

Received November 25, 2019, accepted December 3, 2019, date of publication December 10, 2019, date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2958831

# Combined Use of FCN and Harris Corner Detection for Counting Wheat Ears in Field Conditions

DAOYONG WANG<sup>1,2</sup>, YUANYUAN FU<sup>1</sup>, GUIJUN YANG<sup>1</sup>, XIAODONG YANG<sup>1</sup>,  
DONG LIANG<sup>2</sup>, CHENGQUAN ZHOU<sup>3</sup>, NING ZHANG<sup>1</sup>, HONGYA WU<sup>4</sup>,  
AND DONGYAN ZHANG<sup>2</sup>

<sup>1</sup>Key Laboratory of Quantitative Remote Sensing in Agriculture of Ministry of Agriculture, Beijing Research Center for Information Technology in Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

<sup>2</sup>National Engineering Research Center for Agro-Ecological Big Data Analysis and Application, Anhui University, Hefei 230601, China

<sup>3</sup>Institute of Agricultural Equipment, Zhejiang Academy of Agricultural Sciences (ZAAS), Hangzhou 310021, China

<sup>4</sup>Institute of Agricultural Sciences of Lixiahe District, Yangzhou 225100, China

Corresponding authors: Xiaodong Yang (yangxd7@hotmail.com) and Dongyan Zhang (zhangdy@ahu.edu.cn)

This work was supported in part by the Natural Science Foundation of China under Grant 41771469, Grant 41801225, and Grant 41771463, and in part by the Project of the Beijing Postdoctoral Research Foundation under Grant 2018-ZZ-066.

**ABSTRACT** Accurate counting of wheat ears in field conditions is vital to predict yield and for crop breeding. To quickly and accurately obtain the number of wheat ears in a field, we propose herein a method to count wheat ears based on fully convolutional network (FCN) and Harris corner detection. The technical procedure consists essentially of 1) constructing a dataset of wheat-ear images from acquired red-green-blue (RGB) images; 2) training a FCN as the wheat-ear segmentation model by using the constructed image dataset; 3) preparing testing images and inputting them into the segmentation model to get the initial segmentation results; 4) binarizing the initial segmentation by using the Otsu algorithm (to facilitate subsequent processing); and 5) applying Harris corner detection after extracting the wheat-ear skeleton to obtain the number of wheat ears in the images. The segmentation results show that the proposed FCN-based segmentation model segments wheat ears with an average accuracy of 0.984 and at low computational cost. An average of only 0.033 s is required to segment a  $256 \times 256$ -pixel wheat-ear image. Moreover, the segmentation result is improved by nearly 10% compared with the previous segmentation methods under conditions of wheat-ear occlusion, leaf occlusion, uneven illumination, and soil disturbance. Subsequently, the proposed counting method achieves good results, with an average accuracy of 0.974, a coefficient of determination ( $R^2$ ) of 0.983, and a root mean square error (RMSE) of 14.043. These metrics are all improved by 10% compared with the previous methods. These results show that the proposed method accurately counts wheat ears even under conditions of wheat-ear adhesion. Furthermore, the results provide an important technique for studying wheat phenotyping.

**INDEX TERMS** Wheat-ear counting, fully convolutional network, wheat-ear adhesion, Harris corner detection, field conditions.

## I. INTRODUCTION

Wheat is an important primary grain [1], and its yield is crucial for national food security [2]. Wheat is in high demand by most of the world's population [3], so research into predicting wheat yield has attracted significant attention [4]. The number of wheat ears per unit area is intimately linked to wheat yield and breeding [5], and two

methods are available to obtain the number of wheat ears per unit area: manual field counting and image-based counting of wheat ears. Because conventional manual counting is time consuming and subjective, image-processing technology has recently gained widely acceptance for counting wheat ears [6]–[8] because of its strong universality and high efficiency [9]. Generally, image-based counting of wheat ears is confronted with two main problems: the accurate segmentation of wheat ears and wheat-ear adhesion [10].

The associate editor coordinating the review of this manuscript and approving it for publication was Jingchang Huang.

Currently, two main methods exist for object segmentation: color-information-based and classifier-based segmentation. Fan *et al.* enhanced the contrast between image colors by fusing color saliency maps with brightness saliency maps and then used the color information to segment a simple dermatology image [11]. Ganesan *et al.* used the CIELAB color space to quantify visual differences in an image, which enabled them to segment images with significant color differences [12]. Color-information-based segmentation methods rely excessively on color information and cannot effectively segment images with small color differences [13]. Classifier-based segmentation methods segment objects by learning the characteristics of a target. Li *et al.* combined the texture features of wheat ears with a neural network to effectively identify wheat ears in a laboratory environment with white-board background [14]. Zhou *et al.* combined multi-feature optimization with twin-support vector machine to identify wheat ears in field conditions; however this method was greatly affected by soil background [13]. Classifier-based segmentation methods need to extract target features and then input them into classifiers to achieve segmentation. The quality of the features chosen for segmentation plays an important role in segmentation. Normally, it is difficult to design a feature-extraction algorithm to fully mine the exclusive characteristics of targets in complex conditions. In a bid to overcome this difficulty, deep learning has recently become widely used in image segmentation because it can extract a complex hierarchy of features from images by self-learning [15]. For example, Bargoti *et al.* used convolutional neural networks to segment fruit in an orchard environment [16]. In deep learning, a semantic segmentation based on FCN has been widely used in image segmentation under complex conditions [17]. For example, Martin-Abadal *et al.* obtained high-precision semantic segmentation of the *Posidonia Oceanica* meadows in sea-floor images that was more reliable than manual marking of images [18]. In other work, Bai *et al.* proposed a deep learning method to semantically segment remote-sensing images on complex backgrounds by using U-Net to map the damage after a tsunami disaster. This method greatly improved upon the practice of using application values to respond to operational disaster [19]. Jiang *et al.* proposed an end-to-end personal segmentation network structure that fuses a person-detection network with a FCN, which allowed them to accurately segment a person in a natural scene [20]. As a result, because of its high segmentation accuracy and low computational cost, semantic segmentation has become widely used in complex conditions [21].

After segmenting wheat ears, the adhesion of wheat ears becomes particularly important. Fang *et al.* used watershed segmentation to deal with adhesion in plant disease recognition, but this method over-segments in multiple-adhesion situations [22]. Ning *et al.* effectively segmented images of adhesive ores based on pit matching. The adhesion situation in their study was relatively simple [23]. The methods mentioned above for dealing with adhesion introduce segmentation errors and cannot cope with the complicated adhesion

of wheat ears in field conditions. Corner detection can effectively identify point features of image-edge contour intersection, giving it the potential to solve multi-target adhesion [24]. Yan *et al.* used Shi-Tomasi corner detection to effectively solve the problem of adhesive vehicles [25]. Wang *et al.* used the curvature scale space corner detection method to successfully separate overlapped cells [26]. Therefore, corner detection might be one way to solve adhesion problems in the counting of wheat ears.

The main objectives of this study are (1) to explore the ability of FCN to segment wheat ears in field conditions; and (2) demonstrate whether corner detection can solve adhesion problems in wheat-ear counting. This paper first marks and cuts the acquired RGB images to train the wheat-ear segmentation model. The trained model is then used to segment the input image. Next, the wheat ears are binarized by using the Otsu algorithm to highlight the area of wheat ears. Finally, the wheat-ear skeletons are extracted, and the number of wheat ears is obtained by using Harris corner detection [27].

The rest of this paper is arranged as follows: Section II introduces the study area and the process used to generate wheat-ear image datasets for segmentation model training and testing. Section III outlines the proposed method to count wheat ears, and Sec. IV presents the experimental results for counting wheat ears in field conditions. Section V discusses the effectiveness of the method, and Sec. VI gives the main conclusions.

