

Segmenting ears of winter wheat at flowering stage using digital images and deep learning

Juncheng Ma^a, Yunxia Li^b, Keming Du^a, Feixiang Zheng^a, Lingxian Zhang^{b,*}, Zhihong Gong^c, Weihua Jiao^d

^a Institute of Environment and Sustainable Development in Agriculture, Chinese Academy of Agricultural Sciences, Beijing 100081, China

^b College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

^c Tianjin Climate Center, Tianjin 300074, China

^d Center for Agricultural and Rural Economic Research, Shandong University of Finance and Economics, Jinan 250014, China



ARTICLE INFO

Keywords:

Ears of winter wheat
RGB images
Deep convolutional neural network
Fully convolutional network
Segmentation

ABSTRACT

Segmenting ears of winter wheat from canopy images was considered to be an important procedure prior to the extraction of related traits. Current segmentation method based on computer vision was susceptible to noise, which is limited in practical applications. In this study, a two-stage segmentation method for ears of winter wheat based on digital images of unit ground area and the state-of-the-art deep learning techniques was proposed. In the coarse segmentation stage, a deep convolutional neural network (DCNN) was constructed to classify the superpixels generated by entropy rate superpixel algorithm, achieving the coarse results. In the fine segmentation stage, a fully convolutional network (FCN) allowing pixel-wise semantic segmentation was constructed to eliminate the non-ear pixels in the coarse results. To compare the results of the proposed two-stage segmentation method, conventionally adopted methods for image segmentation were used. Results showed that the proposed two-stage segmentation method was able to accurately segmenting ears of winter wheat from canopy images captured at flowering stage ($Q_{seg} = 0.7197$, $F1$ score = 83.70%, $SSIM = 0.8605$), outperforming the other compared methods. Generalization tests were conducted to evaluate the utility of the proposed two-stage segmentation method. Results showed that the two-stage segmentation method was still capable of accurately segmenting ears of winter wheat, even though the performance slightly decreased. Change of winter wheat cultivar and lack of descriptive information were two factors that could degrade the performance of the two-stage segmentation method. Tests of the methods on Unmanned Aerial Vehicle (UAV) based RGB images showed the Fully Convolutional Network stride 8 predictions (FCN-8s) had a good chance to achieve satisfactory performances on UAV based canopy images.

1. Introduction

Traits associated with ears of winter wheat, such as size and shape, are critical in phenotyping tasks. However, conventional methods for field phenotyping of ears are tedious and costly of human efforts, which are limited in the large-scale and high-throughput applications. With the development of computer vision technology, image-based techniques provide a promising way to extract traits related to ears of winter wheat (Fernandez-Gallego et al., 2018; Madec et al., 2018; Zhou et al., 2018; Lu et al., 2017; Xiong et al., 2017; Zhu et al., 2016). Extracting ears from canopy images captured by handheld mobile device or automatic phenotyping platforms enabled to measure the traits associated

with ears under field conditions, thus increasing the throughput for field phenotyping (Rico-Fernández et al., 2019; Fernandez-Gallego et al., 2018; Zhou et al., 2018; Hu et al., 2017; Xiong et al., 2017; Duan et al., 2015). Therefore, segmenting ears of winter wheat from canopy images was considered to be an important procedure prior to the extraction of the traits. Zhu et al. (2016) proposed a wheat ear detection method to observe the heading stage of wheat. The images of wheat canopy were captured by in-field image acquisition devices and processed by a collection of image processing techniques, generating a feature set for classification by shallow machine learning method. Fernandez-Gallego et al., (2018) proposed an ear-counting algorithm to estimate ear density under field conditions based on digital image

* Corresponding author at: College of Information and Electrical Engineering, China Agricultural University, 17, Qinghua East Road, Haidian District, Beijing 100083, China.

E-mail address: zhanglx@cau.edu.cn (L. Zhang).

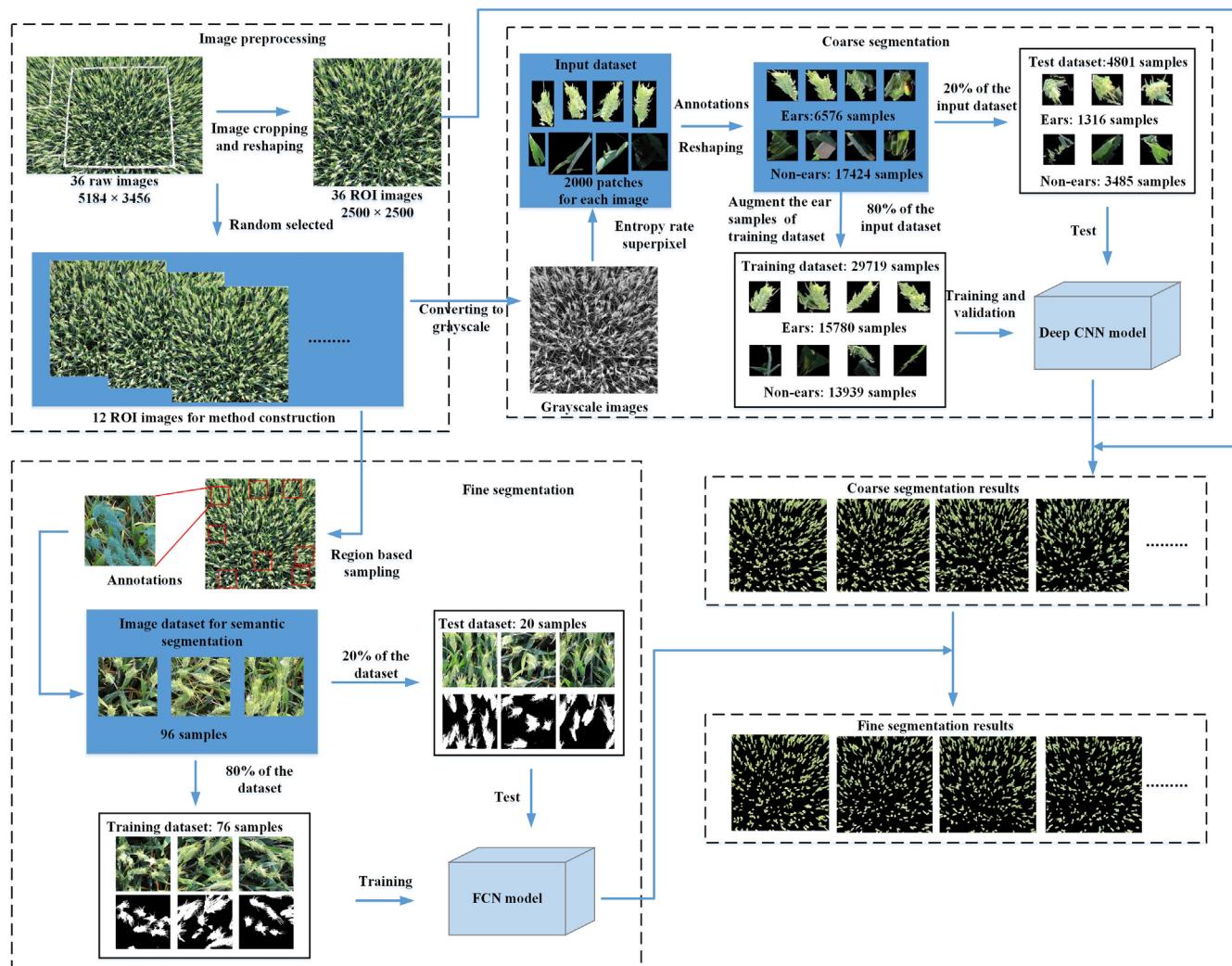


Fig. 1. Pipeline of the two-stage wheat ear segmentation method.

processing. The method can automatically process the images captured in field conditions and count the number of ears per unit ground area without human intervention. Zhou et al. (2018) proposed a computer vision based method to recognize wheat ears from images captured from the side view at 45° above the horizontal. Following the approximately same pipelines in Zhu et al., (2016), the method extracted low-level image features and adopted shallow machine learning, twin-support-vector-machine, to classify the image pixels into two categories. Although good accuracy was reported in the above work, it was known that noises from field conditions, such as strong illumination and clutter background, would severely influence the accuracy of image processing techniques (Ma et al., 2018, 2017; Hamuda et al., 2017), moreover, the manually designed image features had weak generalization ability (Ma et al., 2019; Uzal et al., 2018; Tang et al., 2017), thus hindering the practical application of these methods.

Deep Learning is currently the state-of-the-art technique for image processing, which has been proved to be a powerful tool for field phenotyping of ears (Ma et al., 2019; Ghosal et al., 2018; Madec et al., 2018; Ferreira et al., 2017; Lu et al., 2017; Xiong et al., 2017). Xiong et al. (2017) proposed a segmentation method for rice panicle in field conditions. The authors used SLIC superpixel to build the input dataset for Convolutional Neural Network (CNN). In the test stage, the method adopted entropy rate superpixel for optimization, achieving better segmentation results than the compared approaches. However, the image collection of the method required an imaging bracket with black

background to screen the excessive light and unwanted image pixels, which limited the practical application in field conditions. Madec et al. (2018) proposed an ear density estimating method by using Faster-RCNN (Ren et al., 2015) and RGB images of high spatial resolution. In order to show the validity of the proposed estimating method, the authors adopted TasselNet proposed by Lu et al. (2017) for comparison. TasselNet was a local count regression network based on CNN, which was constructed for in-field counting of maize tassels. Both of the two approaches achieved good accuracy, however, these two methods were focusing on counting the number of ears. The TasselNet directly estimated the value of the estimated ear density from a single canopy image while the Faster-RCNN marked the position of all the ears in a single canopy image, both of which were not able to extract ears from a single canopy image precisely. Therefore, it was necessary to explore a more feasible and precise approach.

In this paper, a two-stage segmentation method for ears of winter wheat based on digital images of unit ground area and the state-of-the-art deep learning techniques was proposed. In the coarse segmentation stage, the entropy rate superpixel algorithm was performed to generate patches of ear and non-ear, aiming to highlight the ears from canopy image and build the input dataset for CNN classification. Based on the input dataset, a deep convolutional neural network (DCNN) was constructed to perform coarse segmentation. In the fine segmentation stage, a fully convolutional network (FCN) which allowed pixel-wise semantic segmentation was constructed to eliminate the non-ear pixels,

thus optimizing the coarse segmentation results.

2. Methods

2.1. Experiment setup and image collection

The experiment field was located in the field station of Shangqiu Academy of Agriculture and Forestry Sciences in Shuangba, Henan, China (Lat: 34°32'21.1884"; N, Long: 115°43'8.0868"; E). Twelve plots of winter wheat, whose size was 5 m in length and 2.4 m in width, were sown on 14th October 2017. Three target plant densities, i.e., 120, 270, and 420 plants/m², were adopted, which meant that each plant density was applied to four replicates. The same amount of fertilizers was applied to all the plots. Within each plot, three white markers of size 1 m × 1 m were laid down on the canopy with certain distances between every two of them. For more details about the Experiment setup, readers are referred to Ma et al. (2019).

