IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 18, 2019, accepted May 12, 2019, date of publication May 16, 2019, date of current version May 30, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2916931

Soybean Seed Counting Based on Pod Image Using Two-Column Convolution Neural Network

YUE LI[1,2], JINGDUN JIA[1], LI ZHANG[1,2], ABDUL MATEEN KHATTAK[2,3], SHI SUN[4], WANLIN GAO[1,2], AND MINJUAN WANG[1,2]

[1]Key Laboratory of Agricultural Informatization Standardization, Ministry of Agriculture and Rural Affairs, China Agricultural University, Beijing 100083, China
[2]College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
[3]Department of Horticulture, The University of Agriculture, Peshawar 25120, Pakistan
[4]Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Corresponding authors: Wanlin Gao (wanlin_cau@163.com) and Minjuan Wang (minjuan@cau.edu.cn)

This work was supported in part by the Project of Scientific Operating Expenses from the Ministry of Education of China under Grant 2017PT19, and in part by the China Postdoctoral Science Foundation under Grant 2018M630222.

ABSTRACT China's soybean supply and demand are seriously imbalanced. It is crucial to improve the level of soybean breeding. Hundred-grain weight is one of the most essential phenotypic parameters for crop breeding. Accurate soybean seed counting is a key step for 100-grain weight. There are several seed counting methods, which have their own limitations one way or the other. Among these, manual counting is time-consuming, electronic automatic seed counter devices are expensive and their counting speed is very slow, and the traditional digital image processing techniques are not suitable for seed counting based on individual pod images. This paper attempted to develop a method that would combine the density estimation-based methods and the convolution neural network (CNN)-based methods to accurately estimate the seed count from an individual soybean pod image with a single perspective. In this paper, we first introduced a new large-scale seed counting dataset, named Soybean-pod. The dataset contains 500 annotated pod images with a total of 32 126 seeds and is the largest annotated dataset for soybean seed counting so far. Simultaneously, we used annotation information to generate a ground-truth density map by convolving a Gaussian kernel and, then, devised a simple but effective method that would elucidate pod images to a seed density map using a two-column CNN (TCNN) and thus accomplish seed counting ultimately. We conducted relevant experiments from three aspects on the new dataset to verify the effectiveness of our model and method, which provided 13.21 mean absolute error (MAE) and 17.62 mean squared error (mse). In addition, our research results showed that deep learning techniques can be easily adapted to precision tasks for plant phenotyping and breeding purposes.

INDEX TERMS Convolution neural network, density map, Gaussian kernel, pod image, soybean seed counting.

I. INTRODUCTION

Soybean is one of China's main grain crops, and it has a great potential for further development as an important source of oils, proteins, and health-care substances. China has always been a big soybean consumer, and its consumption ranks first in the world. With the population growth, improved living standards and changes in dietary structure, the demand for soybeans in China is increasing tremendously. China's annual soybean imports are more than 50 million tons. Besides, there are prominent fluctuations between supply and demand in the consumer market. In recent years, through the unremitting efforts of soybean breeders, the level of soybean yield in China has improved significantly. However, compared with the United States, there is still a considerable gap in production. In order to narrow this gap and solve the serious issue of imbalance between soybean supply and demand, it is vital to improve the level of soybean breeding in China in order to develop high-yielding cultivars. It is essential to expedite the development and cultivation of high-yielding, high-quality, multi-resistant new varieties. The plant breeding process requires phenotypic analysis of large populations. Unfortunately, phenotyping is typically a manual task, which is laborious, costly, and time-consuming. Even more, due to

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Shorif Uddin.

the subjective initiative of the observer, visual assessment of many plants is very difficult and error-prone. As a consequence, phenotyping has become a bottleneck for plant breeding programs [1]–[4].