## II. STUDY AREA AND PREPARATION OF EXPERIMENTAL DATA

### A. STUDY AREA AND DATA ACQUISITION

The experiment field was located at the wheat breeding base of Agricultural Science of the Lixiahe District (Yangzhou, China, 32°25'18.97" latitude North, 119°31'25.66" longitude East). The experiment was conducted under clear windless weather from 12 a.m. to 2 p.m. on May 9, 2018, so the possibility of image distortion due to weather conditions was eliminated. A total of 180 wheat images were acquired by using a Panasonic DC-GF9 digital camera (resolution: 4592×3448 pixels, aperture: f/6.3, exposure time: 1/250 s) at 2.0 m above the vertical canopy, and with a 3 m × 2.25 m view (Figure 1). Concurrently, the number of wheat ears was counted manually in the same view field.

### B. WHEAT-EAR IMAGE DATASET FOR TRAINING AND TESTING SEGMENTATION MODEL

To train the proposed segmentation model, we selected 120 original wheat-field images to construct wheat-ear image datasets. The wheat ears in these images were annotated by four graduate students in 19 days. First, the wheat-ear contours were traced in red (R:225, G:0, B:0). Next, the contour of each wheat ear was filled by hole filling [28] to mark the entire wheat ear. The manually marked wheat ears in the original image were then used as the ground truth during training and testing of the segmentation model.

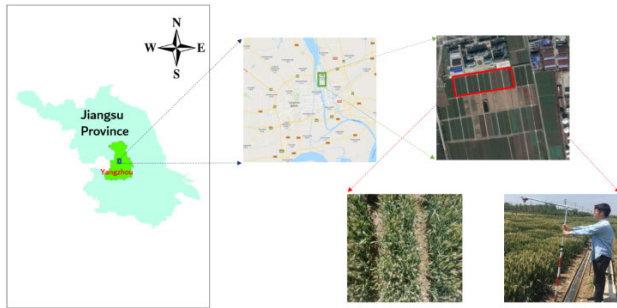


FIGURE 1. Schematic diagram of study area and data collection.

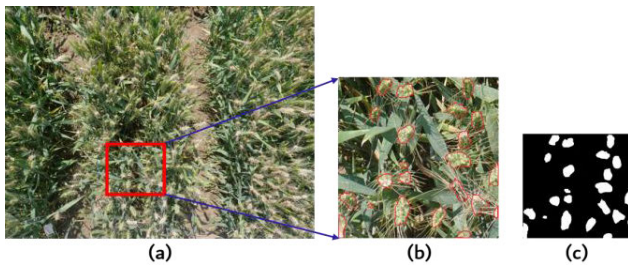


FIGURE 2. Generation of wheat-ear image dataset: (a) original image, (b) manually annotated sub-image, (c) resampled and gray-scaled sub-image. The red box represents a  $768 \times 768$ -pixel sliding window.

The original image was large and not conducive to training the segmentation model, so it was cut into sub-images. To ensure that wheat ears and background were as visible as possible, a  $768 \times 768$ -pixel sliding window was slid over the original image with a step of  $0.5 \times 768$  pixels to obtain a total of 6500 sub-images. The sub-images were resampled to  $256 \times 256$  pixels by using bilinear interpolation [29], and then the resampled sub-images were gray-scaled [30] for training segmentation models (Figure 2).

The input images were slightly processed to better test them. First, the input image was resampled to  $(6 \times 768) \times (5 \times 768)$  pixels by bilinear interpolation. Next, by sliding a  $768 \times 768$ -pixel window over an input image with a step of 768, a total of 30 sub-images were obtained. The sub-images were resampled to  $256 \times 256$  pixels by bilinear interpolation, and then the resampled sub-images were gray-scaled to serve as test images.

### III. METHODS

To properly segment wheat ears in complex conditions, we trained the segmentation model based on a FCN. Based on the segmentation results, the images are binarized by applying the Otsu algorithm to highlight the wheat-ear in the images, following which the wheat-ear skeletons are extracted. Finally, Harris corner detection was used to count the wheat ears. The critical steps of the method are as follows (Figure 3):

- 1) Train the wheat-ear segmentation model based on the constructed dataset of wheat-ear images.
- 2) Use the wheat-ear segmentation model to segment the test images. Next, stitch the segmented sub-images

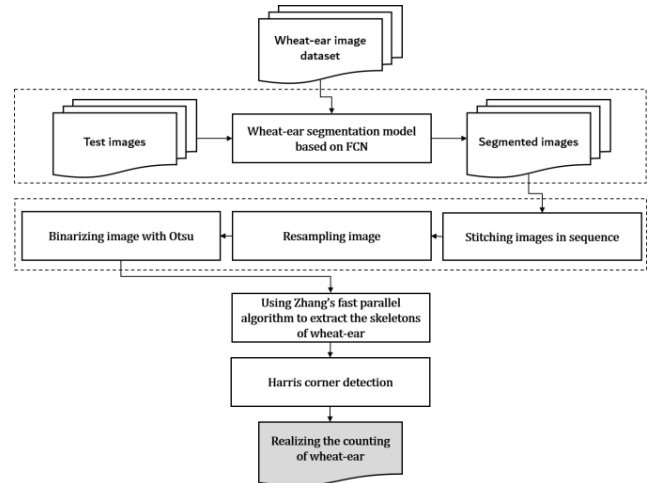


FIGURE 3. Flowchart of proposed wheat-ear counting method.

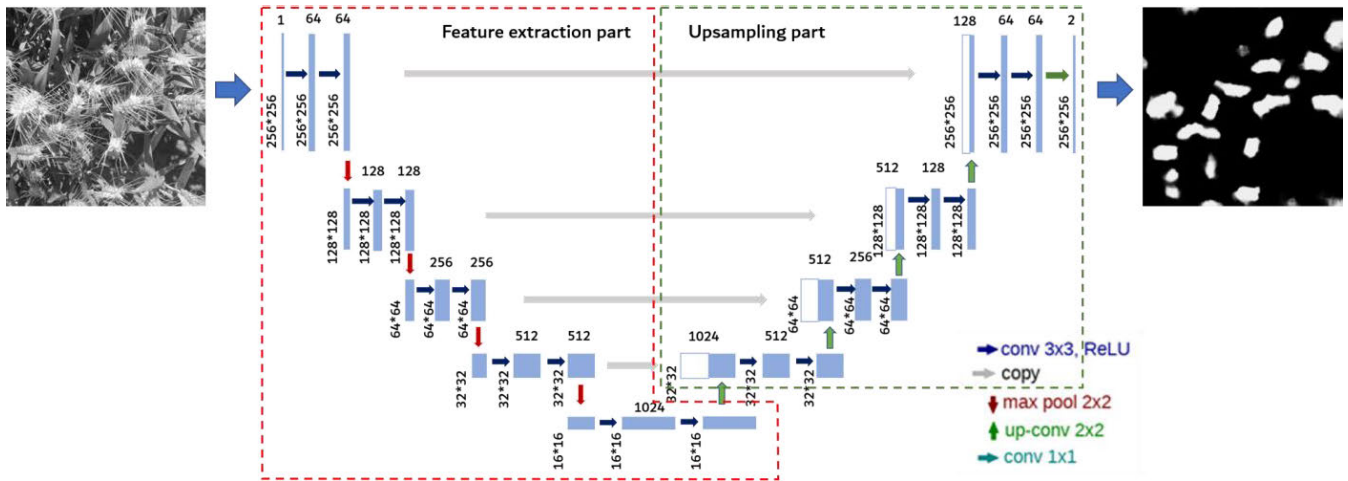
- 3) Use Zhang's fast parallel algorithm [31] to identify the wheat-ear skeletons in the binary image.
- 4) Use Harris corner detection to detect and then realize the counting of wheat-ear.