The images were captured by a Canon EOS 600D digital camera which was mounted on a tripod. The camera oriented vertically downwards over the canopy at a distance of 1.5 m. Images were taken at a focal length of 18 mm with an aperture of f/4, resulting in a field of view of approximately seven rows of each plot. The canopy images were captured on May 2, 2018, which was the flowering stage of winter wheat.

According to the experimental setup, a dataset for season 2018 was constructed, consisting of 36 raw images (12 plots × 3 markers). The images were captured at 5184 × 3456 pixel resolution, which was then manually cropped to eliminate the pixels outside the markers and reshaped to 2500 × 2500 pixel resolution, generating 36 ROI images with a ground resolution of 0.16 mm (Fig. 1). Of the three ROI images for each plot, one ROI image was randomly selected, resulting in a dataset containing 12 ROI images for constructing the input dataset of deep learning methods.

2.2. Coarse segmentation

DCNN was able to achieve good classification with an input dataset covering as much variability in field conditions as possible. Many researchers have utilized superpixel segmentation to construct the input dataset for CNN, especially the Simple Linear Iterative Clustering (SLIC) superpixels (Zhou et al., 2018; Lu et al., 2017; Achanta et al., 2012). In this study, the coarse segmentation was expected to accurately detect all the ears in an ROI image, as well as to eliminate as many non-ear pixels in the coarse segmentation results as possible. Therefore, entropy rate superpixel (Liu et al., 2011) segmentation was adopted to build the input dataset for the DCNN in coarse segmentation since it was particularly useful in preserving edges and keeping the intrinsic homogeneity within the superpixels. Taking a number of 2000 superpixels for each ROI image, a dataset containing 24,000 images was constructed, which was then interactively labeled. Due to the entropy rate superpixel algorithm, the generated superpixels were homogeneous without too many noise pixels belonging to the other class, thus saving efforts for the interactive image labeling. The number of images for the two classes in the input dataset, i.e., ear and non-ear, was 6576 and 17,424 respectively, which was then divided into training and test datasets in a ratio of 8:2 by random selection. The training dataset consisted of 19,199 images, 5260 of which belonged to class ear and 13,939 belonged to class non-ear. The test dataset consisted of 4801 images, 1316 of which belonged to class ear and 3485 belonged to class non-ear. The statistics of the datasets used for the construction of the DCNN are shown in Table 1.

It is shown in Table 1 that the number of images for the two classes in the training dataset was severely unbalanced, which may cause the situation that the DCNN performs better on class non-ear than that on class ear (Ma et al., 2018; Ferreira et al., 2017). Therefore, it was necessary to balance the number for the two classes in the training dataset

prior to construction of the DCNN. For the purpose of improving the number of images for class ear, data augmentation was conducted. The augmentation method was used to translate the original image horizontally and vertically by 10 pixels respectively, which allowed for two augmented datasets. As a result of the augmentation, the number of images for class ear was improved to 15,780, roughly equaling to the number of images for class non-ear. Sequentially, an augmented training dataset consisting of 29,719 images was constructed, 23,775 of which were used for training and 5944 for validation (Table 1).

Considering the pixel resolution of the images in the input dataset, the architecture of the DCNN proposed in Ghosal et al. (2018) was adopted in this paper due to its extraordinary abilities to accurately classify images of small size (Fig. 2). The DCNN for coarse segmentation consisted of five convolution layers, five batch normalization layers, four max pooling layers, and two fully connected layers. For the detailed information about network structure and parameter setting, readers are referred to Ghosal et al. (2018).

Prior to DCNN training, all the images were reshaped to 64 × 64 pixel resolution to suit the input layer. The DCNN was then trained using a NVIDIA Quadro P4000 (8 GB memory) with CUDA 9.0. The stochastic gradient descent with momentum (SGDM) was used to optimize the network weights, whose momentum was set to 0.9 and remained constant for the training process. The learning rate was initialized as 0.001 and dropped every 10 epochs by a drop factor of 0.1. A mini-batch of 128 was used. L2 Regularization with a regularization factor of 0.0005 was adopted to decrease the chance of overfitting. The maximum number of epochs used for training was set to 200. When the test finished, the coarse segmentation of a test ROI image can be achieved by utilizing the pretrained DCNN to perform classification of the generated superpixels. As mentioned above, the superpixels were generated using the entropy rate superpixels algorithm, however, there were still pixels representing non-ear elements, such as the stem. Therefore, following the coarse segmentation of the ROI image, it was necessary to optimize the segmentation results to eliminate the non-ear pixels, i.e., the fine segmentation.

2.3. Fine segmentation

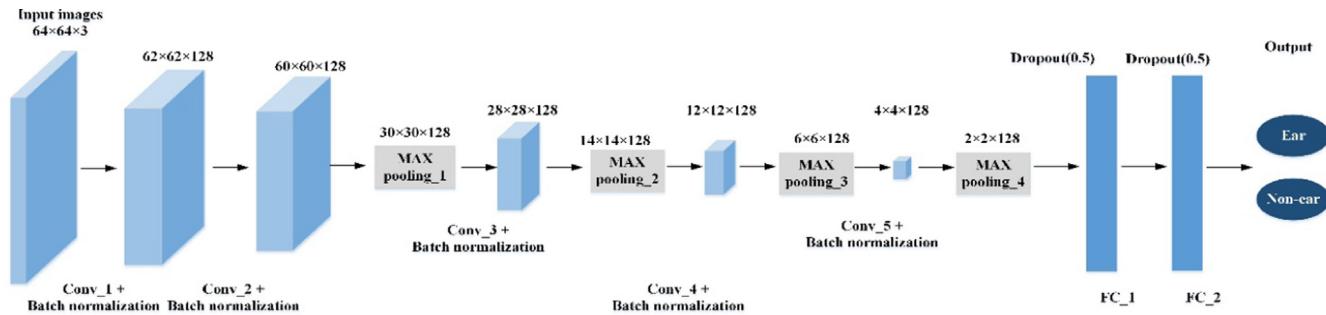
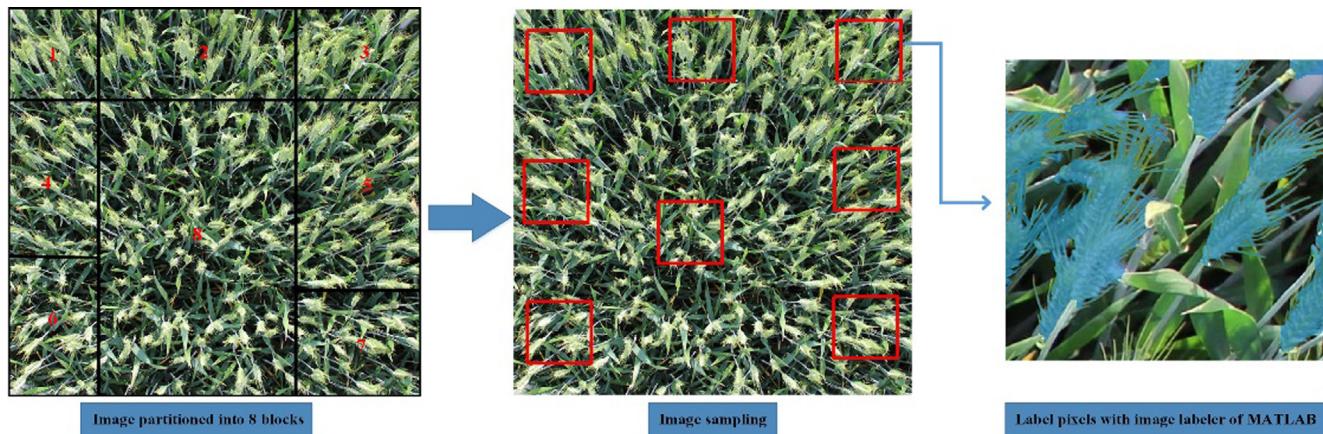
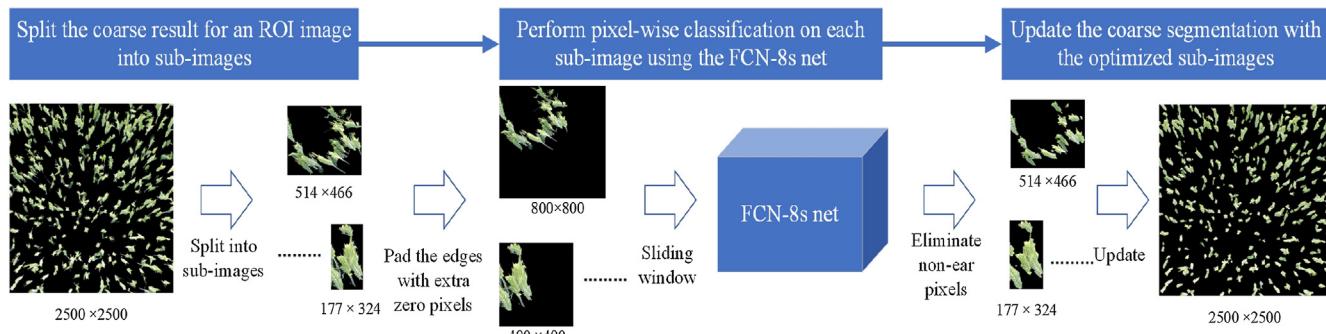
It was expected that the method for the fine segmentation should be capable of performing pixel-wise classification so that the noise pixels in the coarse segmentation results can be eliminated. Given the requirement of the fine segmentation, the fully convolutional network (FCN) (Long et al., 2015) was adopted for its remarkable abilities of pixel-wise prediction. In order to train an FCN, it was necessary to construct the labeled pixel dataset at first, which also utilized the 12 ROI images for method construction (Fig. 1). However, if all the pixels of the 12 ROI images were utilized for the labeled pixel dataset, a tremendous amount of work would go into image labeling. In order to improve the efficiency of the segmentation method, an image sampling strategy based on the ear orientation was developed (Fig. 3). It can be seen in Fig. 3 that an ROI image can be roughly partitioned into 8 blocks according to the ear orientation, thus making it possible to use one sub-image from each block to represent the whole ROI image.