For soybean breeders, to achieve high-yield cultivars, the recording and tracking of a huge amount of phenotypic data is an essential part of the breeding process. Grain weight is an important component of crop yields [5]. Soybean yields largely depend on three major components: the number of pods per plant (PN), the number of seeds per pod (SPP) and the seed size [6]. Hundred-grain weight is also one of the most intuitive and valuable phenotypic parameters and an effective measure of evaluation in soybean breeding and its cultivation at large [7]. For measuring the hundred-grain weight of soybean, the key is to correctly count the seeds. From the perspective of quickly obtaining the phenotypic parameters of soybeans, it is essential to discover a quick method for estimating the number of soybean seeds. Acquiring a fast access to this trait helps facilitate data recording and ultimately contribute towards the improvement of China's soybean research. Common counting methods are manual counting [8], electronic automatic seed counter devices [9], and traditional digital image analysis technologies [10], [11]. Manual counting is not only time-consuming and laborious but also error-prone and unable to meet application requirements. Electronic seed counting devices use vibration sequencing and photoelectric sensors. However, their counting speed is slow, the equipment is expensive, and the operation is cumbersome and complex. The cost of grain counting through traditional digital image analysis technologies is lower. Nevertheless, the process is more troublesome and cumbersome to manually extract features at the early stage. Moreover, the traditional image processing techniques cannot be fully automated.

Recently, deep learning [12], [13] has become a hot topic in the field of computer vision. Compared with traditional digital image processing technology, the advantage of deep learning is that the network automatically learns and extracts related features instead of manual extraction. Deep learning shows great potential in image-based effective object detection, classification, and counting. For example, Aich and Stavness [14] used state-of-the-art deep learning architectures, i.e. a de-convolutional network for initial segmentation and a convolutional network for leaf counting. The evaluation was performed on the leaf counting challenge dataset at CVPPP-2017. Zhang *et al.* [15] developed a method that could accurately estimate the crowd count from an individual image with arbitrary crowd density and arbitrary perspective. They proposed a simple but effective Multi-Column Convolutional Neural Network (MCNN) architecture to map the image to its crowd density map. Zheng *et al.* [16] proposed a deep convolutional neural network (DCNN) to identify four cucumber diseases. With the augmented datasets containing 14,208 symptom images, the DCNN achieved good recognition results, with an accuracy of 93.4%.

The common image-based counting method is to count the number of seeds in the image. Estimating the number of seeds per pod is a laborious and difficult job, requiring visual inspection of each pod by the breeder [17]. Therefore, the present work is intended to devise a method that would enable to count the seed within the soybean pod using its image. This would not only solve the problem of estimating the number of soybean seeds per pod, even before threshing but also resolve the seed counting issues of other leguminous crops at large.

The aim was to develop a method that can accurately estimate the soybean seed number from an individual pod image with a single perspective. In this study, we first built a new dataset named Soybean-pod containing 500 pod images, with a total of 32126 annotated seeds. Then developed a simple and novel method for counting seeds from soybean pod images using a Two-column Convolution Neural Network.

This paper is organized into the following sections: Section 1 and 2 introduce the motivation of this research, and also some related works. Section 3 details the new dataset we created for seed counting. Section 4 describes the design of the algorithm to estimate seeds from the images. Section 5 shows an experimental result obtained by our algorithm and Section 6 concludes this research and discusses future work.

II. RELATED WORK

The number of seeds is not only a common agronomic parameter but also an important determinant of grain yield. Seed counting is an indispensable part of the soybean breeding process. Many methods have been developed so far to calculate the number of seeds. These studies date back to the development of photoelectric seed detector [18] that further evolved into a completely automated method as described by Severini *et al.* [19]. However, it is very expensive to count seeds using the electronic or laser-based particle counters. The earliest and most commonly used method was that people counted the seeds through naked eyes, but it was very time-consuming and involved high chances of human error [20].

The potential solutions for image-based object counting [29], [32] can be classified into two categories: detection-based methods, and density estimation-based methods.

A. DETECTION-BASED METHODS

In recent years, image analysis technology has been studied and applied. The detection-based methods mainly use digital image processing technology to achieve challenging segmentation tasks, thereby completing the targeted counting [21]. Zhong *et al.* [10] proposed a novel algorithm based on watershed and concavities to segment the large-scale clustered slender-particles efficiently that enabled such quantitative analysis, which was previously infeasible. Mussadiq *et al.* [20] presented and evaluated the performance of four open-source image-analysis programs to count crop seeds from digital images. Although they concluded that seed counts prediction by selected image analysis programs had high accuracy, their performance was easily distinguishable by selected program and crop species. Dorj *et al.* [22] developed a new citrus recognition and counting algorithm, which utilized color features (or schemes) to present an estimate of citrus yield and obtained a correlation coefficient (R2) of 0.93 between the counting algorithm and counting performed through human observation. Many early studies focus on detection-based methods that detect people and count their number through a detector [23]–[25]. These methods need to train a detector capturing information, which is very complex, and it is generally not effective in high-density crowd scenes [26]. In general, detection-based methods have poor stability and scalability.

