Prediction and counting of field wheat based on LC-DcVgg

Zhang Yiwen[1]
#School of Information Engineering, Sichuan Agricultural University
*Key Laboratory of Agricultural Information Engineering of Sichuan Province
86-18383571801
jiedepinghengqi@126.com

Shu Baiyi[2]
#School of Information Engineering, Sichuan Agricultural University
*Key Laboratory of Agricultural Information Engineering of Sichuan Province
86- 18227592917
Shubaiyi100@stu.sicau.edu.cn

Xue Ziwei[3]
#School of Information Engineering, Sichuan Agricultural University
*Key Laboratory of Agricultural Information Engineering of Sichuan Province
86-13600636456
1021543870@qq.com

Wang Yue[4]
#School of Information Engineering, Sichuan Agricultural University
*Key Laboratory of Agricultural Information Engineering of Sichuan Province
86- 18681734858
1012252306@qq.com

Mu Jiong[5]*
#School of Information Engineering, Sichuan Agricultural University
*Key Laboratory of Agricultural Information Engineering of Sichuan Province
86-13340608699
jmu@sicau.edu

ABSTRACT
The number of wheat spikes per unit area is an important parameter for assessing wheat yield and wheat planting density. At present, the methods of intelligent counting of wheat include remote sensing technology and machine learning technology, but all have shortcomings such as poor stability, strong limitations, and poor versatility. And however, the existing object detection neural network algorithm requires a large amount of manpower to produce data sets. And it is not possible to identify too dense wheat spikes. In this paper, a new structure called direct connection is proposed to improve the algorithm of Vggnet, which makes it better combined with localization-based counting loss. Direct connection can fuse the features of shallow layer with those of deep layers, which can make the network retain the original picture information and make the localization-based counting loss play a better role in wheat spike counting. The model has a good recognition effect on the wheat spikes that are stuck together. For dense wheat spikes, the model can achieve MAE of 11.857, an accuracy rate of 90.4%, RMSE of 16.985.

CCS Concepts
• **Computing methodologies~Artificial intelligence~Computer vision ~Computer vision problems~Image segmentation**

Keywords
Deep learning;Point supervision;Spikes count;Yield prediction

ICMAI 2020, April 10–13, 2020, Chengdu, China

ACM ISBN 978-1-4503-7707-2/20/04…$15.00
https://doi.org/10.1145/3395260.3395299

1. INTRODUCTION
Wheat is the most important grain crop in world trade and one of the main grain crops in China. It is an important task in agricultural production to estimate wheat yield by the number of wheat spikes per unit area.

Under the requirement of fast and accurate wheat yield counting, it is very necessary and reasonable to develop an automatic counting network model to realize agricultural automation.

This paper proposes a point label based CNN neural network wheat spikes counting method and obtains a suitable model for wheat spikes number. First, it breaks the conventional traditional machine learning method and uses the deep learning algorithm to identify wheat spikes pictures. Second, the new network structure is improved. Third，The model has a good recognition effect on the wheat spikes that are stuck together.

The noise in the field environment can be overcome effectively, the effective characteristics of wheat spikes can be extracted and the number of wheat spikes can be counted automatically and efficiently. For dense wheat spikes, the mae of 11.857 can be obtained by this model.

2. RELATED WORK
There are three main ways to predict wheat growth.

The first is the traditional artificial area survey method, which has disadvantages such as slow speed, high cost and low accuracy, and is not suitable for large-scale wheat growth assessment[1][2].

The second is to predict wheat growth through remote sensing technology [3][4][5]. In China, Li weiguo et al. [6] built a remote sensing estimation model of winter wheat yield with the research method of combining the ground GPS positioning survey and p-6 satellite remote sensing data. The model has good estimation performance. Abroad, JA Fernandez Gallego et al[7] used thermal images to automatically count wheat spikes and proved the high correlation between thermal image counting and manual counting. This method provides some technical support for wheat yield

estimation, but it is not suitable for small-scale wheat growth estimation due to its weak stability and lack of versatility.

The third is the traditional machine learning method [8]. Fan mengyang et al. [9] adopted the learning method of support vector machine to accurately extract wheat spikes contour to count wheat spikes. Liu zhe et al. [10] proposed the wheat spikes count method based on improved k-means. The accuracy of wheat ear count reached 94.69%. These methods all cost a lot of time and manpower in the image preprocessing stage, and are not robust enough to noise such as uneven illumination and complex background in field environment, so it is difficult to expand the application.