Based on the above image sampling strategy, eight sub-images corresponding to eight blocks were used to represent one ROI image, resulting in a labeled pixel dataset containing 96 sub-images (twelve ROI images × eight blocks). The sub-image had a size of 400 × 400, which was then labeled using the Image Labeler App of Matlab (MathWorks Inc., USA). When the image labeling finished, the labeled pixel dataset was divided into two subsets, i.e., training and test. The training dataset consisted of 76 sub-images and the corresponding pixel labels. The test dataset consisted of 20 sub-images and the corresponding pixel labels. Image augmentation was applied to the training dataset. The augmentation method was to flip the image vertically and horizontally, as well as to translate the images vertically and horizontally by 10 pixels, resulting in 4 augmented datasets. The architecture of

Table 1

Statistics of the datasets used for the construction of the DCNN.

Categories	Original dataset		Augmentation		Augmented dataset		
	Training	Test	Training	Test	Training	Validation	Test
Ear	5260	1316	Two datasets	\	12,624	3156	1316
Non-ear	13,939	3485	\	\	11,151	2788	3485
Total	19,199	4801	10,520	\	23,775	5944	4801

**Fig. 2.** Architecture of DCNN (Ghosal et al., 2018).**Fig. 3.** Pipeline of the image sampling strategy.**Fig. 4.** Pipelines for fine segmentation.**Table 2**

Confusion matrix of the DCNN for coarse segmentation.

Class	Ear	Non-ear	Precision (%)	Sensitivity (%)	F1 score (%)
Ear	1249	108	92.0	94.9	93.4
Non-ear	67	3377	98.1	96.9	97.5
Accuracy (%)	96.4				

FCN in this paper was constructed using the VGG 16-layer net (Long et al., 2015; Simonyan and Zisserman, 2014). Specifically, the FCN-8s net was adopted for its excellent performance in semantic segmentation. The same GPU for DCNN training was also utilized to train the FCN. The network weights were optimized using the SGDM with a momentum of 0.9. The learning rate was initialized as 0.001 and remained constant for the training process. A mini-batch of 4 was used, as well as the L2 Regularization with a regularization factor of 0.0005. The

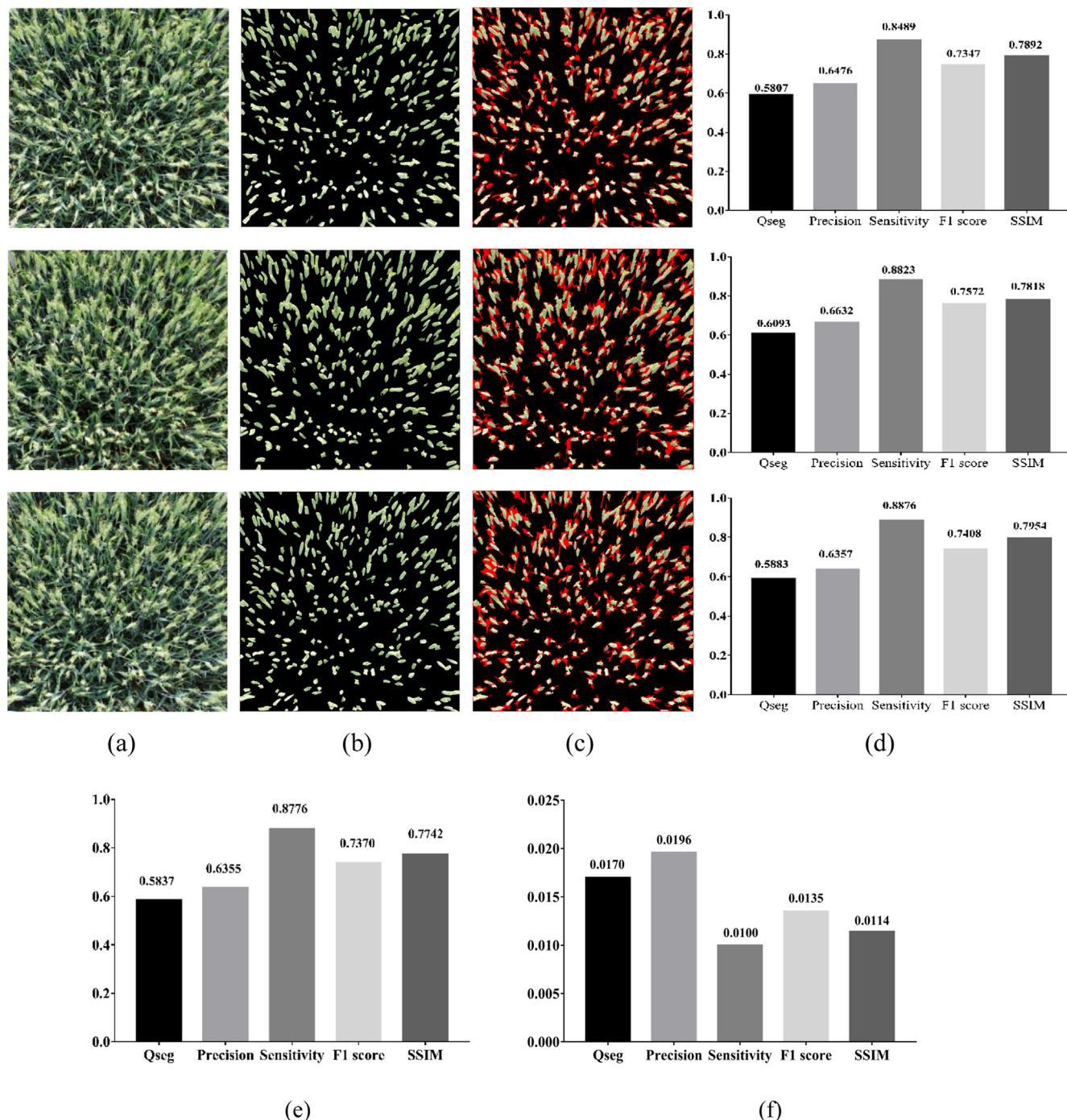


Fig. 5. Coarse segmentation results, (a) Original images, (b) Ground truth images, (c) Coarse segmentation results, the red pixels were the incorrectly classified pixels, (d) Evaluation metrics for the corresponding image, (e) Mean values of the evaluation metrics for coarse segmentation, (f) Standard deviations of the evaluation metrics for coarse segmentation.

Table 3
Confusion matrix of the FCN-8s net for fine segmentation.

Class	Ear	Non-ear	Precision (%)	Sensitivity (%)	F1 score (%)
Ear	760,733	135,038	84.92	86.75	85.83
Non-ear	116,162	2,188,067	94.96	94.19	94.57
Accuracy (%)	92.15				

maximum number of epochs used for training was set to 30. When the test of the FCN-8s net finished, the fine segmentation result for an ROI image can be achieved by following the pipelines (Fig. 4).

- (1) Spilt the coarse segmentation result of an ROI image into sub-images according to the number of connected components.
- (2) For each sub-image, enlarge the image size by padding the edges with extra zero pixels. Use a sliding window and the FCN-8s net to perform pixel-wise classification, obtaining the optimized result of the sub-image.
- (3) Update the coarse segmentation result with the optimized sub-

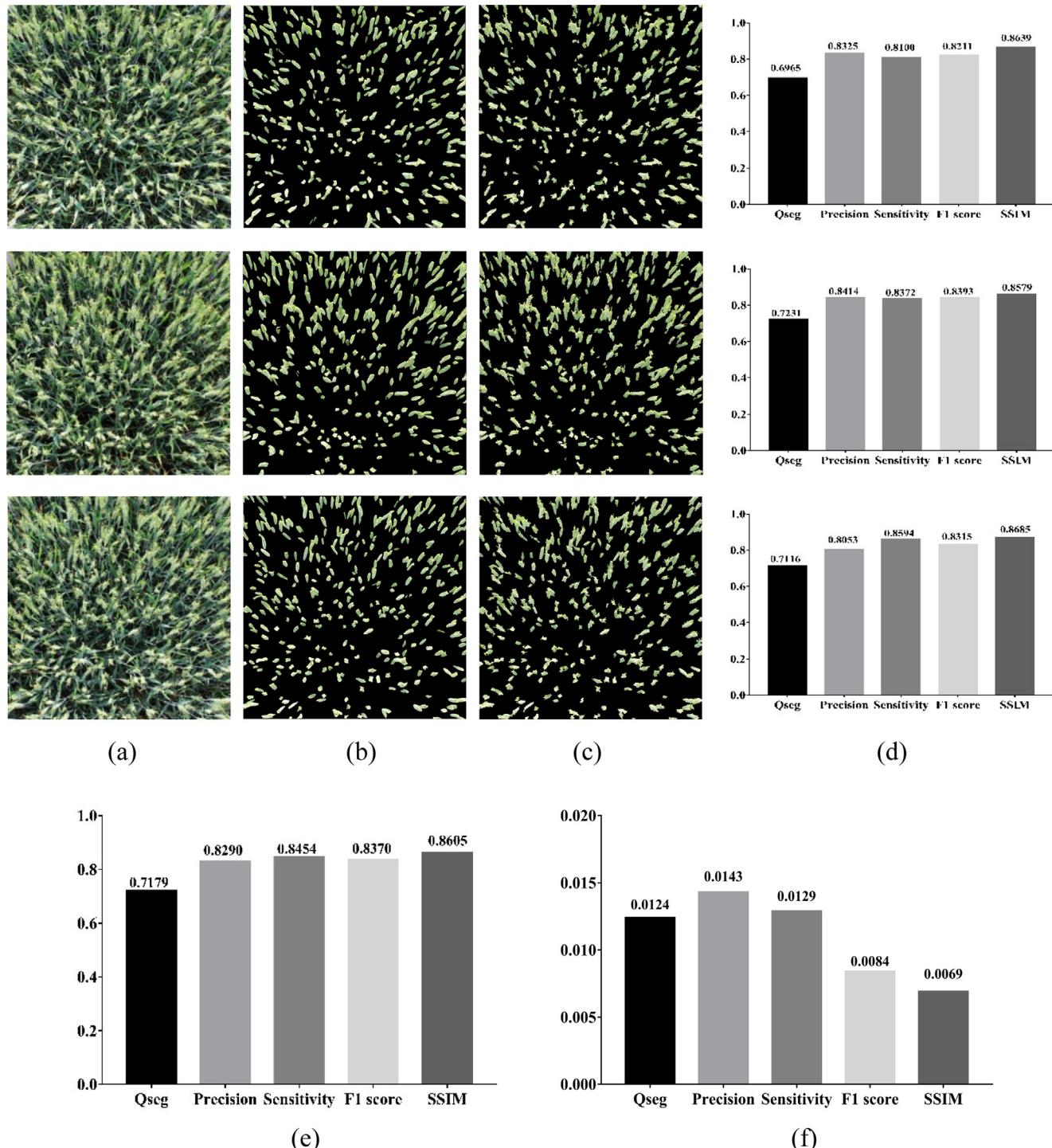


Fig. 6. Fine segmentation results, (a) Original images, (b) Ground truth images, (c) Fine segmentation results, (d) evaluation metrics, (e) Mean values of the evaluation metrics for fine segmentation, (f) Standard deviations of the evaluation metrics for fine segmentation.