#### A. CONSTRUCTION OF WHEAT-EAR SEGMENTATION MODEL BASED ON FULLY CONVOLUTION NETWORK

The field wheat-ear segmentation model is realized by using a fine-tuned U-Net network, which semantically segments wheat ears in the field. The U-Net network is built on the architecture of a FCN [32]. In the feature-extraction module, the deep features of the target are learned by the successive convolution layer. The output of the feature-extraction module is then merged in the up-sampling module. The feature-extraction part consists of ten  $3 \times 3$  convolution layers and four  $2 \times 2$  maximum pooling layers, and the up-sampling part contains eight  $3 \times 3$  convolution layers, a  $1 \times 1$  convolution layer, and four  $2 \times 2$  deconvolution layers.

The network function is powerful and suited for multi-scale segmentation of large images in complex conditions. In this paper, the input to the network is a  $256 \times 256$ -pixel grayscale image. The output is a  $256 \times 256$ -pixel segmented grayscale image of wheat ears in the field. Using the rectified linear unit nonlinearity (ReLU) as activation function. The convolution result of each convolution layer is filled with zeros to ensure that the input and output size remains unchanged to avoid cropping when up-sampling the output part of the feature extraction. Figure 4 shows the field wheat-ear segmentation network.

The wheat-ear segmentation model is trained by the constructed wheat-ear image dataset, which contains 4550 training sets and 1950 verification sets. The specific parameters are (i) learning rate = 0.001, (ii) batch size = 20, (iii) epochs = 30, and (iv) steps\_per\_epoch = 1000.



**FIGURE 4.** Field wheat-ear segmentation network. The red dotted frame shows the feature-extraction part of the network, and the green dotted frame shows the up-sampling part of the network.

The learning rate determines how fast the parameter moves to the optimal value, the batch size indicates the number of samples taken from the training set for each training batch, the epoch indicates the number of rounds used for training, and the steps\_per\_epoch indicates the number of batches sent to be trained in an epoch. The training time is 10.06 hours, the segmentation accuracy is 0.984, the loss is 0.038, and the segmentation of a 256×256 image takes 0.033 s. The segmentation model can be well used for the segmentation of wheat ears in the field.

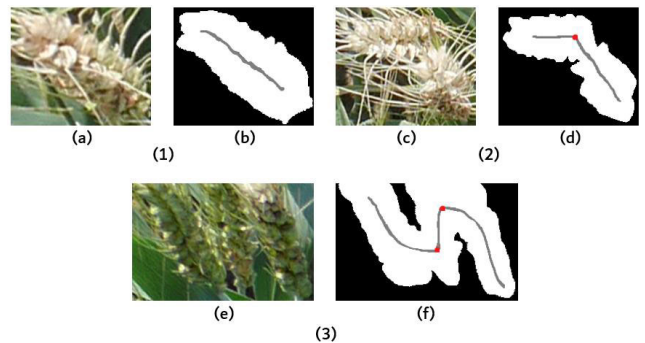
**B. IMAGE BINARIZATION USING OTSU ALGORITHM**

The output of the above segmentation model is a probability distribution map, which is a grayscale image. Therefore, threshold processing must be applied to the probability-distribution image to highlight the regions that are probably wheat ears. The resulting binary image of the wheat ears is more convenient for counting. The maximum inter-group variance method was proposed by Otsu [33], which is an adaptive threshold-determination method and is called the Otsu algorithm. This algorithm divides the image into two parts, background and target, by looking for the threshold  $T$  based on the grayscale characteristics of the image. The segmentation threshold of image foreground and background is denoted  $T$ . The fraction of foreground pixel points in the whole image is  $w_0$ , and its average grayscale is  $\mu_0$ . The fraction of background pixel points in the whole image is  $w_1$ , and its average grayscale is  $\mu_1$ .  $T$  is calculated as follows:

$$T = w_0 \times w_1 \times (\mu_1 - \mu_0) \times (\mu_0 - \mu_1) \quad (1)$$

**C. HARRIS-CORNER-DETECTION-BASED WHEAT-EAR COUNTING**

Counting wheat ears requires first identifying the wheat-ear skeletons, which is done by progressively scanning the binary wheat-ear image to determine boundary points. All boundary-point pixels are added to the boundary-point series.



**FIGURE 5.** Harris corner detection of wheat ears: (1) single wheat ear, (2) two adhered wheat ears, (3) three adhered wheat ears. Panels (a), (c), and (e) show RGB images of adhered wheat ears. Panels (b), (d), and (f) show corner-detection images of adhered wheat ears. The figure shows examples of various adhesion conditions. White is the wheat-ear area, gray is the wheat-ear skeleton, and red points are corner points.

Next, Zhang’s fast parallel algorithm is used to judge whether the boundary point can be deleted: If so, the point is directly deleted; if not, the point is retained. After scanning, the image is refined. Next, Harris corner detection is used to count the number of adhesive wheat ears. As shown in Figure 5, a single wheat ear has no corner point; two adhered wheat ears have one corner point; three adhered wheat ears have two corner points, etc. Thus,  $n$  corner points in a single connected region indicate  $n + 1$  wheat ears. In this way, the number of corner points in the image is counted to obtain the number of wheat ears in the image.

**D. CRITERIA FOR EVALUATION ALGORITHM**

Four metrics: SSIM, Precision, Recall, and F-measure, were used to evaluate segmentation quality [34], [35]. SSIM describes the similarity between the segmented image and the real image: the higher its value, the more the two images are similar. Precision measures the accuracy of the segmentation algorithm, and Recall measures the integrity of the segmentation image. Finally, F-measure is used to balance Precision

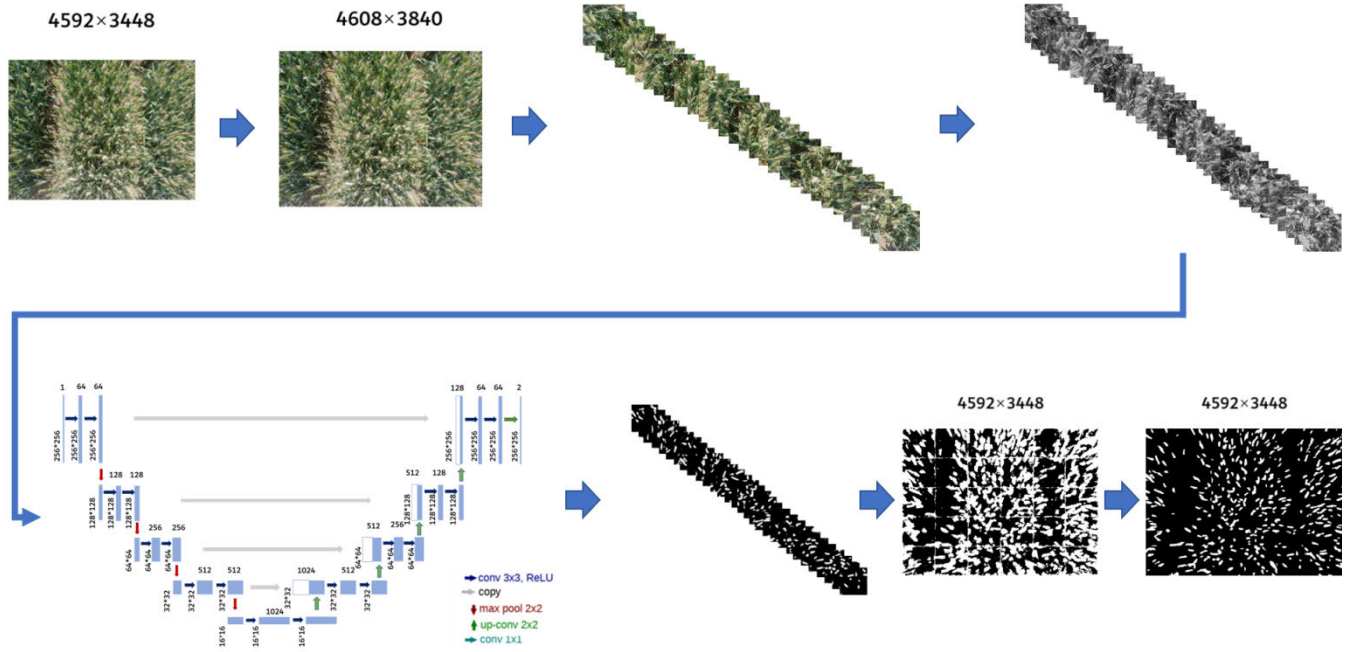


FIGURE 6. Wheat-ear segmentation.