Convolutional neural network (CNN) is a kind of deep learning model that is applicable to image analysis and emerged in recent years [11][12]. Xiong etc. [13] proposes a rice segmentation is based on pixel segmentation and CNN - Panicle - SEG, achieved the identification of rice grain in the field environment. Chengquan Zhou[14] et al,used the model based on multi-feature optimization and twsvm to test the ear of wheat in the field, the final precision of the model is 0.79 – 0.82. However, these algorithms all need to build a large number of sample bases for training, which takes a long time.

Therefore, the existing counting methods are still not perfect, and it is urgent to develop more efficient and accurate technology to realize wheat yield prediction.

3. MATERIALS AND METHODS

3.1 Dataset

The wheat datasets used in this paper are all obtained by ourselves The specific image processing process is shown in Fig. 1. The original image size is 4000x2250, and the standard image after cutting is 480x640 Then we use the point marking method to mark the data set, which marks the center position of the wheat ear and sets all other positions to zero (we mark all the wheat spikes who is visible in the image, including the wheat spikes that are blocked and in the shadow). Point level annotation is easy to annotate because it requires less manpower than bounding box annotation and per pixel annotation Point level annotations provide a rough estimate of the location of an object but do not provide its size or shape.

There are 1254 Wheat Images in the dataset, which are divided into training set, verification set and test set The training set has 659 wheatear images, the verification set has 210 wheatear images, and the test set has 190 wheatear images Training set, verification set, and test set have the same resolution and image size.

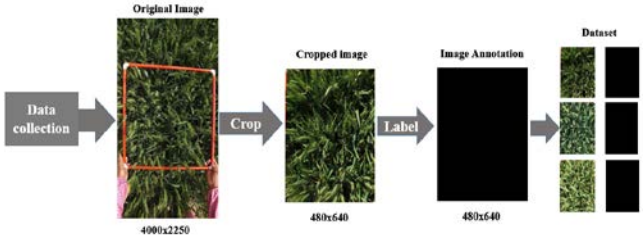

Figure 1 Data processing flow

3.1.1 Data Augmentation

Due to the uncertainty of light and weather in the wheat field, the field illumination conditions are very complicated. To enrich the image training set and avoid overfitting, we have enhanced the data set. Image enhancement eliminates lighting and clearly identifies the foreground of the image. In this study, in order to facilitate the generalization of the training model in the future, we used the following methods to process the original image: (a) brightness enhancement and attenuation, (b) chroma enhancement and attenuation, and (c) contrast enhancement and attenuation.At the same time, this paper uses the left and right flip, 180° rotation, etc. to expand the data set.

3.2 Localization-based counting loss (LC)[15]

Localization-based counting loss (LC)[15] is a function that encourages the network to output a single blob per object instance using point-level annotations only.Detection model needs to learn the location, size,and shape of object instances that are possibly heavily occluded.detection but the idea of LC is that there is no need to get precise boundaries for counting,it is only necessary to make sure we have a positive region on each object and a negative region between object.Therefore, it is very suitable for situations with severe occlusion, such as wheat spikes in the field.localization-based counting loss consists of four distinct terms:

$$L(S,T) = \underbrace{L_I(S,T)}_{Image-level-loss} + \underbrace{L_P(S,T)}_{Point-level-loss} + \underbrace{L_S(S,T)}_{Split-level-loss} + \underbrace{L_F(S,T)}_{False-positive-loss} \quad (1)$$

The first two terms, the image-level and the point-level loss, enforces the model to predict the semantic segmentation labels for each pixel in the image,the split-level loss and the false-positive loss encourage the model to output a unique blob for each object instance and remove blobs that have no object instances.The last two terms,the split-level loss and the false-positive loss encourage the model to output a unique blob for each object instance and remove blobs that have no object instances.

3.2.1 Image-level loss

$$L_I(S,T) = -\frac{1}{|C_e|}\sum_{c\in C_e}\log(S_{t_c C}) - \frac{1}{|C_{-e}|}\sum_{c\in C_{-e}}\log(1-S_{t_c c}) \quad (2)$$

C_e the set of classes present in the image

For each category that appears in the image, at least one pixel should be labeled as that classes. No pixel should belong to a classes that does not exist in the image

3.2.2 Point-level loss

$$L_p(S,T) = -\sum_{i\in\Gamma_s}\log(S_{iT_i}) \quad (3)$$

Γ_s represents the lo-cations of the object instances

T_i represents the true label of pixeli.