B. DENSITY ESTIMATION-BASED METHODS

There are many studies on the problem of target counting on density estimation-based methods, especially in the field of crowd counting [27], [30], [31]. However, in terms of seed counting, research on density estimation-based methods is very rare. Recently, by combining deep learning, the convolutional neural network (CNN)-based solutions have achieved greater ability in this task and have outperformed detection-based methods pertaining to high-density crowd scenes. Zhang *et al.* [28] proposed a scale-adaptive CNN architecture and the results demonstrated significant improvements of SaCNN over the state-of-the-art. Uzal *et al.* [17] developed a classic approach to estimate the number of seed per pod, based on tailored features extraction (FE) followed by a Support Vector Machines (SVM) classification model, and also the referred CNNs.

The seed counting task is very similar to the crowd counting task in some respects. There is little difference in seed size of the same pod image we collected. Therefore, the task of seed counting is relatively simpler than crowd counting. Our approach combines density estimation-based methods and CNN-based methods to accurately estimate the soybean seed number from an individual pod image.

III. DATASET

As the existing datasets are not entirely suitable for the evaluation of soybean seed counting task from pod image, in this work we introduced a new soybean seed counting dataset named Soybean-pod. This contains 500 annotated pod images, with a total of 32126 seeds with their centers annotated, and is released with labeled ground truth. This dataset is the largest so far produced for the annotated soybean seeds counting. All the pod images in the dataset were taken with the iPhone 8 camera during daytime under natural light conditions during mid to late October at the Beijing Shunyi Base of the Institute of Crop Sciences, Chinese Academy of Agricultural Sciences.

A. PLANT MATERIAL AND IMAGE CAPTURE

The pods of selected soybeans, from the Beijing Shunyi Base of the Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, were manually prepared by removing any dirt and subsequently subjected to digital imaging.

Soybean pods of different varieties were randomly spread on a piece of black cloth in such a way that the pods did not overlap or touch each other. The original digital images of pods were taken with a camera at daytime under natural light conditions during mid to late October at the Beijing Shunyi Base of the Institute of Crop Sciences, Chinese Academy of Agricultural Sciences. The camera lens was fixed 30-40 cm above a black background cloth. For this purpose, iPhone 8 camera with 12 million pixels, f / 1.8 aperture, manufactured by Apple Inc. was used. For taking a photo, we would turn on the camera mode on the iPhone, select ''photo'', set the focal length to 4 mm, ISO speed to 40, turn off the flash and live, turn on the anti-shake function.

B. DATA ANNOTATION

Since the quality of the captured image was not as good as expected, a series of image processing techniques, such as cropping, flipping, resizing, and adjusting brightness, etc. were performed before labeling. Since the goal was to develop the density map of soybean seeds that would enable the actual seed counts ultimately, the point annotation format was chosen. First the center of the soybean seed in the image was determined and marked as the location. Then Labelme was selected as the annotation tool because it had the ability to implement point annotation and was easy to operate. The obtained point coordinates were used to generate the ground truth density map. The Labelme was installed in Windows 10, Anaconda environment. Figure 1 shows the Labelme visual operation interface, where the seed was marked on the interface.

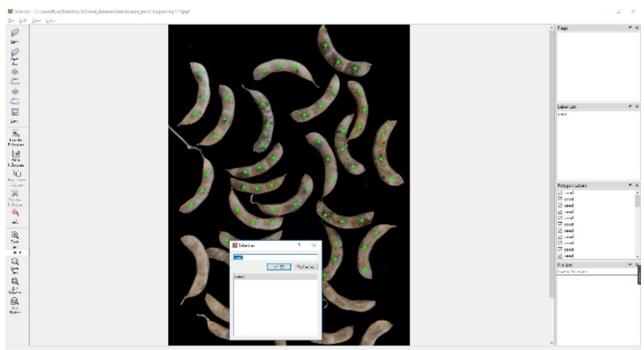


FIGURE 1. The Labelme visual operation interface.