and Recall. A higher F-measure is indicative of better segmentation. The segmentation time is used to evaluate the efficiency of the algorithm. These metrics can be calculated as follows:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\%. \quad (5)$$

The accuracy, average accuracy (Acc),  $R^2$ , and RMSE are used as metrics to evaluate the counting performance [8]. Accuracy, Acc, and  $R^2$  values closer to unity indicate better performance, as do smaller RMSE values:

$$Accuracy = \frac{|z_i - c_i|}{z_i} \times 100\% \quad (6)$$

$$Acc = \frac{1}{n} \sum_{i=1}^n \frac{|z_i - c_i|}{z_i} \times 100\% \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (z_i - c_i)^2}{\sum_{i=1}^n (z_i - \bar{z}_i)^2} \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (z_i - c_i)^2}{n}} \quad (9)$$

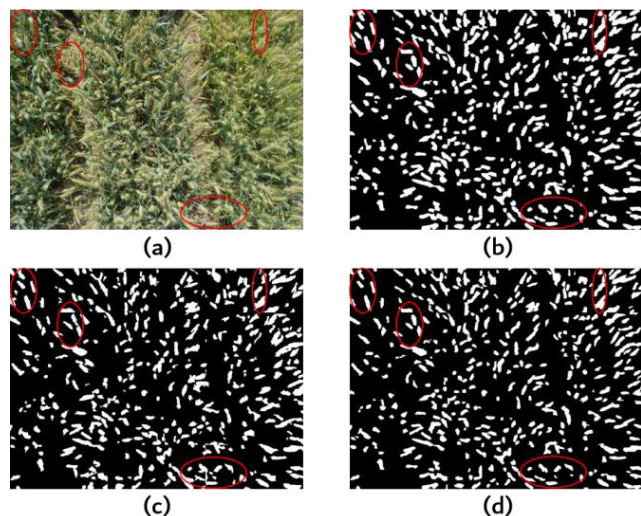
where  $x$  and  $y$  refer to the two images being compared,  $\mu_x$  ( $\mu_y$ ) is the mean of image  $x$  ( $y$ ),  $\sigma_x^2$  ( $\sigma_y^2$ ) is the variance of image  $x$  ( $y$ ),  $\sigma_{xy}^2$  is the covariance of image  $x$  and  $y$ , and  $c_1 = k_1^2 L^2$ ,  $c_2 = k_2^2 L^2$  are constants that keeps things stable, with  $L$  being the dynamic range of pixel values, and  $k_1 = 0.01$ ,  $k_2 = 0.03$ .  $TP$  is the predicted number of wheat ears, and the corresponding real results are all wheat-ear pixels.  $FP$  (false positive) is the number of pixels classified as wheat-ear pixels, but the real results of these pixels are the background.  $FN$  (false negative) is the number of pixels that belongs to the real results but are not correctly identified. Finally,  $n$  is the number of images in the testing set,  $z_i$  is the number of wheat ears counted in image  $i$ ,  $\bar{z}_i$  is the average number of wheat ears per image, and  $c_i$  is the predicted number of wheat ears in image  $i$ .

#### IV. RESULTS

The proposed method is tested by comparing its results to the number of wheat ears counted on the ground. The wheat-ear segmentation model was developed and coded in PyCharm 2017 (Python 3.5.4, OS: Ubuntu 18.04 64-bit, CPU: Intel i7-6800K 3.40GHz, GPU: Nvidia GeForce GTX 1080Ti, RAM: 16 GB). The wheat-ear counting algorithm was developed in Matlab R2017a (Windows 10, CPU: Intel i7-6800K 3.40GHz, GPU: Nvidia GeForce GTX 1080Ti, RAM: 16 GB).

##### A. RESULTS OF WHEAT-EAR SEGMENTATION

The testing images were input into the segmentation model, and the segmentation result was spliced and resampled to 4592x3448 pixels. The Otsu algorithm was applied to binarize the image (Figure 6).



**FIGURE 7.** Example of field wheat-ear segmentation: (a) original image, (b) ground truth of original image, (c) segmentation result of method of Zhou *et al.*, (d) segmentation result of method in this paper. The red elliptical regions indicate the segmentation errors.

This work uses the segmentation method proposed by Zhou *et al.* [13] to analyze the advantages and disadvantages of the proposed segmentation method. Zhou *et al.* compared and analyzed a variety of segmentation methods, and their results were better than those of the other methods. Thus, we compare the results of the proposed method only with those of the method of Zhou *et al.* Figure 7 shows the segmentation results of the proposed method.

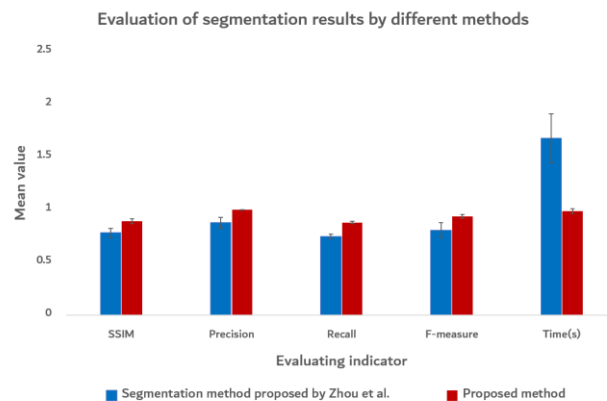
As shown in Figure 7, both the segmentation method proposed in this paper and the segmentation method of Zhou *et al.* produce a certain segmentation error when segmenting wheat ears. To better evaluate the segmentation results, the SSIM, Precision, Recall, and F-measure are used to evaluate the 60 image-segmentation results. In addition, the segmentation time is also used to evaluate the efficiency of the proposed algorithm [36]. The results are shown in Figure 8.

As shown in Figure 8, the SSIM, Precision, Recall, F-measure, and average segmentation time of the proposed method are 0.890, 0.999, 0.878, 0.935, and 0.984 s, respectively. The proposed segmentation method thus performs better than the method of Zhou *et al.* for segmenting wheat ears in the field. Compared with the method of Zhou *et al.*, there is more than a 10% improvement in each evaluation index. The proposed segmentation method has little up-down deviation and good stability, which allows efficient segmentation of wheat ears in field conditions.

## B. RESULTS OF WHEAT-EAR COUNTING

Based on the binary image of the segmentation results, the wheat-ear skeletons are identified. The number of wheat ears is then counted by using Harris corner detection (Figure 9).

As shown in Figure 9, Harris corner detection can effectively detect the adhered wheat ears and thereby properly count the number of wheat ears. Sixty wheat-ear images are used to compare the performance of the proposed method



**FIGURE 8.** Evaluation of segmentation results of each method. Each column represents the mean, and the error bars represent the standard deviation in the testing data.

with that of Zhou *et al.* [13]. Table 1 presents the results of counting, and Figure 10 combines the results of manual statistics,  $R^2$ , and the RMSE of each method.