This loss ignores all the pixels that are not annotated.

3.2.3 Split-level loss

$$L_S(S,T) = -\sum_{i\in T_b}\alpha_i\log(S_{i0}) \quad (4)$$

S_{i0} is the probability that pixel i belongs to the background class and α is the number of point-annotations in the blob in which pixel i lies.

There are two ways to split, one is line split, The other is watershed split[18].In the paper, we use the watershed split.This loss function encourages the model to focus on segmenting the spots with the most point-level annotations. Just like training the model with the help of the watershed segmentation algorithm

3.2.4 *False Positive loss*

$$L_F(S,T) = -\sum_{i \in B_{fp}} \log(S_{i0}) \quad (5)$$

B_{fp} is the set of pixels constituting the blobs predicted for each class,except the background class

S_{i0} is the probability that pixel i belongs to the background class

This loss discourages the model to predict a blob with no point annotations, to reduce the number of false-positive predictions.

3.3 Network structure

Our model framework is based on the classic vggnet16, which is the classic base model for deep learning. The traditional vgg network consists of 5 maximum pooling layers, 13 convolution layers, 3 full connection layers, and a SoftMax classifier layer. The convolution kernel is 3 × 3, which can better extract image details and enhance the nonlinear expression of the network (step = 1). The maximum pooling technique is used in all pooling layers, and the size of the pooling window is 2 x 2 (step = 2). All hidden layers are added along with the ReLU layer.

In this paper, we replaced the vggnet's fully connected layer and the SoftMax classifier layer with a layer of upsampling, so that the model can return a complete image. It is found through experiments that the original vgg network can not achieve good results. Because the LC uses the traditional segmentation algorithm—the watershed algorithm. Therefore,the characteristics of the network transmission to loss need to original image features.The basic features,can not be too abstract, otherwise the loss can not work, and inspired by the shot cut in resnet[16], we have improved the original vgg network structure, so that it can be well-matched with LC, get better results. The specific network structure is shown in Figure 2.

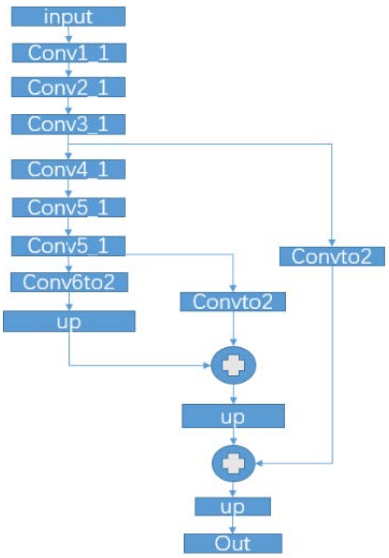

Figure 2 Direct connection

Based on vggnet16, we get the earlier features of the model and the later-derived features to the same dimension through the convolution block and the upsampling layer and then connect them. We call this structure direct connection.The network can enable the basic image information required for the Split-level loss while maintaining the effective extraction of the features of the image so that the Split-level loss can be put to good use. According to this principle, densenet[17] should also have a good effect, but according to our experiments, the actual effect of densent is still not as good as our improved network. See Section 4 for a detailed experimental comparison. See Figure 3 for the overall model structure.

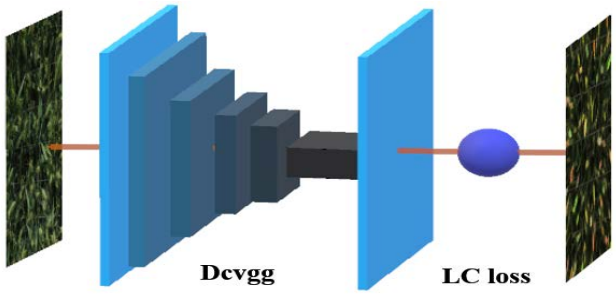

Figure 3 Model structure

4. RESULTS AND DISCUSSION

In this section, we describe the evaluation indicators, the training process, and introduce the experimental results and discussion.

4.1 Metric

MAE is a commonly used metric for evaluating object counting methods.Since we are using point labels for labeling, we can easily get the number of wheat spikes in each image, and our model is not like the traditional target detection model, our model is easy to get results larger than the actual num ber of objects., so we chose mae as a metric to assess the accuracy of the model.

