92	96.08
2	87	89	-2.30	97.70
3	89	95	-6.74	93.26
4	90	98	-8.89	91.11
5	97	96	1.03	98.97
6	101	103	-1.98	98.02
7	104	99	4.81	95.19
8	117	125	-6.84	93.16
9	123	120	2.44	97.56
10	133	146	-9.77	90.23
11	143	139	2.80	97.20
12	144	129	10.42	89.58
13	151	158	-4.64	95.36
14	167	164	1.80	98.20
15	177	191	-7.91	92.09
16	190	187	1.58	98.42
17	197	190	3.56	96.44
18	211	207	1.90	98.10
19	220	233	-5.91	94.09
20	234	223	4.70	95.30

Authorized licensed use limited to: Oestfold University College. Downloaded on February 20,2023 at 18:15:34 UTC from IEEE Xplore. Restrictions apply.

IV. Conclusion

As for the problems of operation trouble, low efficiency and low accuracy in counting wheat ear image by traditional methods, we propose to use the deep learning model EfficientDet-D3 to detect the target of ear, and the quantity of target is counted at the same time so as to achieve the goal of automatic counting of ear. From the experimental results, the EfficientDet algorithm can quickly carry out target detection of wheat ear image and the accuracy rate reaches 92.92% and the counting accuracy reaches 95.30%. With high accuracy, it is also able to adapt to a variety of relatively complex lighting, different shooting angles and different ear densities, with strong adaptability and extension. We can only train EfficientDet-D3 so far, because of the limiting experimental environment, EfficientDet-D4 to EfficientDet-D7 are unable to train anymore. Theoretically, EfficientDet-D7 can achieve the highest detection accuracy, in the future work, we will continue to improve our method to achieve a better effect.

References

[1] Gao Yunpeng. Study on detection method of wheat ear in field based on deep neural network [D].Beijing Forestry University.2019.,

[2] Zhang Lingxian, Chen Yunqiang, Li Yunxia, Ma Juncheng, Du Keming. Detection and Counting System for Winter Wheat Ears Based on Convolutional Neural Network [J].Transactions of The Chinese Society of Agricultural Machinery, 2019, 50(03):144-150.

[3] Liu Tao, Sun Chengming, Wang Lijian, Zhong Xiaochun, Zhu Xinkai, Guo Wenshan. In-field Wheatear Counting Based on Image Processing Technology [J].Transactions of The Chinese Society of Agricultural Machinery, 2014, 45(02):282-290.

[4] Fan Mengyang, Ma Qin, Liu Junming, Wang Qing, Wang Yue, Duan Xiongchun. Counting Method of Wheatear in Field Based on Machine Vision Technology[J].Transactions of The Chinese Society of Agricultural Machinery,2015,46(S1):234-239.

[5] Zhao Feng. Reserch on wheat spike identification based on color features and improved AdaBoost algorithm [D]. Agricultural University of Hebei, 2014.

[6] Liu Zhe, Huang Wenzhun, Wang Liping. Field wheat ear counting automatically based on improved K-means clustering algorithm, 2019, 35(03):174-181.

[7] Li Yinian, Du Shiwei, Yao Min, Yi Yingwu, Yang Jianfeng, Ding Qishuo, He Ruiyin. Method for wheatear counting and yield predicting based on image of wheatear population in field[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE),2018,34(21):185-194.

[8] Fathollah Bistouni, Mohsen Jahanshahi (2020). Impact of Raising Switching Stages on the Reliability of Interconnection Networks. Journal of the Institute of Electronics and Computer, 2, 93-120.

Authorized licensed use limited to: Oestfold University College. Downloaded on February 20,2023 at 18:15:34 UTC from IEEE Xplore. Restrictions apply.