Table 1 and Figure 10 show that the highest accuracy of the counting method in this work is unity, and the minimum accuracy is 0.924. Furthermore, Acc,  $R^2$ , and RMSE are 0.974, 0.983, and 14.043, respectively. For all images, the manual counting result is positively and strongly correlated with and automated counting result. Moreover, compared with the method of Zhou *et al.*, various metrics are improved upon using the proposed method.

## V. DISCUSSION

### A. ACCURACY AND ANTI-INTERFERENCE ANALYSIS OF WHEAT-EAR SEGMENTATION MODEL IN FIELD

We now discuss how to better evaluate the segmentation accuracy and anti-interference of the model. Figure 11 analyzes the performance of the segmentation model under complex conditions, such as wheat-ear occlusion, leaf occlusion, uneven illumination, and soil interference. The green, blue, brown, purple, and yellow frames indicate occlusions of wheat, uneven illumination, leaf occlusion, soil influence, and areas of segmentation error, respectively. Red frames indicate incorrect segmentations that affect the count. Because many segmentation errors appear in the figure, some are not marked. We used the proposed segmentation model and the method of Zhou *et al.* to segment 30 images with these interference factors. The segmentation results (see Figure 12) are then evaluated by SSIM, Precision, Recall, and F-measure.

According to Figure 11, the wheat-ear segmentation model proposed herein effectively segments the wheat ears and has a certain degree of anti-interference. Although some errors appear, the model basically segments each wheat ear. Faced with these interfering factors, the method of Zhou *et al.* produces large segmentation errors. As shown in Figure 12, the segmentation results of the proposed method are better than those of the method of Zhou *et al.* as per SSIM, Precision, Recall, and F-measure. The proposed segmentation model obtains  $SSIM = 0.983$ , indicating that the segmentation results are basically similar to the ground truth.

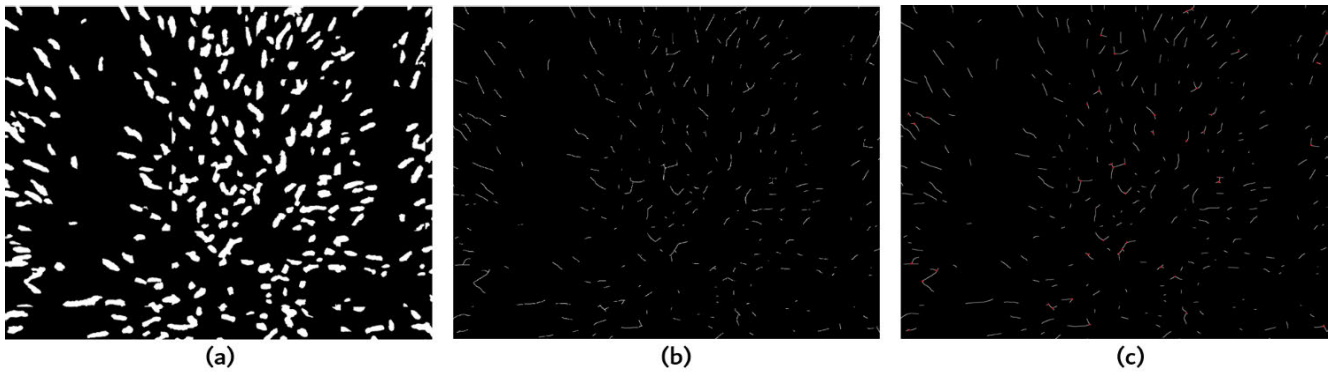


FIGURE 9. Wheat ears count results. (a) binary image of segmentation results, (b) wheat ears skeleton, (c) Harris corner.

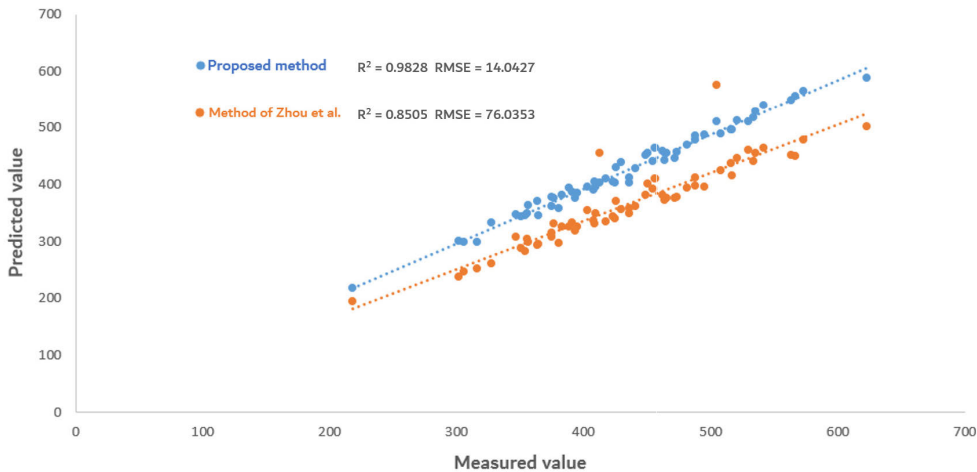


FIGURE 10. Each method counts  $R^2$  and RMSE of the results. Each color dot represents each method.

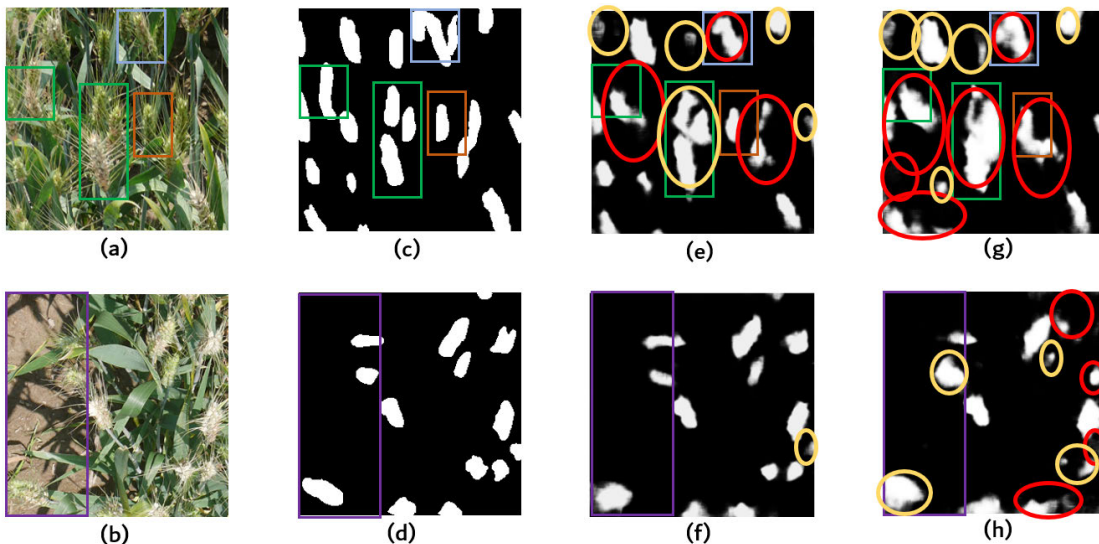
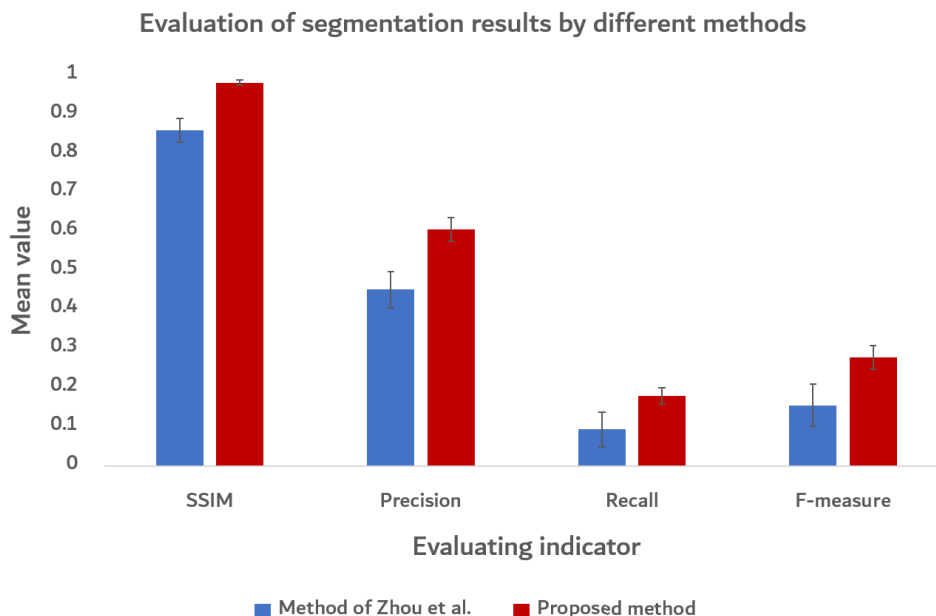