Table 4

The details of the down-sampled image datasets.

Dataset	Down sampling factor	Ground resolution	Range of the NOSPs (interval: 500)
Dataset for 2018	/	0.16 mm	[500,4000]
Down-sampled dataset_1	0.7	0.33 mm	[500,3500]
Down-sampled dataset_2	0.5	0.64 mm	[500,3000]
Down-sampled dataset_3	0.4	1 mm	[500,2000]

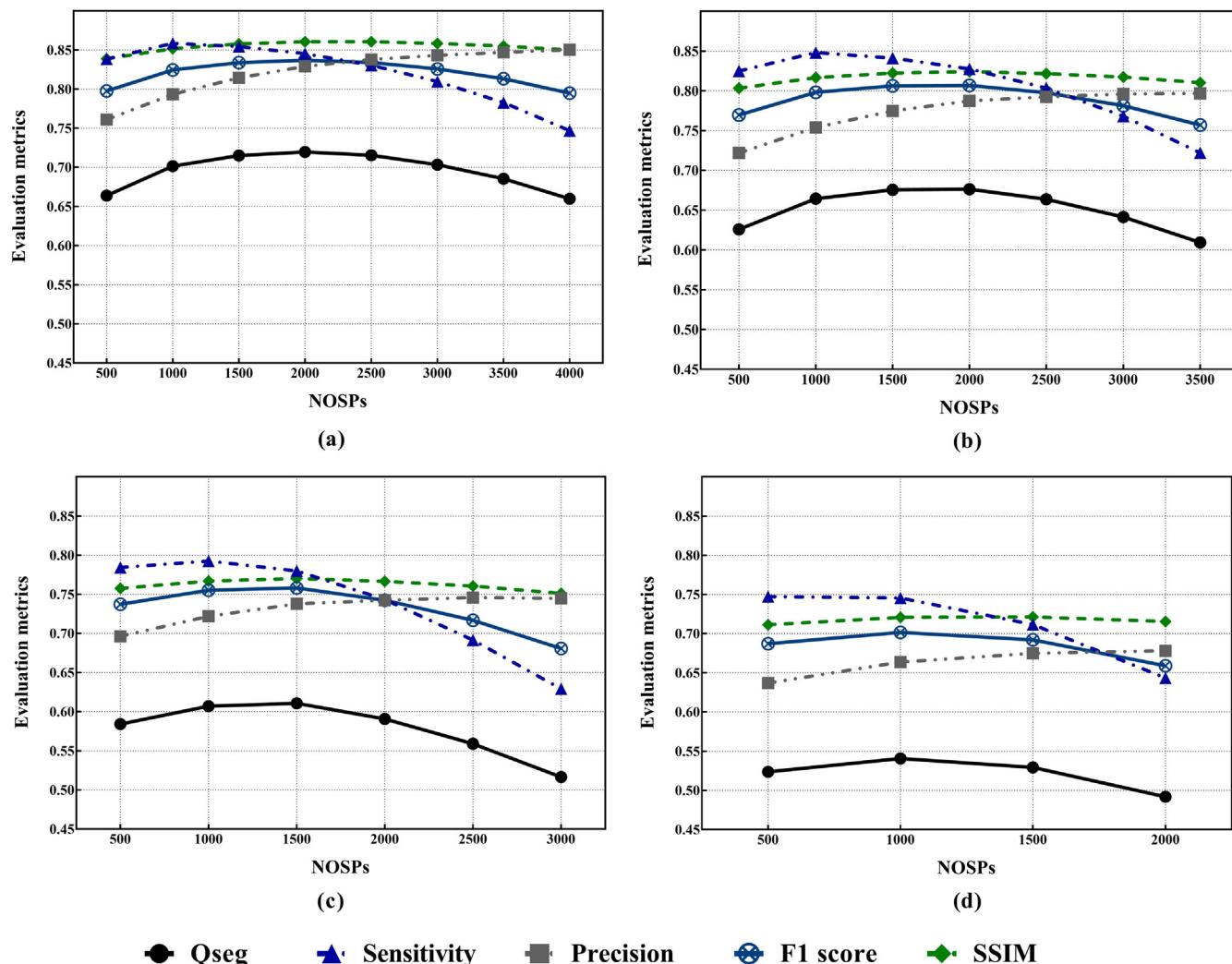


Fig. 7. Performances of the two-stage segmentation method over the down-sampled datasets, (a) dataset for 2018 with ground resolution of 0.16 mm (no down sampling), (b) Down-sampled dataset_1 with ground resolution of 0.33 mm, (c) Down-sampled dataset_2 with ground resolution of 0.64 mm, (d) Down-sampled dataset_3 with ground resolution of 1 mm.

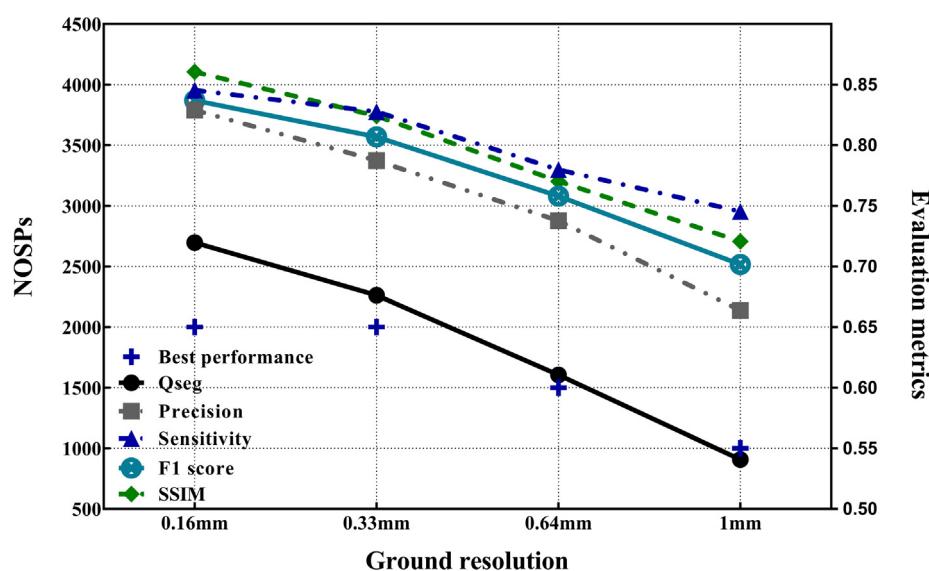


Fig. 8. NOSPs adopted in the best performances of the two-stage segmentation method over the down-sampled datasets (left Y axis), the right Y axis showed the evaluation metrics of the corresponding best performance.



Fig. 9. Some of the misclassified superpixels by the DCNN in Down-sampled dataset_3, the NOSPs was 1000, the bottom row was the size of the corresponding superpixel.

images.

2.4. Performance evaluation

In order to show the validity of the two-stage segmentation method, as well as to understand the limitations, comparative tests and generalization tests were performed. For the comparative tests, the two-stage segmentation method was compared to other widely used segmentation methods. The K-means clustering algorithm was selected as one of the compared methods for its ability to perform automatic segmentation (Ma et al., 2017). For the purpose of segmentation using automatic clustering, the ROI images were converted to CIELab color space (Ma et al., 2017; García-Mateos et al., 2015). The ear counting algorithm proposed by Fernandez-Gallego et al. (2018) was adopted as one of the compared methods. Since the pixel resolution of the images was different, the original window size used in the ear counting algorithm was not applicable to canopy images in this study. Based on the test results, the window size was finally determined as 32. Another compared method was random forest (Breiman, 2001), which was constructed using the color features of the superpixels. The color features included the minimum, maximum, mean and standard deviation values of each channel of RGB, HSV and CIELab color spaces. As mentioned in 2.3, the FCN-8s net was able to perform pixel-wise classification, making it a feasible method to segment ears of winter wheat from digital images. Therefore, in the comparative tests, the FCN-8s net which was constructed for the fine segmentation was applied directly to the ROI images. Given that the input image size for the FCN-8s net was 400×400 pixels, the ROI images were divided into sub-images using a sliding window. No overlap between the sub-images was necessary since the FCN-8s net had the advantage of pixel-wise classification. For the generalization tests, the two-stage segmentation method was generalized to image datasets with new variability. In order to quantitatively evaluate the performance of the segmentation methods, five widely used evaluation metrics, i.e. Qseg, Precision, Sensitivity, F1 score, and structural similarity index (SSIM), were adopted in this study (Zhou et al., 2018; Ma et al., 2018, 2017; Ferreira et al., 2017; Xiong et al., 2017). For the four evaluation metrics, i.e. Qseg, Precision, Sensitivity and F1 score, confusion matrices were involved for the calculation (Powers, 2007). For the SSIM, the method proposed by Wang et al. (2004) was used.

$$Q\text{seg} = \frac{TP}{TP + FP + FN} \quad (1)$$

Table 5
The statistics of the evaluation metrics for comparative tests (mean value \pm standard deviation).

Methods	Qseg	Precision (%)	Sensitivity (%)	F1 score (%)	SSIM
Proposed method	0.7197 \pm 0.0124	82.9 \pm 1.43	84.54 \pm 1.29	83.7 \pm 0.84	0.8605 \pm 0.0069
Auto-clustering	0.3594 \pm 0.0222	36.56 \pm 2.32	95.49 \pm 0.79	52.83 \pm 2.42	0.4684 \pm 0.0212
Ear counting algorithm	0.3650 \pm 0.0269	69.36 \pm 2.88	43.68 \pm 4.12	53.42 \pm 2.87	0.7848 \pm 0.0136
Random forest	0.4097 \pm 0.0222	44.84 \pm 2.39	82.81 \pm 4.72	58.09 \pm 2.25	0.6471 \pm 0.0223
FCN-8s net	0.6917 \pm 0.0124	72.32 \pm 1.49	94.09 \pm 0.78	81.77 \pm 0.87	0.8326 \pm 0.0075

Bold values indicated the best performance by the methods.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (3)$$

$$F1 \text{ score} = 2 * \frac{\text{Precision} * \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (4)$$

where TP was the number of ear pixels that were correctly classified as lass ear and FP was the number of background pixels that were incorrectly classified as class ear. FN was the number of ear pixels that were incorrectly classified as class background and FP was the number of background pixels that were correctly classified as class background.

3. Results and discussion

3.1. Construction of the two-stage segmentation method

3.1.1. Coarse segmentation

When training finished, the test dataset was used to test the DCNN. The test results were shown in Table 2.