As far as we know, this dataset was the first one devised for soybean seed counting based on pod imaging and also the largest one in terms of the number of annotated seeds. The dataset was further divided into three sets, i.e. a training set, a validation set, and a testing set. For training, 309 images were used that contained 16530 labeled seeds. For validation, 34 images were used that contained 1720 labeled seeds. For testing, 157 images were used that contained 13876 labeled seeds. The details of sets are given in Table 1, which reveals the number of images, the total number of labeled seeds, the minimum number of labeled seeds contained in an image,

TABLE 1. The detail of dataset.

Set	Num	Total	Min	Max	Resolution	Ave
train	309	16530	1	186	Different	53.50
validation	34	1720	2	162	Different	50.59
test	157	13876	3	176	Different	88.38
Soybean-pod	500	32126	1	186	Different	64.25

Dataset properties: Num stands for the number of images, Total is the number of labeled seeds, Min represents the minimum number of labeled seeds contained in a single image. Max represents the maximum number of labeled seeds contained in a single image, and Ave denotes the average seed count

FIGURE 2. The top row indicates representative images of the train set in the new 'Soybean-pod' seeds dataset. The bottom row indicates representative images of test set in the new 'Soybean-pod' seeds dataset.

the maximum number of labeled seeds contained in an image, and the average seed count for the mentioned sets. Also, the color images are shown in Figure 2.

C. DATA AUGMENTATION

There were two problems in analyzing the captured images, 1. the pixels of the entire image were too large to be suitable for network training, and 2. the number of training sets used to train the network was relatively small that could provide poor network test results. Considering the above two problems, it was decided to augment the training dataset. The way the data augmentation was used was to extend the training data set. The specific methods were to reduce the pixels of the input image for the convolutional neural network and increase the number of input images.

In order to increase the number of input images as much as possible with the limited number of original images, crop 9 patches were considered from each original image at different locations. The input image with too small pixel would provide bad training result. In order to avoid that, the size of the patch was set to 1/4 of the original image. Thus, the number of patches used to train the network was 3087, much larger than the original number. The importance and necessity of using data enhancement are demonstrated in the fifth part of this article. The example patches are shown in Figure 3.

IV. OUR APPROACH

A. DENSITY MAP BASED SEED COUNTING

There are three kinds of Convolutional Neural Networks (CNN) estimating the number of seeds in a given color

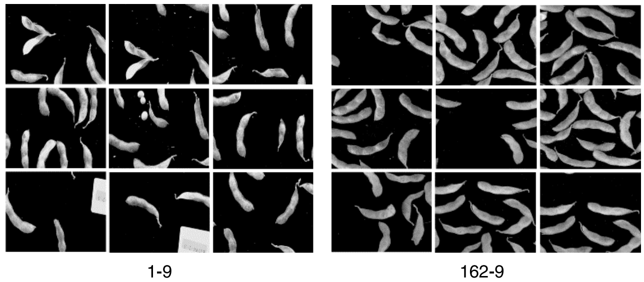


FIGURE 3. The left indicates the 9 example patches cropped on the first image, the right indicates the 9 example patches cropped on the 162nd image.

pod image. The first one is a network that identifies pod types that contain different numbers of seeds and count the same kind of pods, then use the formula $N = 1x_1 + 2x_2 + 3x_3 + 4x_4$ to calculate the number of seeds in each color pod image(SPP means seed per pod, i-SPP means that the number of seed per pod is i, x_i is the number of i-SPP). The second one is a network whose input is the image and output is the estimated seed count. The third one is to output a density map of the seed (say how many seed per square meter), and then obtain the seed count by integration.

In this study, the third kind was favored for the following reasons: 1. Density map preserves more information, such as grain position. 2. Compared to the network that identifies and counts pod, a density map of the seed can direct counting without completing recognition and classification tasks. 3. The soybean seeds could not be directly seen in the pod image, hence the second method of directly identifying seeds was not feasible. So task for estimating the number of seeds based on pod images, we designed the network that input is the pod map, output is seed estimation density map, then integrated this map to count the seeds.