FIGURE 11. Accuracy and anti-interference analysis of wheat-ear segmentation model in field conditions. The images in the left column are with interference factors, the ground truth binary images are in the second column, the grayscale images output by the proposed segmentation model are in the third column, and the grayscale images output by the method of Zhou et al. are in the fourth column. The green, blue, brown, purple, and yellow frames indicate the areas where the wheat is occluded, uneven illumination areas, occlusion of leaves, soil-influence areas, and areas of segmentation error, respectively. The red frames indicate incorrect segmentations that affect the count.

Because deep learning uses deep neural networks and a large number of training designs, it can better learn and mine the deep features of images [37]. Compared with traditional

machine learning methods, it has significant advantages in terms of target recognition. Therefore, deep learning is used to segment wheat ears based on leveraging a large number



**FIGURE 12.** Anti-interference analysis of each method. Each column indicates a mean, and the error bars indicate the up-down deviation.

of manually labeled datasets. Its segmentation accuracy and anti-interference capability are better than those of traditional machine learning methods. However, Precision, Recall, and F-measure are poor, indicating that there are a lot of errors in the segmentation results compared with the ground truth.

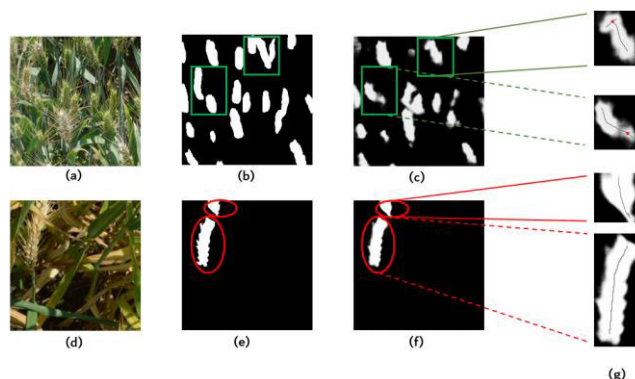
The model segmentation error for the proposed method may be related to network or dataset selection. The error may be due to the relatively simple network structure and the fact that wheat-ear depth features cannot be fully mined, leading to the error of the network extraction of wheat-ear features. Further study should use a more complex network such as Mask r-cnn [38] or DeepLab v3 [39]. In addition, the dataset does not adequately reflect the wheat ears in the field environment. Therefore, the dataset should be expanded to learn the characteristics of wheat ears in the field.

**B. ANALYSIS OF WHEAT-EAR COUNT**

Table 1 gives the count for each image is displayed. Because the segmentation model accurately segments the wheat ears in the field, the counts obtained by the proposed method are close to the ground truth, and the accuracy exceeds that of the method of Zhou *et al.*

In this work, most of the image counts obtained by using the proposed method are higher than the ground truth, which may be because leaves split complete wheat ears and increase the connected region, leading to counts that exceed the ground truth. Most of the image counts are lower than the ground truth. This may be due to occlusions and to large shadows of wheat ears, leading to segmentation error and resulting in slightly lower counts.

As shown in Figure 13, wheat-ear occlusion may cause under-segmentation of adhered wheat ears, leading to counting error. Shadows may also cause segmentation errors,



**FIGURE 13.** Analysis of counting results for each image. The left column shows images with interference factors, the second column shows ground truth binary images, the third column shows the grayscale image from the proposed segmentation model, and the fourth column shows wheat-ear skeletons and Harris corner detection. The green rectangles show occluded wheat ears and shadows that affect a larger part of the image, and the red ovals show the segmentation of a wheat ear separated by a leaf.

causing more under-segmentation of the wheat-ear portion of an image. In Figure 13, these effects caused under-segmentation of the wheat ears, resulting in a lower count. The wheat ear occluded by the leaf may split in two a single connected area, resulting in multiple counts.

**C. EFFECT OF DIFFERENT FIELD OF VIEW ON RECOGNITION ACCURACY**

To study how field size affects the recognition accuracy, we determine the center points of 60 images. The side length of the square area is determined using 0.6 to 2.2 m intervals. The number of wheat ears in each area is counted. The influence of the field size on recognition accuracy is determined from the counts obtained by the proposed method.



TABLE 1. Wheat ears count results of each image.

Image number	Measured	Proposed method	Accuracy	Method of Zhou et al.	Accuracy	Image number	Measured	Proposed method	Accuracy	Method of Zhou et al.	Accuracy
1	563	549	0.975	451	0.822	31	418	409	0.979	334	0.799
2	413	402	0.973	455	0.898	32	395	385	0.979	326	0.825
3	364	371	0.981	293	0.805	33	356	349	0.980	304	0.854
4	306	299	0.977	247	0.826	34	423	404	0.955	343	0.811
5	302	301	0.997	238	0.788	35	391	387	0.990	332	0.849
6	328	333	0.985	261	0.796	36	354	346	0.977	283	0.799
7	516	496	0.961	436	0.845	37	383	381	0.995	326	0.851
8	465	455	0.979	375	0.807	38	357	364	0.980	299	0.838
9	534	517	0.968	441	0.826	39	472	445	0.943	375	0.795
10	567	555	0.978	450	0.794	40	573	564	0.984	478	0.834
11	530	511	0.964	460	0.868	41	521	513	0.984	446	0.856
12	623	587	0.942	502	0.806	42	535	528	0.987	455	0.851
13	542	539	0.995	463	0.854	43	449	452	0.993	382	0.851
14	505	510	0.990	575	0.861	44	381	358	0.940	297	0.780
15	488	486	0.996	412	0.844	45	316	299	0.943	252	0.798
16	456	464	0.983	410	0.899	46	365	345	0.945	295	0.808
17	473	456	0.963	378	0.799	47	410	396	0.965	348	0.849
18	403	396	0.983	354	0.878	48	441	427	0.968	362	0.821
19	394	375	0.952	318	0.807	49	218	217	0.995	194	0.890
20	517	496	0.959	415	0.803	50	455	440	0.966	392	0.862
21	508	489	0.963	424	0.835	51	389	393	0.990	326	0.838
22	464	443	0.953	373	0.804	52	425	402	0.946	340	0.800
23	462	458	0.991	381	0.825	53	409	404	0.988	330	0.807
24	481	470	0.977	394	0.819	54	347	347	1	308	0.888
25	495	488	0.986	396	0.800	55	436	411	0.943	349	0.801
26	451	455	0.991	401	0.889	56	375	377	0.995	315	0.840
27	430	439	0.979	356	0.828	57	377	376	0.997	331	0.878
28	426	429	0.993	370	0.869	58	436	403	0.924	358	0.821
29	488	478	0.980	397	0.814	59	408	390	0.956	336	0.824
30	351	343	0.977	288	0.821	60	375	361	0.963	307	0.819

As shown in Figure 14, the maximum  $R^2$  and RMSE are 0.912 and 22.841 when the field of view is 2.2 m $\times$ 2.2 m. As shown in Figure 15, the  $R^2$  significantly increases when the view increases from 0.8 to 1.0 m, from 1.6 to 1.8 m, and from 2.0 to 2.2 m. The RMSE increases significantly when the view increases from 0.6 to 0.8 m and from 1.0 to 2.0 m. It is concluded that the field of view is more reasonable at 1.0 m and 2.2 m, and the field of view is optimal at 2.2 m.