It can be seen from the confusion matrix that 92.0% of the ear predictions were correct, as well as 94.9% of the ear cases, revealing that the DCNN achieved an F1 score of 93.4% for class ear. Similar results can be observed for class non-ear. The results showed that DCNN achieved an accuracy of 96.4%, indicating that the DCNN was capable of accurately distinguishing the two classes. In the following step, the pretrained DCNN was used to classify the superpixels generated from the ROI image into the two classes, achieving the coarse segmentation result for the corresponding ROI image. Some coarse segmentation results were shown in Fig. 5.

As expected, non-ear pixels can be observed in the coarse segmentation results (Fig. 5c, the red pixels), which agreed with the corresponding results of evaluation metrics presented in Fig. 5d. Quantitative evaluation of the coarse segmentation results was performed using the evaluation metrics mentioned in 2.4 (Fig. 5e, and f). The results showed that the coarse segmentation did not perform well on the metrics of Qseg, Precision, F1 score, and SSIM, with the mean values of 0.5837, 63.55%, 73.70%, and 0.7742, respectively, however, it achieved a high score of sensitivity, with a mean value of 87.76% (Fig. 5e). The detailed results of coarse segmentation on the dataset for 2018 were shown in Appendix A, Table S1. Overall, the coarse segmentation was far from a good segmentation, requiring the fine segmentation to further optimize the results. It was revealed form the coarse segmentation results that the canopy images captured in field conditions had a lot of noise that the brightness was not robust enough to accurately distinguish ear from the background. Although the entropy rate superpixels algorithm which had an advantage of keeping the intrinsic homogeneity was used to generate superpixels for the ROI images, it was unavoidable that there were noise pixels in the results. Nevertheless, what was significant about the coarse segmentation was that most of the ear pixels in ROI images, i.e., nearly 90%, were correctly segmented so that the fine segmentation could just focus on eliminating the non-ear pixels in the coarse segmentation results.

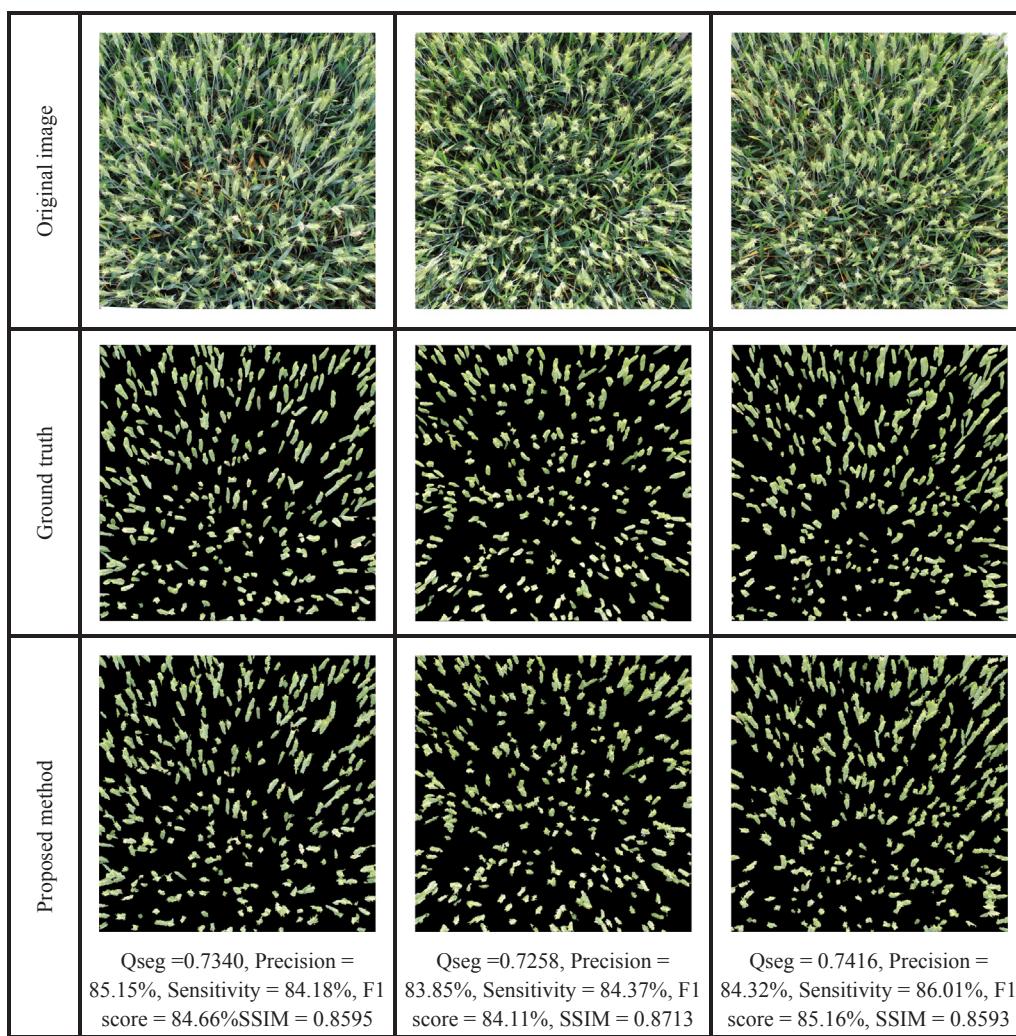


Fig. 10. Examples of segmentation results.

3.1.2. Fine segmentation

In order to test the performance of the FCN-8s net for fine segmentation, the confusion matrix was used. The results were shown in Table 3.

It can be seen from the confusion matrix that the FCN-8s net achieved an overall accuracy of 92.15%. In the performance per class, the performance of the FCN-8s net on class non-ear was better than that on class ear. The precision values for each class were 84.92%, and 94.96% respectively, and the sensitivity values for each class were 86.75% and 94.19%. The test results showed that the FCN-8s net was able to eliminate the noise pixels. The fine segmentation results were shown in Fig. 6.

Significant improvement in the segmentation results can be observed (Fig. 6c). Quantitative evaluation of the fine segmentation results was also performed using the evaluation metrics (Fig. 6d, e). Comparing the two results, it can be seen that the fine segmentation achieved improved results on the metrics of Qseg, Precision, F1 score, and SSIM, with mean values of 0.7197, 82.90%, 83.70%, and 0.8605 respectively, while the result on the sensitivity slightly decreased, with a mean value of 84.54% (Fig. 6e). The detailed results of fine segmentation on the dataset for 2018 were shown in Appendix A, Table S2. The observed improvement in the segmentation results could be attributed to the good performance of FCN-8s net. With the ability to perform pixel-wise classification, FCN-8s net optimized the coarse segmentation results by correcting the classification results of the non-ear pixels. For the results of the FCN-8s net over the test dataset, it was

not necessary to be concerned about the fact that performance on class background was better than that on class ear. Previous studies indicated that deep learning techniques can achieve satisfactory results with a large mass of data (Ferreira et al., 2017; Ma et al., 2019, 2018). It was obvious that the data size of the two classes in the training set for FCN-8s net was severely unbalanced. The data size of class ear was heavily outnumbered by that of class background, thus making it expectable that the FCN-8s net might achieve better performance on class background. The results in this section indicated that the proposed two-stage wheat ear segmentation method achieved accurate segmentation results of the canopy images captured at flowering stage of winter wheat. In the following section, the factors that might influence the performance of the proposed two-stage segmentation method were discussed.

3.1.3. Preferable NOSPs for canopy images of different ground resolutions

The number of superpixels (NOSPs) was an important hyperparameter, determining the size of the superpixels. Taking a large number resulted in small-sized superpixels that contained less descriptive information for the DCNN, increasing the chances of misclassification. In the opposite case, the superpixels might contain noise pixels degrading the accuracy of FCN-8s net. In field use of the two-stage segmentation method, multiple devices might be involved, resulting in canopy images of different ground resolutions. It was therefore necessary to determine the suitable NOSPs for canopy images of different ground resolutions. To understand the influence of NOSPs and ground resolution of canopy images to the performance of the two-stage segmentation method, as

Auto clustering	Qseg = 0.3884, Precision = 39.60%, Sensitivity = 95.29%, F1 score = 55.95%, SSIM = 0.4879	Qseg = 0.3456, Precision = 35.24%, Sensitivity = 95.36%, F1 score = 51.47%, SSIM = 0.4677	Qseg = 0.3808, Precision = 38.72%, Sensitivity = 95.84%, F1 score = 55.16%, SSIM = 0.4658
Ear counting algorithm			
	Qseg = 0.3542, Precision = 74.50%, Sensitivity = 40.31%, F1 score = 52.31%, SSIM = 0.7789	Qseg = 0.3774, Precision = 72.57%, Sensitivity = 44.02%, F1 score = 54.80%, SSIM = 0.8037	Qseg = 0.3650, Precision = 69.10%, Sensitivity = 43.62%, F1 score = 53.48%, SSIM = 0.7704
FCN-8s net			
	Qseg = 0.7067, Precision = 74.13%, Sensitivity = 93.80%, F1 score = 82.18%, SSIM = 0.8302	Qseg = 0.6918, Precision = 72.50%, Sensitivity = 93.80%, F1 score = 81.79%, SSIM = 0.8411	Qseg = 0.7114 Precision = 74.25%, Sensitivity = 94.45%, F1 score = 83.14%, SSIM = 0.8303
Random forest			
	Qseg = 0.4374, Precision = 47.86%, Sensitivity = 83.56%, F1 score = 60.86%, SSIM = 0.6535	Qseg = 0.4112, Precision = 44.20%, Sensitivity = 85.51%, F1 score = 58.27%, SSIM = 0.6578	Qseg = 0.4415, Precision = 48.25%, Sensitivity = 83.87%, F1 score = 61.26%, SSIM = 0.6532

Fig. 10. (continued)

well as to determine the preferable NOSPs for canopy images of different ground resolutions, grid search were conducted.

To simulate canopy images of different ground resolutions, the image dataset for 2018 was down sampled by factors of 0.7, 0.5, and 0.4 using bicubic interpolation. The down-sampled image datasets had ground resolutions of 0.33 mm, 0.64 mm and 1 mm respectively.

Details of the down-sampled image datasets can be found in Table 4. The performances of the two-stage segmentation method tested over the down-sampled datasets were shown in Fig. 7. The detailed results on the down-sampled image datasets were shown in Appendix B, Table S3-S6.