4.2 Training setting

The experiment was run on two NVIDIA GeForce RTX 2080Ti graphics cards. Based on the pytorch deep learning framework, we use Python to train and test the target recognition network model. In this paper, we use the stochastic gradient descent method to train the network in an end-to-end joint manner. All data is obtained after the model converges. We use the Adam optimizer with a learning rate of 10^{-5} and a weight decay of 5×10^{-5}.

4.3 Results

Effect pictures after experiment recognition are as follow:

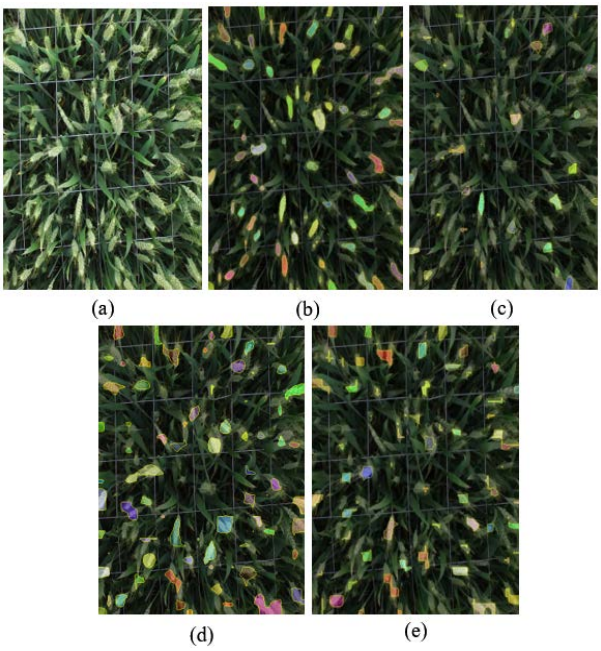

Figure 4 (a) is the original picture,(b) Is the result of the DcVgg model,(c) Is the result of the Vgg model ,(d) Is the result of the Vgg-sc model,(e) Is the result of the densenet model

Table 1 Comparison of indicators between models

	Vgg	Vgg-sc	Dcvgg	Densenet-40	Actual value
Pre-value	36	76	**85**	73	126
Best MAE	13.671	**10.442**	11.857	13.968	----
Best Acc-rate	0.879	**0.908**	0.904	0.876	----
Best RMSE	20.731	17.229	**16.985**	20.363	----
Best NRMSE	0.319	0.229	**0.207**	0.2828	----

More test data information is as follows:

Image no	Predictive value	Actual value
TEST_0005.jpg	107	116
TEST_0009.jpg	76	113
TEST_0015.jpg	99	145
TEST_0022.jpg	119	94
TEST_0027.jpg	109	127
TEST_0033.jpg	95	88
TEST_0040.jpg	119	105
TEST_0045.jpg	117	118
TEST_0050.jpg	105	86

From figure 4 and table 1, we can conclude that for LC-Dcvggnet, Vgg-sc(Vggnet include shortcut), and Densenet, our model can get good results. Compared with the traditional target detection model, our model can better solve the problem of overlapping targets that cannot be solved by common target detection, and also has better recognition for the shaded and incompletely displayed wheat spikes. For semantic segmentation, the size of the wheatear is too small, the boundary is not smooth, and the general segmentation model is difficult to complete the segmentation. LC-Dcvgg can learn the edge features of wheat well and segment the shape of wheat ears well.

4.4 Analysis

Compared to other models, our Dcvgg can get the most accurate results(see the figure 4). And the shape obtained by model segmentation is also most similar to wheat. Although the loss of vgg-sc is the smallest, its actual effect is far less than that of Dcvgg, the segmentation shape does not match the edge of the wheatear, and the sticky wheat ear can't be divided very well. The reason should be that although the shortcut structure can play the role of skipping the useless module and maintaining the gradient correlation, it does not retain the basic information of the image well, which causes split-level loss to not play its role well. For the densenet model, although the dense connection structure can make good use of image features, by observing (e) in Figure 4, we can clearly find that the segmented image presents many obvious irregular squares. The main reason should be the use of the bottleneck layer results in too little parameter, the model can not learn the edge information of the wheat well, and can only divide the image into irregular squares according to the position of the point label.

4.5 Existing Problems

Figure 5 Error details

As can be seen from the results, for the occluded and shaded wheat spikes, although our model has a good identifiable effect, there are still many errors, (see figure 5), such as the presence of multiple sticky spikes of wheat. It is recognized as one case, or the blade is mistakenly identified as wheat spikes, and the range of the mark does not cover the entire wheat ear. And in some cases, a single wheat ear is identified as multiple spikes of wheat. This will be an important direction for our subsequent improvement of the model.