Because the field of view in this paper is limited to 0.6 to 2.2 m, we cannot discuss the counts for other fields of view. However, the results for  $R^2$  and RMSE with the field of

view between 0.6 and 2.2 m indicate that  $R^2$  and RMSE may increase with a larger field of view. However, the trend may change; for example, a larger field of view may result in a larger segmentation error and therefore a worse result. These results provide a certain point of reference for studying the optimal result within the field of view.

#### D. COMPARISON WITH OTHER COUNTING TECHNIQUE

Table 2 compares our results with those of the MCNN method [40], where MCNN has 4450 training sets and 1950

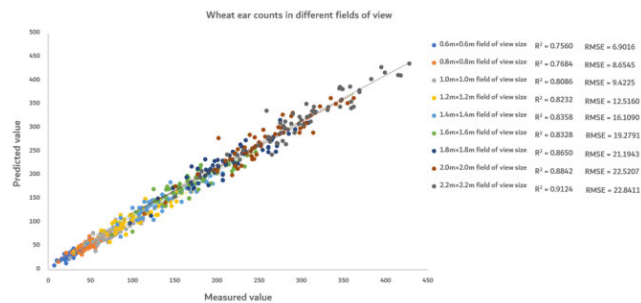


FIGURE 14. R<sup>2</sup> and RMSE for different field-of-view sizes.

TABLE 2. Wheat-ear counting from proposed method compared with mcn method.

Methods	Acc	R <sup>2</sup>	RMSE
MCNN	0.853	0.869	67.037
Proposed method	0.974	0.983	14.043

validation sets during training, with the parameters learning rate = 0.001, each patch = 10, epochs = 10. MCNN is designed to solve counting problems in cluster environments or dense populations by density regression. Because MCNN is based on density and differs from the processing used herein, we compare and analyze only the final counting results.

As shown in Table 2, MCNN provides good counting results for wheat ears, but the results are inferior to those of the proposed method. Because the training data used herein is a sub-image obtained by cropping, the wheat density in the sub-image is small and may not be applicable to the MCNN. Second, MCNN may have fewer training sessions, and the learning rate may not be optimal, which impacts the final counting results. In this experiment, MCNN counts the wheat ears in the sub-image and adds the count results of the 30 sub-images to obtain the final count result. Since a large number of the same wheat ears are cropped into the two sub-images when the original image is cropped, the final counting result is increased to cause a large counting error.

The proposed method divides wheat-ear counting into two processes. Considering the efficiency and accuracy of the detection, the two processes can be combined into one deep learning model. For example, Yu *et al.* proposed a deep convolutional neural network that supports region of interest; it combines the region of interest subnet and the classified subnet into a single deep learning model to effectively identify apple leaf disease [41]. Inspired by this method, wheat-ear counting and image segmentation can be combined into one model and then trained in an end-to-end manner. However, to properly merge the two subnets, the loss function setting is important, and significant experimentation is required to determine the best loss function.

This type of thinking guides our next experiment, which is designed to determine whether the subnet of the wheat-ear

count can be realized by deleting the classification layer of the VGG subnet and then adding the regression layer. The number of wheat ears can be entered into the regression layer to provide the true value on the ground. A loss function is also designed to merge the two subnets and achieve more accurate counts.

## VI. CONCLUSION

This research investigates wheat-ear counting in a field environment and proposes an efficient automated counting method that uses RGB images. First, we constructed a dataset of 120 wheat-ear images manually marked. Next, a wheat-ear segmentation model was established based on a FCN. The model offers high segmentation accuracy and good anti-interference capacity for wheat-ear segmentation under a complex environment. The wheat ears are counted by applying Harris corner detection to wheat-ear skeletons. This work compares the proposed wheat-ear segmentation model with previous machine learning methods. The results show that the proposed segmentation model offers a significant advantage for segmenting wheat ears. The proposed method provides accurate counts and robust anti-interference capacity, which provides effective data support for estimating crop yield and for breeding wheat varieties.

This study demonstrates that FCN can be effectively applied to wheat-ear segmentation under field conditions, and that the wheat-ear count can be detected by using Harris corner detection. In addition, the method has low implementation costs. Provided the wheat-ear image data acquired in the field are collected by a digital camera, the corresponding number of wheat ears can be obtained. The results of this study also support wheat phenotype and breeding research.

## ACKNOWLEDGMENT

D. Wang analyzed the data and drafted the article. Y. Fu, G. Yang, D. Zhang, C. Zhou, and X. Yang reviewed and edited the article. D. Liang and X. Yang designed the experiments. N. Zhang and H. Wu collected experimental data. All authors gave final approval for publication. The funders had no role in choosing the study design or in the collection, analysis, and interpretation of the data, in the writing of the report, or in the decision to submit the article for publication.

## REFERENCES

- [1] T. Curtis and N. G. Halford, "Food security: The challenge of increasing wheat yield and the importance of not compromising food safety," *Ann. Appl. Biol.*, vol. 164, no. 3, pp. 354–372, May 2014.
- [2] S. Ayas, H. Dogan, E. Gedikli, and M. Ekinci, "Microscopic image segmentation based on firefly algorithm for detection of tuberculosis bacteria," in *Proc. 23rd Signal Process. Commun. Appl. Conf. (SIU)*, Malatya, Turkey, May 2015, pp. 851–854, doi: 10.1109/SIU.2015.7129962.
- [3] P. D. Hollins, P. S. Kettlewell, S. T. Parsons, and M. D. Atkinson, "The impact of supply, demand and grain quality on the UK bread and feed wheat price differential in the UK," *J. Agricult. Sci.*, vol. 144, no. 144, pp. 411–419, Oct. 2006.
- [4] P. Bognár, A. Kern, S. Pásztor, J. Lichtenberger, D. Koronczay, and C. Ferencz, "Yield estimation and forecasting for winter wheat in Hungary using time series of MODIS data," *Int. J. Remote Sens.*, vol. 38, no. 11, pp. 3394–3414, May 2017.