It can be seen from the results that the two-stages segmentation

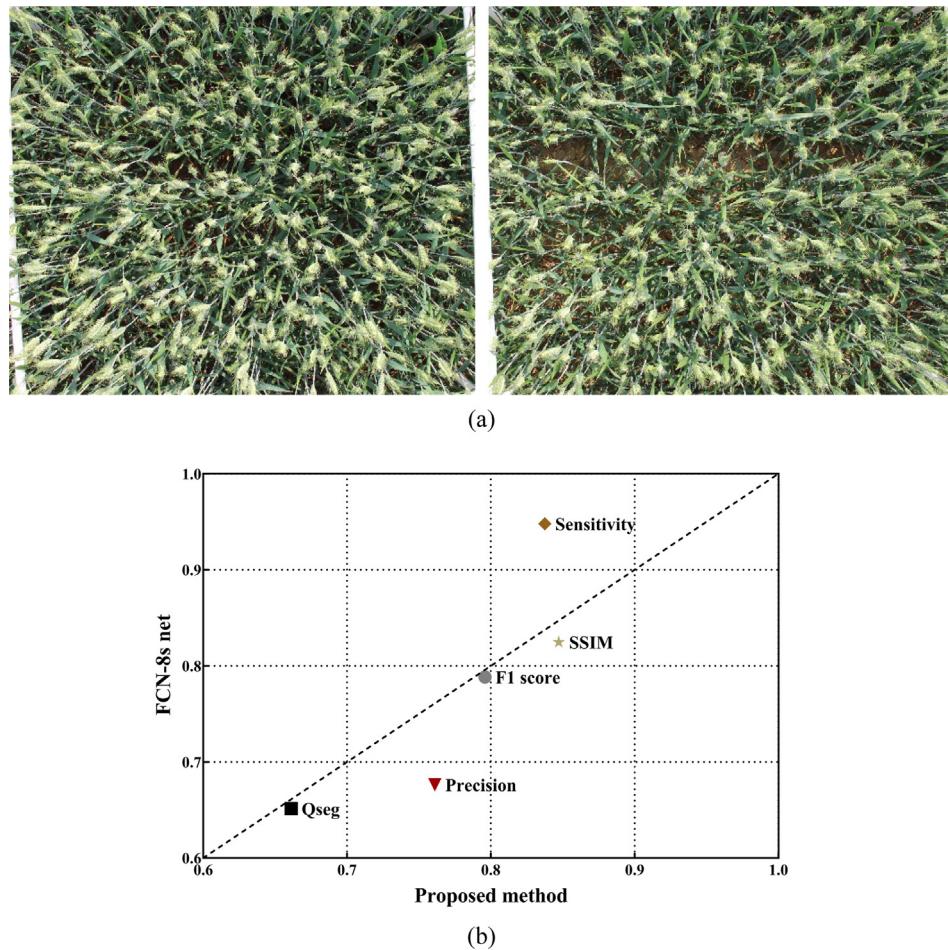


Fig. 11. The comparison results of the two methods on the dataset for season 2019.

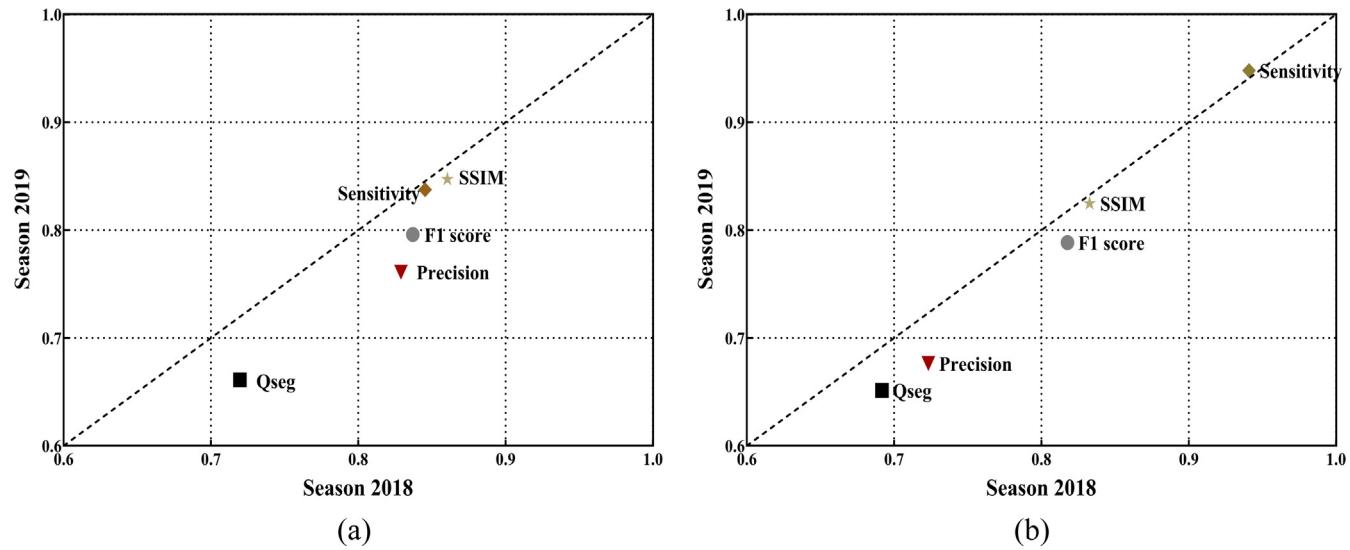


Fig. 12. The performances of the two methods on the dataset for the two seasons, (a) the proposed two-stage segmentation method, (b) FCN-8s net.

method achieved the best performance over the dataset for 2018 when adopting the NOSPs of 2000. Similarly, the best performances over the down-sampled datasets were achieved using the NOSPs of 2000, 1500, and 1000 respectively. The results derived from the grid search (Fig. 7) showed that the NOSPs had a second-order effect on the performances of the two-stage segmentation method over the down-sampled datasets, thus making it possible to confirm the most suitable NOSPs for canopy

images of ground resolution ranging from 0.16 mm to 1 mm for future reference (Fig. 8).

It also can be seen from Fig. 8 that the performance of the two-stage segmentation degraded with the decrease of the ground resolution. This degradation in the performances was mostly explained by the facts that the superpixels generated from canopy images of relatively low ground resolution contained less descriptive information than those generated

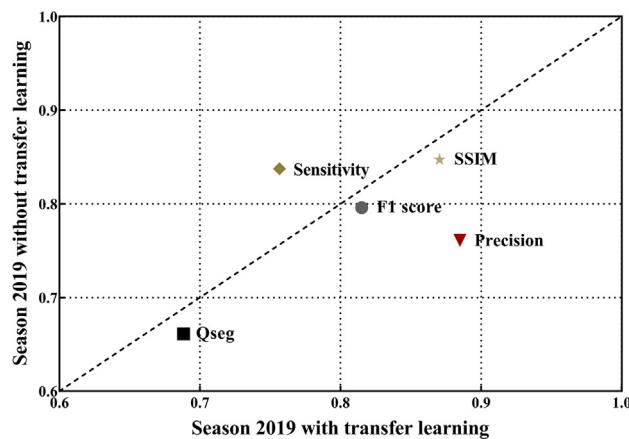


Fig. 13. The performance of the two-stage method over the dataset for season 2019 by transfer learning.

from canopy images of relatively high ground resolution for the DCNN. This can be seen in Fig. 9. The size of the misclassified superpixels by the DCNN was small, moreover, the percentage of ear pixels was low in these small-size superpixels.

3.2. Comparative tests with widely adopted segmentation methods

In this section, the results of the comparative tests were presented. The statistics of the evaluation metrics for the comparative tests were shown in Table 5. Some of the segmentation results were shown in Fig. 10. It can be seen that the proposed two-stage segmentation method demonstrated superior results to the compared methods in the evaluation metrics of Qseg, Precision, F1 score, and SSIM. The proposed method had the four evaluation metrics equal to 0.7197, 82.90%,

83.70%, and 0.8605 respectively. As for the evaluation metric of sensitivity, the auto-clustering algorithm outperformed the other segmentation methods, with a mean value of 95.49%. The sensitivity value of the proposed two-stage segmentation method was 84.54%. For the detailed results of the compared methods on the dataset for season 2018, readers were referred to Appendix C, Table S7a and S7b.

The results revealed that both of the proposed two-stage segmentation method and the FCN-8s net were able to correctly segment the winter wheat ears, moreover, by comparison, the proposed two-stage segmentation method performed better than the FCN-8s net. The reason that the proposed two-stage segmentation method outperformed the FCN-8s net was rooted in the coarse segmentation stage, which protected the fine segmentation stage from being exposed to the variability that was not covered in the training dataset. A large amount of non-ear pixels, i.e., leaves and stems, can be observed in the segmentation results of the auto-clustering algorithm and the random forest method, indicating that the method was severely affected by the clutter background (Fig. 10). A possible explanation for the poor performances of the two methods might be that they were counting on the color information to discriminate ears from background in ROI images. As mentioned in Section 3.1, the canopy images captured under field conditions had a lot of noise, which had severe influences on the color. Besides, from the perspective of color, the ears of winter wheat at flowering stage were very similar to leaves. In the case of the ear counting algorithm, the above explanation also applied, moreover, it was using a fixed value for the parameters, which cannot cover all the variability of different image datasets. In summary, the results of the comparative tests suggested the proposed two-stage segmentation method was a reliable method for canopy images captured at flowering stage of winter wheat.

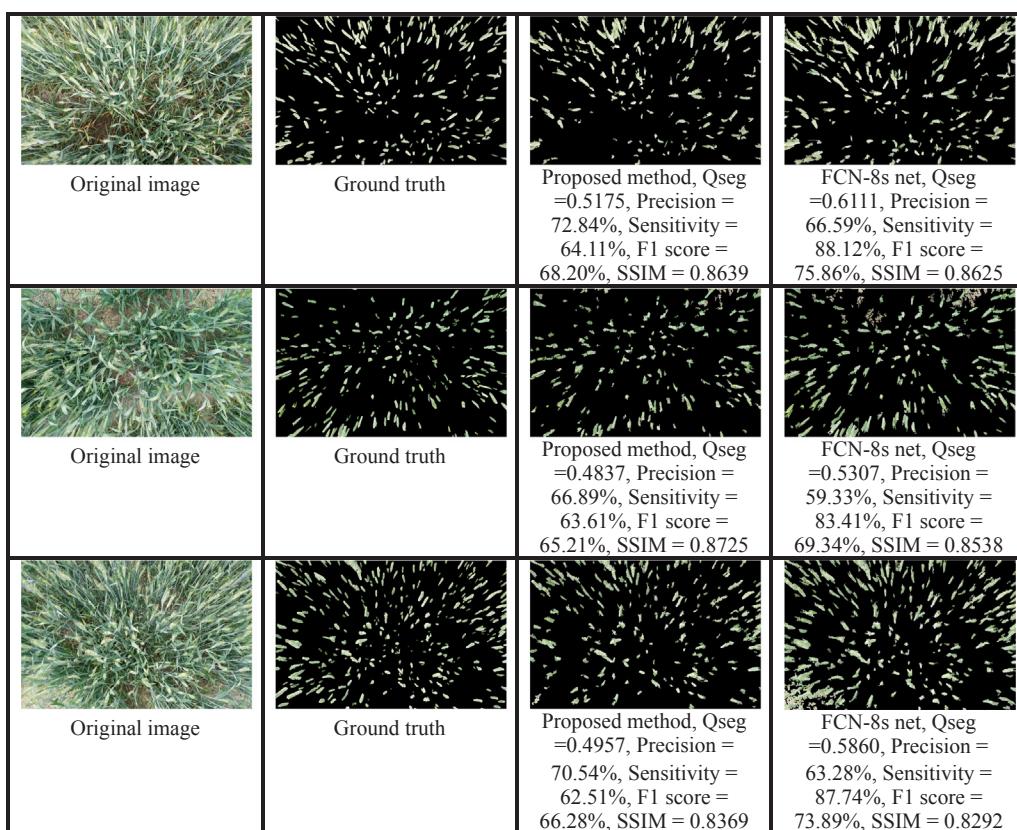


Fig. 14. The performances of the two methods over the low ground resolution images.