B. GROUND TRUTH DENSITY MAP

Since CNN-based approaches predict the density map from an input image by training CNN, it was necessary to provide a high quality of ground truth density map in the training data in order to improve the performance of our method.

The ground truth seeds density map was generated by convolving annotated points with a Gaussian kernel:

$$D_i^{gt}(p \mid I_i) = \sum_{\mu \in A_i^{gt}} \mathcal{N}^{gt}(p; \mu, \sigma) \quad \forall p \in I_i$$

where A_i^{gt} is the set of 2D points annotated for the image I_i, indicating the ground truth seeds positions in the image. The ground truth density map D_i^{gt} at a specific pixel p of I_i is generated by convolving annotated points with a Gaussian kernel, $\mathcal{N}^{gt}(p; \mu, \sigma)$ represents the evaluation of a normalized 2D Gaussian function, with mean μ and isotropic covariance matrix σ (spread parameter). The sum of the density map is equivalent to the total number of seed in an image.

The spread parameter σ is determined based on the size of the seed within the image. Since the seed is relatively sparse, the same spread parameter is used in Gaussian kernel to

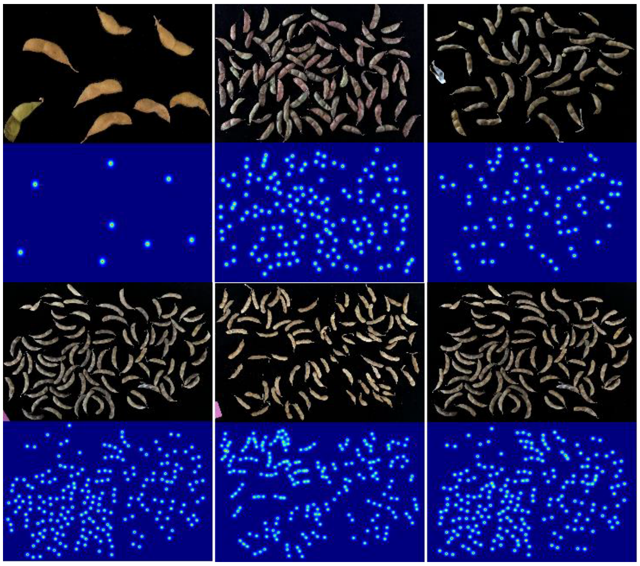


FIGURE 4. The first row indicates images of the test set. The second indicates the corresponding ground truth density map. The third indicates images of the train set. The fourth row indicates the corresponding ground truth density map.

generate the ground truth density maps. The example ground truth density maps are shown in Figure 4.

C. COUNTING ARCHITECTURE

In designing the network architectures, two points were considered: (1) soybean seed counting has problems as seeds are usually small compared to the whole image. At the same time, the distance between the camera and the pod is fixed at 30-40cm. Therefore, deep networks that represent deeply semantic information are not necessary; (2) based on this, we considered to adjust the number of columns based on MCNN. We also considered two parallel CNNs architectures. However, since the soybean pods are not threshed in the image, it is relatively difficult to count the seed through the pod image. The interest was to find out whether networks of this simplicity would be able to deal with these complications or not?

Motivated by the success of Multi-column CNN (MCNN), we also proposed to use filters with different sizes of the local receptive field to learn the map from the raw pixels to the density maps. To simplify, a network architecture is presented, which has two columns CNNs with the same network structures, namely TCNN-seed. Figure 5 gives an overview of the network architecture. As the figure shows, two parallel CNNs structures are almost similar, which are Conv2d-Maxpooling-Conv2d-Maxpooling-Conv2d-Conv2d. The difference is the sizes and numbers of filters, we used 3×3 and 5×5 sizes of local receptive fields to capture seeds at different scales detecting blob. Max pooling layer has a 2×2 size kernel. All the previous convolutional layers are followed by normalization layers and rectified linear unit (ReLU). We used ReLU as an activation function because it makes the model converge fast and maintain a steady state. We stacked the output feature maps of the two columns CNNs. Finally, the last Conv2d map the total feature map to a density map with another 1×1 filter for the input patch. Note that since our architecture combines CNN with additional nonlinear and normalization layers, there is no need to impose strict requirements on the shape of the input image.