5. CONCLUSION

Accurately estimating wheat yield requires accurate statistics of the number of wheat spikes per unit area. However, the success rate of wheat ear identification is seriously hindered by the complex field environment, the occlusion between wheat spikes and the change and constant change of natural light. In this paper, field wheat spikes are used as research objects, and point labels are used for object labeling. Through the direct connection method, the LC-vgg network structure was improved, which can effectively solve the problem of difficulty in identifying sticky wheat spikes and the difficulty in identifying incomplete wheat spikes in the shadows. The rapid identification and accurate counting of wheat spikes were achieved, with a final mae of 11.85 and an NRMSE of 0.207.The method can effectively overcome the noise in the field environment, extract the effective features of the wheat spikes, directly predict the wheat yield information, and have strong adaptability and expandability in realizing the automatic counting of wheat, realizing the unit area of the wheat

field. Through the automatic measurement of wheat, it provides an accurate reference model for the prediction of wheat yield. For future work, we plan to continue improving the network structure and exploring different FCN architectures to improve the effectiveness of the model and compress the model so that it can be better applied to actual agricultural production.

6. ACKNOWLEDGMENTS

This paper was supported by Key Laboratory of Agricultural Information Engineering of Sichuan Province, Sichuan Agricultural University, 2019 Google-supported industry-university collaboration collaborative education project student project (PJ190499#) "Wheat ears recognition and counting based on deep learning under Tensorflow platform" and Wheat growth assessment in farmland based on neural network algorithm,College Students' innovation and Entrepreneurship Training Program(201910626026). Furthermore, thanks to CERNET Innovation Project (No.NGII20170611) and Science and Technology Innovation Miaozi Project (No.2019025)

7. REFERENCES

[1] Zhu, Y., Cao, Z., Lu, H., Li, Y., and Xiao, Y. (2016). In-field automatic observation of wheat heading stage using computer vision. Biosystem Engineering, 143:28–41.

Agricultural Machinery, 2015, 46(12): 234−239. (in Chinese with English abstract)

[2] bayberry, Li Guang. Simulation of wheat yield prediction model. Journal of wheat crops [J], 2013, 30 (10): 382-385.

[3] Feng Qi, Wu Shengjun. Research progress of crop remote sensing yield estimation in China [J]. World science and technology research and development, 2006, 28 (3): 32-36

[4] Fan Mengyang, Ma Qin, Liu Junming, et al. Counting method of wheat ear in field based on machine vision technology[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(12): 234−239. (in Chinese with English abstract)

[5] Liu Huimin, he binfang, Zhang hongqun. Study on remote sensing monitoring and evaluation method of winter wheat growth in Anhui Province [J]. China agronomy bulletin, 2011, 27 (33): 18-22.

[6] Li Weiguo, Wang Jihua, Zhao Chunjiang, et al. Remote sensing estimation of winter wheat yield based on ecological factors [J]. Journal of wheat crops, 2009, 29 (5): 906-909.

[7] Fernandez-Gallego J A., Buchaillot M L, Gutiérrez N A ,et al.Automatic Wheat Ear Counting Using Thermal Imagery.Remote Sens. 2019, 11(7), 751.

[8] GAO Zhenyu, WANG An, LIU Yong, et al. Intelligent fresh-tea-leaves sorting system research based on convolution neural network[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(7): 53-58.（in Chinese）.

[9] Fan Mengyang, Ma Qin, Liu Junming, et al. Counting method of wheat ear in field based on machine vision technology[J]. Transactions of the Chinese Society for

[10] Liu Zhe, Huang wenzhun, Wang Liping. Automatic counting of wheat ear in field based on improved k-means clustering

[11] .Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436－444

[12] Kuang Ling. Prediction of grain yield by RBF neural network [J]. Computer simulation, 2011, 28 (11): 189-200.

[13] Xiong X, Duan L, Liu L, et al. Panicle-SEG: A robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization[J]. Plant Methods, 2017, 13(1): 104－113.

[14] Zhou C, Liang D, Yang X, et al. Wheat ears counting in field conditions based on multi-feature optimization and TWSVM[J]. Frontiers in plant science, 2018, 9: 1024.

[15] Laradji I H, Rostamzadeh N, Pinheiro P O, et al. Where are the blobs: Counting by localization with point supervision[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 547-562.

[16] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[17] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

[18] algorithm [J]. Journal of agricultural engineering, 2019,35 (03): 174-181