			
Ears: 0.0418 Background: 0.9582	Ears: 0.2179 Background: 0.7821	Ears: 0.3282 Background: 0.6718	Ears: 0.4283 Background: 0.5717

3.3. Generalization tests using image datasets with new variabilities

In order to validate the utility of the two-stage segmentation method, experiments of the proposed method on a dataset for the season 2019 were conducted. The dataset for the season 2019 containing 36 ROI images was constructed by following the same protocols of the dataset for season 2018. Since the FCN-8s net demonstrated great potential for segmenting ears of winter wheat, it was adopted as the compared method. It was worth noting that the growth of winter wheat in season 2019 was affected by the frost damage, adding new variability to the images in the dataset, such as different ear sizes and extra soil pixels due to a less dense canopy (Fig. 11a). The comparison results of the two methods over the dataset for season 2019 were shown in Fig. 11b. The NOSPs for the two-stage segmentation method was 2000.

The results showed that there were four evaluation metrics below the 1:1 line, i.e., Qseg, Precision, F1 score, and SSIM, indicating that the proposed two-stage segmentation method achieved superior results to the FCN-8s net on the four evaluation metrics. The proposed method had the four evaluation metrics equal to 0.6612, 76.10%, 79.59%, and 0.8472 respectively, and the FCN-8s net had the evaluation metrics equal to 0.6514, 67.64%, 78.84%, and 0.8247 respectively. The sensitivity was above the 1:1 line, revealing that the FCN-8s net outperformed the proposed two-stage segmentation method on this evaluation metric. The proposed method had the sensitivity equal to 83.75%, and the FCN-8s net had the sensitivity equal to 94.80%. The detailed results on the dataset for season 2019 were shown in Appendix D, Table S8. It can be seen that the comparison results that the performance of the two-stage segmentation method over the dataset for season 2019 was superior to that of the FCN-8s net agreed with that on the dataset for season 2018. The performances of the two methods on the dataset for the two seasons were shown in Fig. 12.

For the evaluation results of the proposed two-stage segmentation method (Fig. 12a), it can be observed that the sensitivity was very close to the 1:1 line while the other four evaluation metrics of Qseg, Precision, F1 score, and SSIM were below the 1:1 line, indicating that the method maintained its ability to correctly segment all the ear pixels. Compared to the performance on the dataset for season 2018, the proposed two-stage method made more errors of predicting non-ear pixels as ear pixels. Although the performance of the proposed two-stage segmentation method on the dataset for season 2019 slightly decreased (Qseg = 0.6612, F1 score = 79.59%, SSIM = 0.8472), it was still capable of segmenting winter wheat ears from canopy images captured at flowering stage. The similar results can be observed on the evaluation results of the FCN-8s net (Fig. 12b). In order to improve the performance the proposed two-stage method over the dataset for season 2019, transfer learning was adopted. The training, validation and test sets were constructed by following the same protocols for the dataset for season 2018. The results were shown in Fig. 13. The detailed results on the dataset for season 2019 by transfer learning were shown in Appendix D, Table S9.

It can be seen from the results that there were four evaluation metrics of Qseg, Precision, F1 score, and SSIM were below the 1:1 line, indicating that the method was able to segment the ear pixels with more precision. The evaluation metric of sensitivity was above the 1:1 line, indicating that some of the ear pixels were misclassified as background during the process of optimization. In general, transfer learning did

Fig. 15. Some of the superpixels misclassified by the DCNN, the bottom row demonstrated the predicted scores of the corresponding superpixels.

improve the performance of the proposed two-stage method over the dataset for season 2019 ($\text{Qseg} = 0.6884$, $\text{F1 score} = 81.51\%$, $\text{SSIM} = 0.8704$), enabling it to segment ears from the datasets for both season 2018 and season 2019.

To further assess the generalization ability of the two-stage segmentation method, tests were performed using canopy image of coarser resolution. The images were captured at the experimental station of Tianjin climate center, Tianjin, China. The winter wheat cultivar was Jimai 22. The device for image capture was a smart phone (MEIZU M5 note), which was at a height of 1 m and oriented vertically downwards. The pixel resolution of the images was 1440×2560 , resulting in a ground resolution of 0.5 mm. Nine canopy images were adopted to test the performance of the two methods, and some of the results were shown in Fig. 14 According to Fig. 8, NOSPs of 2000 was adopted for the two-stage segmentation method. The detailed results on the nine canopy images were shown in Appendix D, Table S10

Results showed decreased performances for both methods as evaluated over the nine canopy images (Table 5). This result was expected since the images presented new variability, such as relatively low ground resolution and different winter wheat cultivars. It was believed that the decreased performances for the two methods were caused by different reasons. For the FCN-8s net, it can be seen from Fig. 14 that there were noise pixels that misclassified as class ear while the FCN-8s net achieved a mean sensitivity value of 86.34%. It was therefore concluded that the decreased performance for FCN-8s net was caused by the pixels which were not covered in the training set. For the case of the two-stage method, it can be seen in Fig. 14 that some of the ear pixels were misclassified as class background, indicating that the DCNN in the coarse segmentation failed to correctly classify these superpixels. Fig. 15 demonstrated some of the superpixels misclassified by the DCNN.

It was revealed that the winter wheat cultivar had an influence on the two-stage segmentation method since the superpixels in the figure were containing few noise pixels. However, due to the fact that the two-stage segmentation method correctly classified over 65% ear pixels (Appendix D, Table S10), the reason for the misclassifications was not a single factor, but integrating other factors. According to the predicted scores of the misclassified superpixels (Fig. 15, bottom row), one can see that the DCNN performed better on the superpixels with rich details of ear, indicating the lack of descriptive information caused by low ground resolution was the ‘other factor’. Therefore, it was concluded that both the change of winter wheat cultivar and the lack of descriptive information contributed to the misclassification of the two-stage segmentation method. Nevertheless, the results laid the fact that the proposed two-stage segmentation method had good generalization ability, which was a robust tool for segmenting ears of winter wheat from canopy images captured at the flowering stage.

3.4. FCN-8s net demonstrated good potential for UAV-based canopy images

In order to test the potentials of the method for segmenting ears of winter wheat from UAV-based RGB images, experiments were conducted at the experimental station of the Institute of Environment and Sustainable Development in Agriculture, Chinese Academy of Agricultural Sciences located in Shunyi, Beijing, China. The winter wheat cultivar was Zhongmai 1062. The canopy images of winter wheat

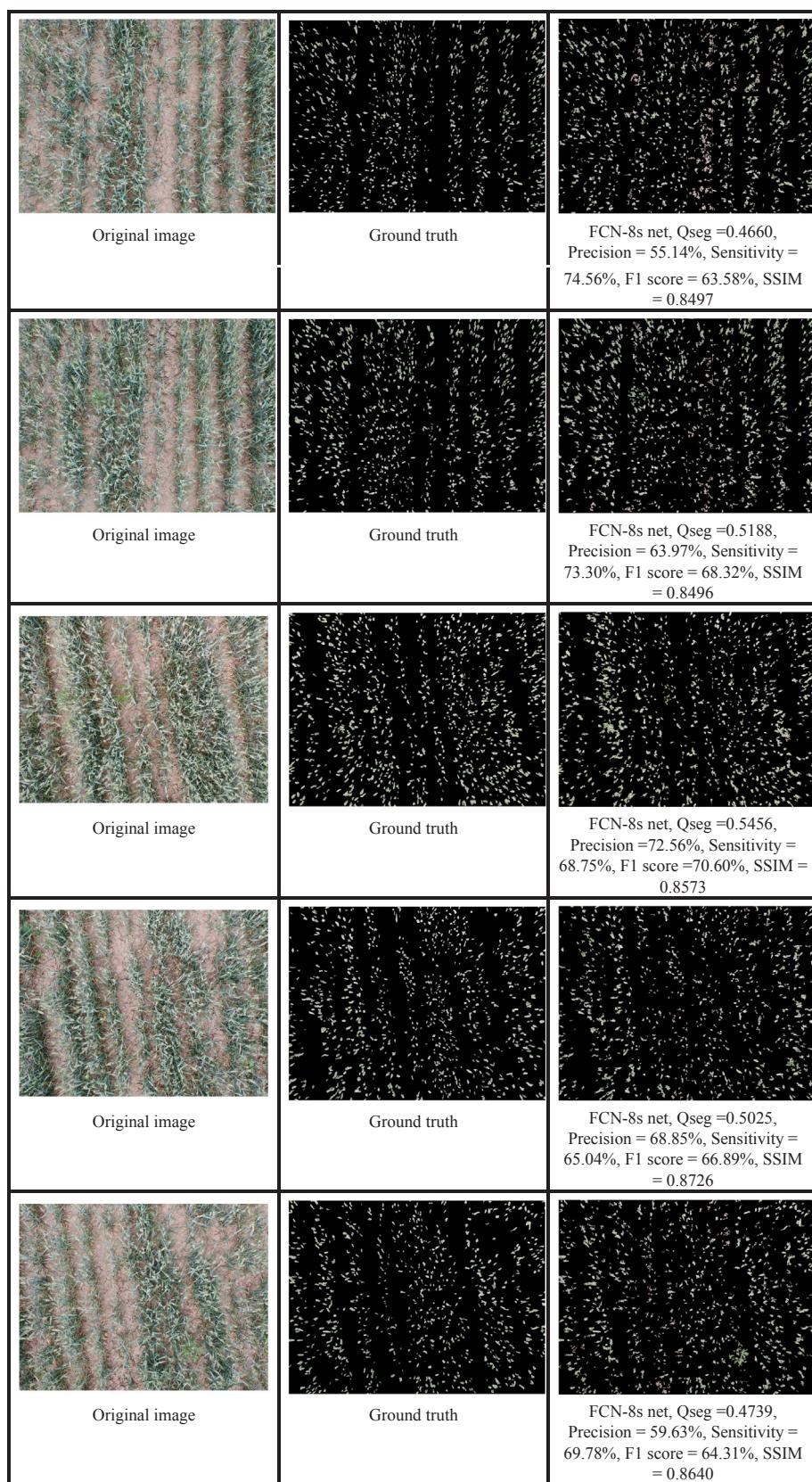


Fig. 16. The performance of the FCN-8s net over the UAV-based canopy images.

were captured by an Unmanned Aerial Vehicle (UAV) on May 17, 2019, which was the early flowering stage of winter wheat. The UAV for image collection was the DJI Mavic 2 pro (SZ DJI Technology Co., Shenzhen, China), which was a consumer-level drone with a digital camera. The images which were captured from the nadir view direction at the height of 3 m had a pixel resolution of 4000×3000 , resulting in a ground resolution of 0.18 mm. The two methods were tested over five UAV-based canopy images, and the results were shown in Fig. 16.