1) CONVOLUTIONAL LAYERS

There are four convolutional layers in each column, and there is a convolutional layer that is used to generate the density map after concatenating the feature map. The kernels of four convolutional layers in one column are 3×3 with the dimensions of 40, 80, 40 and 20. The kernels of four convolutional layers in the other column are 5×5 with the dimensions of 20, 40, 20 and 10.

2) BATCH NORMALIZATION

The input data for each layer is normalized during training: an operation with a mean of 0 and a variance of 1 is called batch normalization (BN), which gives a stable distribution for the input of each layer. The advantage of BN is to speed up the convergence and promote network training. It is usually placed after the convolutional layer and in front of the nonlinear layer (the activation function). The BN is defined as follow:

$$BN_{\gamma,\beta}(x) = \frac{x - E[x]}{\sqrt{Var[x] + \epsilon}} \times \gamma + \beta$$

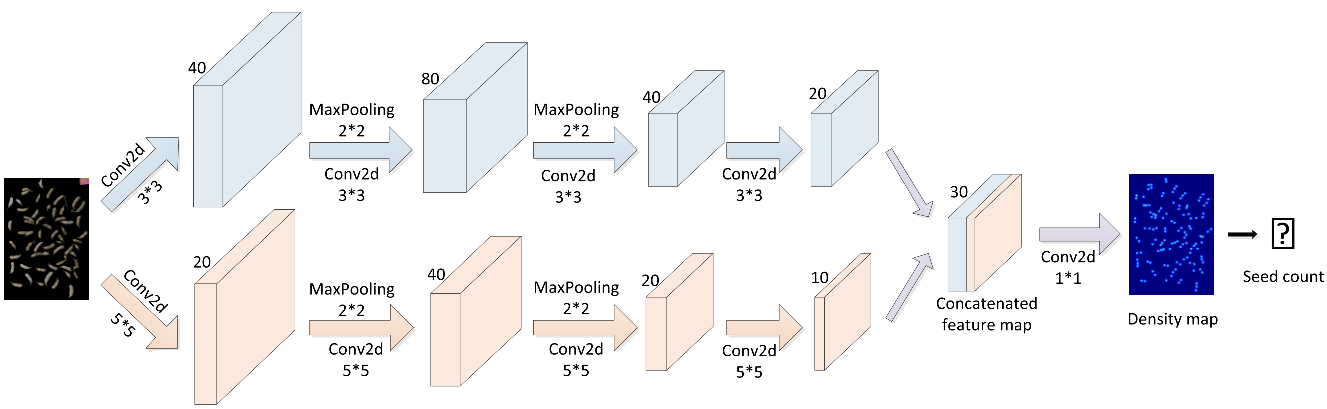


FIGURE 5. TCNN-seed: An overview of our network architecture.

where $E[x]$ is mean of mini-batches, $Var[x]$ is the standard deviation of mini-batches, they can all be calculated during the training network. γ and β are learnable parameter vectors.

3) ACTIVATION FUNCTION

ReLU, as one of the most popular activation function, has become more and more popular during the recent years. There are two advantages to ReLU: 1. Maximum speed of convergence of accelerated stochastic gradient descent method; 2. Its calculation is not complicated. This is why we chose to use it as an activation function. The function is defined as follow:

$$f(x) = max(0, x)$$

The function of ReLU only needs a threshold to get the activation value without other extra mass operations. It simply saves the part larger than 0 and sets the part smaller than 0 to 0.

4) POOLING

The function of the pooling layer is to gradually reduce the spatial size of the data volume, which can reduce parameters, reduce the computational resource cost, and also effectively control over-fitting. Commonly, mean pooling and max pooling are two forms for the pooling layer.

Max pooling is choosing the max value of the image area as the pooled value of this area. In this research, we use two layers of max pooling with the size of 2×2 between two convolution layers.

5) THE FORMULA USED TO CALCULATE THE NUMBER OF SEEDS

The input of the TCNN is the image, and its output is a seed density map, the integral of which gives the overall seed count. The formula used to calculate the number of seeds is as follows:

$$C^{et}(N) = \sum_{i=0,j=0}^{i=I,j=J} d_{ij}^{et}(N)$$

where $C^{et}(N)$ is the estimated number of seeds contained in test image N, $d_{ij}^{et}(N)$ represents the estimated density value on each pixel of the image N obtained by the network optimal model.