- [5] G. Desheva and S. Kachakova, "Correlations between the main structural elements of yield in common wheat cultivars," *Plant Sci.*, to be published.
- [6] Y. Zhu, Z. Cao, H. Lu, Y. Li, and Y. Xiao, "In-field automatic observation of wheat heading stage using computer vision," *Biosyst. Eng.*, vol. 143, pp. 28–41, Mar. 2016.
- [7] J. A. Fernandez-Gallego, S. C. Kefauver, N. A. Gutiérrez, M. T. Nieto-Taladriz, and J. L. Araus, "Wheat ear counting in-field conditions: High throughput and low-cost approach using RGB images," *Plant Methods*, vol. 2018, no. 1, p. 22, Mar. 2018.
- [8] J. Wu, G. Yang, X. Yang, B. Xu, L. Han, and Y. Zhu, "Automatic counting of in situ rice seedlings from UAV images based on a deep fully convolutional neural network," *Remote Sens*, vol. 11, no. 6, p. 691, Mar. 2019.
- [9] W. Maldonado and J. C. Barbosa, "Automatic green fruit counting in orange trees using digital images," *Comput. Electron. Agricult.*, vol. 127, pp. 572–581, Sep. 2016.
- [10] S. H. Mussavi, K. Alamisaid, G. Fathi, A. Siahpoosh, and M. H. Gharineh, "Effect of seed density and molinit rates on barnyardgrass (*Echinochloa crus-galli*) control in direct-seeded rice in Ahwaz," *Agronomy J.*, vol. 1, no. 90, pp. 83–92, 2011.
- [11] H. Fan, F. Xie, Y. Li, Z. Jiang, and J. Liu, "Automatic segmentation of dermoscopy images using saliency combined with Otsu threshold," *Comput. Biol. Med.*, vol. 85, pp. 75–85, Jun. 2017.
- [12] P. Ganesan, V. Rajini, and R. I. Rajkumar, "Segmentation and edge detection of color images using CIELAB color space and edge detectors," in *Proc. INTERACT*, Dec. 2010, pp. 393–397.
- [13] C. Zhou, D. Liang, X. Yang, H. Yang, J. Yue, and G. Yang, "Wheat ears counting in field conditions based on multi-feature optimization and TWSVM," *Frontiers Plant Sci.*, vol. 9, p. 1024, Jul. 2018.
- [14] Q. Li, J. Cai, B. Berger, M. Okamoto, and S. J. Miklavcic, "Detecting spikes of wheat plants using neural networks with laws texture energy," *Plant Methods*, vol. 13, no. 1, p. 83, Oct. 2017.
- [15] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, "Deep learning for brain MRI segmentation: State of the art and future directions," *J. Digit. Imag.*, vol. 30, no. 4, pp. 449–459, 2017.
- [16] S. Bargoti and J. P. Underwood, "Image segmentation for fruit detection and yield estimation in apple orchards," *J. Field Robot.*, to be published.
- [17] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1520–1528.
- [18] M. Martin-Abadal, E. Guerrero-Font, F. Bonin-Font, and Y. Gonzalez-Cid, "Deep semantic segmentation in an AUV for Online Posidonia Oceanica meadows identification," *IEEE Access*, vol. 6, pp. 60956–60967, 2018.
- [19] Y. Bai, E. Mas, and S. Koshimura, "Towards operational satellite-based damage-mapping using U-Net convolutional network: A case study of 2011 Tohoku Earthquake-Tsunami," *Remote Sens.*, vol. 10, no. 10, p. 1626, 2018.
- [20] X. Jiang, Y. Gao, Z. Fang, P. Wang, and B. Huang, "An end-to-end human segmentation by region proposed fully convolutional network," *IEEE Access*, vol. 7, pp. 16395–16405, 2019.
- [21] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P. A. Heng, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.
- [22] Y. Fang and R. P. Ramasamy, "Current and prospective methods for plant disease detection," *Biosensors*, vol. 5, no. 3, pp. 537–561, Aug. 2015.
- [23] Z. Ning, W. Shen, and C. Xiong, "Adhesion ore image separation method based on concave points matching," *Inf. Technol. Intell. Transp. Syst.*, vol. 455, pp. 153–164, Nov. 2016.
- [24] F. Mokhtarian and R. Suomela, "Robust image corner detection through curvature scale space," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1376–1381, Dec. 1998.
- [25] J. Y. Chen, Y. Liu, N. Li, and K. P. Lim, "A method for separating adhesive vehicles based on corner detection," in *Proc. Int. Conf. Inf. Commun. Technol.*, Apr. 2015, pp. 1–5.
- [26] P. Wang, X. Hu, Y. Li, Q. Liu, and X. Zhu, "Automatic cell nuclei segmentation and classification of breast cancer histopathology images," *Signal Process.*, vol. 122, pp. 1–13, May 2015.
- [27] P. Ram and S. Padmavathi, "Analysis of Harris corner detection for color images," in *Proc. Int. Conf. Signal Process., Commun., Power Embedded Syst. (SCOPE5)*, Oct. 2017.
- [28] E. Dougherty and R. A. Lotufo, *Hands-On Morphological Image Processing*. Bellingham, WA, USA: SPIE, 2003, doi: 10.1117/3.501104.
- [29] E. J. Kirkland, "Bilinear interpolation," in *Advanced Computing in Electron Microscopy*. Boston, MA, USA: Springer, 2010, pp. 261–263.
- [30] E. R. Dougherty and R. A. Lotufo, "Gray-scale morphology," Tech. Rep., 2003.
- [31] T. Y. Zhang and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Commun. ACM*, vol. 27, no. 3, pp. 236–239, 1984.
- [32] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Nov. 2015, pp. 234–241.
- [33] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, p. 62–66, Jan. 1979.
- [34] X. Xiong, L. Duan, L. Liu, H. Tu, P. Yang, D. Wu, G. Chen, L. Xiong, W. Yang, and Q. Liu, "Panicle-SEG: A robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization," *Plant Methods*, vol. 13, no. 1, p. 104, Nov. 2017.
- [35] B. Ma, Z. Liu, F. Jiang, Y. Yan, J. Yuan, and S. Bu, "Vehicle detection in aerial images using rotation-invariant cascaded forest," *IEEE Access*, vol. 7, pp. 59613–59623, 2019.
- [36] C. Yu, B. Jin, Y. Lu, X. Chen, Z. Yi, Z. Kai, and S. Wang, "Multi-threshold image segmentation based on firefly algorithm," in *Proc. 9th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Beijing, China, Oct. 2013, pp. 16–18.
- [37] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, Apr. 2016.
- [38] K. He, G. Georgia, D. Piotr, and G. Ross, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2961–6969.
- [39] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *Comput. Sci.*, vol. 2014, no. 4, pp. 357–361, Dec. 2014.
- [40] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 589–597.
- [41] H.-J. Yu and C.-H. Son, "Apple leaf disease identification through region-of-interest-aware deep convolutional neural network," 2019, *arXiv:1903.10356*. [Online]. Available: <https://arxiv.org/abs/1903.10356>



**DAOYONG WANG** is currently pursuing the master's degree with the School of Electronic Information Engineering, Anhui University, and the Beijing Research Center for Information Technology in Agriculture. His current research interests include computer vision, image processing, and deep learning.



**YUANYUAN FU** received the Ph.D. degree from the School of Environment and Resources, Zhejiang University, in 2015. Her research interests include remote sensing image processing and agricultural quantitative remote sensing.



**GUIJUN YANG** received the Ph.D. degree in cartography and geographic information system from the State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Applications (IRSA), Chinese Academy of Sciences (CAS), Beijing, China, in 2008. He is currently a Research Associate with the National Engineering Research Center for Information Technology in Agriculture (NERCITA), Beijing. His research interests include radiative transfer modeling, imagery simulation, atmospheric correction, quantitative inversion, big data, and cloud computing.



**XIAODONG YANG** received the Ph.D. degree from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, in 2008. He is currently an Associate Research Fellow with the Beijing Agricultural Information Technology Research Center. His research interests are agricultural remote sensing and 3S integration.



**NING ZHANG** received the Ph.D. degree from Beijing Forestry University, in 2017. Her research interests include agricultural and forestry pest research.



**DONG LIANG** received the Ph.D. degree from Anhui University, in 2004. He is currently the Dean of the School of Electronic Information Engineering, Anhui University. His research interests include computer signal processing, intelligence, and processing technology.



**HONGYA WU** is currently a Researcher with the Agricultural Institute of Lixiahe, Jiangsu, where he is also the Deputy Director of the Wheat Research Office and the General Secretary of the Communist Youth League. His research interest is wheat genetic breeding.



**CHENGQUAN ZHOU** received the Ph.D. degree from the School of Electronics and Information Engineering, Anhui University, in 2018. His research interests include image processing and crop phenotypic research.



**DONGYAN ZHANG** received the Ph.D. degree from the School of Environment and Resources, Zhejiang University, in 2002. His research interests include remote sensing information processing and analysis, intelligent sensor design and development, and big data mining and application.

...