As expected, it can be observed from Fig. 16 that there were false positive pixels in the segmentation results. The performance of the FCN-8s net decreased as compared to that over the ground-based canopy images ($Q_{seg} = 0.5014$, F1 score = 66.74%, SSIM = 0.8586). The detailed results on the UAV-based canopy images were shown in Appendix D, Table S11. It can be seen from the above figure that the ear density was very low and there were weeds in the field, thus resulting in a huge amount of noise pixels which was not covered in the training set of the FCN-8s net. This may explain the decreased performance of the FCN-8s net. In the case of the two-stage segmentation method, different numbers of superpixels in the coarse segmentation, i.e., 2000, 3500, and 5000, were tested to adapt to the pixel resolution of the UAV-based RGB images. However, the DCNN was not able to distinguish the ears from background no matter which number was adopted, thus making the fine segmentation impossible to achieve valid segmentation results. Although the ground resolution of the UAV-based canopy images was close to that of the ground-based images in datasets for season 2018, the ear density was very low and the size of the ears was relatively small, which was just the scene that would invalidate the two-stage segmentation method based on the analysis in Section 3.3. This may explain the failure of the two-stage segmentation method. On the other hand, the FCN-8s net that trained with limited number of images achieved promising segmentation, which highlighted its potential for high throughput applications. It can be expected that the FCN-8s net will achieve better results with considerable number of training images.

4. Conclusion

A two-stage segmentation method for ears of winter wheat based on digital images and deep learning techniques was proposed in this paper. The results showed that the proposed two-stage segmentation method was able to achieve accurate segmentation of winter wheat ears from canopy images captured at flowering stage ($Q_{seg} = 0.7197$, F1 score = 83.70%, SSIM = 0.8605). The performance of the proposed segmentation method was compared to the widely adopted segmentation methods. Results showed that the proposed two-stage segmentation method outperformed the compared methods, making a reliable tool to segmentation of winter wheat ears from canopy images captured at the flowering stage. The results have also shown that the FCN-8s net demonstrated great potentials for segmentation of winter wheat ears ($Q_{seg} = 0.6917$, F1 score = 81.77%, SSIM = 0.8326). In order to validate the utility of the proposed two-stage segmentation method, generalization tests were constructed. Results showed the two-stage segmentation method was still capable of accurately segmenting ears of winter wheat, even though the performance slightly decreased. Change of winter wheat cultivar and lack of descriptive information were two factors that could degrade the performance of the two-stage segmentation method. Tests of the methods for segmenting ears of winter wheat from UAV-based RGB images showed the FCN-8s had a good chance to achieve satisfactory performances for UAV-based canopy images.

Although the FCN-8s net demonstrated great potentials for segmenting the winter wheat ears, it is very necessary to enlarge the size of the training set, thus covering as much variability expected to occur in practice as possible. In the following research, considerably more work will need to be done to increase the number of images in the training set for the FCN-8s net. Moreover, canopy images captured at other stages will also be included in the dataset so that the method can extend its

application scope.

CRediT authorship contribution statement

Juncheng Ma: Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing, Funding acquisition. **Yunxia Li:** Software, Formal analysis, Data curation. **Keming Du:** Validation, Project administration, Funding acquisition. **Feixiang Zheng:** Validation, Writing - review & editing. **Lingxian Zhang:** Conceptualization, Methodology, Writing - review & editing, Supervision. **Zhihong Gong:** Validation, Resources. **Weihua Jiao:** Validation, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The Authors wish to acknowledge The National Natural Science Foundation of China (31801264), Young Elite Scientists Sponsorship Program by CAST (2018QNRC001) and The National Key Research and Development Program of China (2016YFD0300606) for their funding support of this research.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compag.2019.105159>.

References

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* 34, 2274–2281. <https://doi.org/10.1109/TPAMI.2012.120>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32. https://doi.org/10.1007/978-3-662-56776-0_10.
- Duan, L., Huang, C., Chen, G., Xiong, L., Liu, Q., Yang, W., 2015. Determination of rice panicle numbers during heading by multi-angle imaging. *Crop J.* 3, 211–219. <https://doi.org/10.1016/j.cj.2015.03.002>.
- Fernandez-Gallego, J.A., Kefauver, S.C., Gutiérrez, N.A., Teresa, M., Taladriz, N., Araus, J.L., 2018. Wheat ear counting in - field conditions : high throughput and low - cost approach using RGB images. *Plant Methods* 1–12. <https://doi.org/10.1186/s13007-018-0289-4>.
- Ferreira, A. dos S., Freitas, D.M., Silva, G.G. da, Pistori, H., Folhes, M.T., 2017. Weed detection in soybean crops using ConvNets. *Comput. Electron. Agric.* 143, 314–324. <https://doi.org/10.1016/j.compag.2017.10.027>.
- García-Mateos, G., Hernández-Hernández, J.L., Escarabajal-Henarejos, D., Jaén-Terrones, S., Molina-Martínez, J.M., 2015. Study and comparison of color models for automatic image analysis in irrigation management applications. *Agric. Water Manag.* 151, 158–166. <https://doi.org/10.1016/j.agwat.2014.08.010>.
- Ghosal, S., Blystone, D., Singh, A.K., Ganapathysubramanian, B., Singh, A., 2018. An explainable deep machine vision framework for plant stress phenotyping. *Proc. Natl. Acad. Sci. U. S. A.* 115, 4613–4618. <https://doi.org/10.1073/pnas.1716999115>.
- Hamuda, E., Mc Ginley, B., Glavin, M., Jones, E., 2017. Automatic crop detection under field conditions using the HSV colour space and morphological operations. *Comput. Electron. Agric.* 133, 97–107. <https://doi.org/10.1016/j.compag.2016.11.021>.
- Hu, Q. xia, Tian, J., He, dian, 2017. Wheat leaf lesion color image segmentation with improved multichannel selection based on the Chan-Vese model. *Comput. Electron. Agric.* 135, 260–268. <https://doi.org/10.1016/j.compag.2017.01.016>.
- Liu, M., Tuzel, O., Ramalingam, S., Chellappa, R., 2011. Entropy rate superpixel segmentation. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, 2097–2104. <https://doi.org/10.1109/CVPR.2011.5995323>.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 07–12 June, 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>.
- Lu, H., Cao, Z., Xiao, Y., Zhuang, B., Shen, C., 2017. TasselNet: Counting maize tassels in the wild via local counts regression network. *Plant Methods* 13, 1–17. <https://doi.org/10.1186/s13007-017-0224-0>.
- Ma, J., Du, K., Zhang, L., Zheng, F., Chu, J., Sun, Z., 2017. A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing. *Comput. Electron. Agric.* 142, 110–117. <https://doi.org/10.1016/j.compag.2017.08.023>.

- Ma, J., Du, K., Zheng, F., Zhang, L., Gong, Z., Sun, Z., 2018. A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network. *Comput. Electron. Agric.* 154, 18–24. <https://doi.org/10.1016/j.compag.2018.08.048>.
- Ma, J., Li, Y., Chen, Y., Du, K., Zheng, F., Zhang, L., Sun, Z., 2019. Estimating above ground biomass of winter wheat at early growth stages using digital images and deep convolutional neural network. *Eur. J. Agron.* 103, 117–129. <https://doi.org/10.1016/j.eja.2018.12.004>.
- Madec, S., Frederic, B., Liu, S., De Solan, B., Xiliang, J., Lu, H., Duyume, F., Baret, F., 2018. Ear density estimation from high resolution RGB imagery using deep learning technique. *Agric. For. Meteorol.* 264, 225–234 [https://doi.org/10.1016/j.agrimeet.2018.09.037X](https://doi.org/10.1016/j.agrimeet.2018.09.037).
- Powers, D.M.W., 2007. Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation 24.
- Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (6), 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>.
- Rico-Fernández, M.P., Ríos-Cabrera, R., Castelán, M., Guerrero-Reyes, H.I., Juarez-Maldonado, A., 2019. A contextualized approach for segmentation of foliage in different crop species. *Comput. Electron. Agric.* 156, 378–386. <https://doi.org/10.1016/j.compag.2018.11.033>.
- Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition 1–14. <https://doi.org/10.1016/j.infsof.2008.09.005>.
- Tang, J., Wang, D., Zhang, Z., He, L., Xin, J., Xu, Y., 2017. Weed identification based on K-means feature learning combined with convolutional neural network. *Comput. Electron. Agric.* 135, 63–70. <https://doi.org/10.1016/j.compag.2017.01.001>.
- Uzal, L.C., Grinblat, G.L., Namías, R., Larese, M.G., Bianchi, J.S., Morandi, E.N., Granitto, P.M., 2018. Seed-per-pod estimation for plant breeding using deep learning. *Comput. Electron. Agric.* 150, 196–204. <https://doi.org/10.1016/j.compag.2018.04.024>.
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612.
- Xiong, X., Duan, L., Liu, L., Tu, H., Yang, P., Wu, D., Chen, G., Xiong, L., Yang, W., Liu, Q., 2017. Panicle-SEG: A robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. *Plant Methods* 13, 1–15. <https://doi.org/10.1186/s13007-017-0254-7>.
- Zhou, C., Liang, D., Yang, X., Yang, H., Yue, J., Yang, G., 2018. Wheat ears counting in field conditions based on multi-feature optimization and TWSVM. *Front. Plant Sci.* 9, 1–10. <https://doi.org/10.3389/fpls.2018.01024>.
- Zhu, Y., Cao, Z., Lu, H., Li, Y., Xiao, Y., 2016. In-field automatic observation of wheat heading stage using computer vision. *Biosyst. Eng.* 143, 28–41. <https://doi.org/10.1016/j.biosystemseng.2015.12.015>.