D. OPTIMIZATION OF TCNN

We use the Euclidean distance to measure the difference between the estimated density map generated during the training network and ground truth given in the training data. The loss function that is similar to other works [26,27,28] is defined as follow:

$$L(\Theta) = \frac{1}{2N} \sum_{i=1}^{N} \left\| D(X_i; \Theta) - D_i^{gt} \right\|_2^2$$

where Θ is a set of learnable parameters in the CNN. N is the number of the training images, X_i is the input image and D_i^{gt} stands for the ground truth density map of the image X_i. $D(X_i; \Theta)$ denotes the estimated density map generated by CNN, which is parameterized with Θ for X_i. L is the loss between ground truth density map and estimated density map.

As the number of training samples was very limited and the gradient effect was vanishing for deep neural networks, it was not easy to learn all the parameters simultaneously. Motivated by the success of pre-training, we pre-trained CNN in every single column by directly mapping the outputs of the fourth convolutional layer to the density map. Then used these pre-trained CNNs to initialize CNNs in all columns and fine-tuned all the parameters simultaneously. At the same time, in the process of training the network, we used the validation set to evaluate and optimize the model.

V. EXPERIMENT AND EVALUATION

The experiments were performed on an Ubuntu 18.4 64-bits PC equipped with an Intel®Xeon(R) CPU E5-2630 v4 @ 2.20GHz×20 processor, and 31.3 GB-RAM. For deep learning technology, implementation of the proposed network and its training were based on the PyTorch framework in NVIDIA 1080Ti GPU.

We first presented the evaluation metric that we used to evaluate our method in the experiment. Then, conducted relevant experiments from three aspects of Soybean-pod dataset to verify the effectiveness of our model and method. The three aspects were different training set, different network structure and different test set. Finally, by comparison, we demonstrate the rapidity and low cost of estimating the number of seeds by our method.

A. EVALUATION METRIC

Consistent with previous works, we evaluated our methods and the state-of-the-art counting methods with both the Mean Absolute Error (MAE) and the Mean Squared Error (MSE). MAE is defined as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |e_i - a_i|$$

where N is the number of test samples, e_i is the estimated number of seeds in the i^{th} image predicted by the model being evaluated, and a_i is the actual number of seeds in the i^{th} image from seed labeled annotations. MAE indicates the accuracy of the predicted seed count across the test sequence. MSE indicates the robustness of the predicted count. MSE is defined as follows:

$$\text{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (e_i - a_i)^2}$$

B. DIFFERENT TRAINING SETS: WITHOUT AUGMENTATION VS AUGMENTATION

We separately used images (without augmentation) and patches (augmentation) as training data to train the network (TCNN-seed) introduced in *Our Approach*. In the process of training the network, the validation set was used to evaluate and optimize the model, then the test set was estimated with

TABLE 2. Mean absolute error (MAE) and mean squared error (MSE) of images and patches.

	MAE	MSE
images	31.93	50.34
patches	**13.21**	**17.62**

TABLE 3. Mean absolute error (MAE) and mean squared error (MSE) of three networks.

	MAE	MSE
CNN(3×3)	23.49	34.11
CNN(5×5)	20.76	27.91
TCNN	**13.21**	**17.62**

the optimal model obtained by each. Table 2 shows the relevant result data. We can see that using the data-enhanced patches as the training data, the MAE and MSE obtained in the test are smaller, which proves that data enhancement is very important and necessary because this makes the network model achieve better estimation results.

C. DIFFERENT NETWORK STRUCTURE: SINGLE COLUMN CNNS VS TCNN-SEED

We used 3,087 patches as training sets to train three networks: single column CNN (3 × 3), single column CNN (5 × 5) and TCNN-seed. Table 3 provides the relevant result data. It can be seen that TCNN-seed significantly outperforms each single column CNN for both MAE and MSE.

D. DIFFERENT GROUPS ON THE TEST SET

We divided the test images into 10 groups according to seed count in an increasing order. There were 157 test images. Group 1-5 all had 15 images each, Group 6, 7 and 10 contained 16 images each, and all the other groups had 17 images each. We used the final model obtained with the 3087 patches training network (TCNN-seed) to test 10 groups separately. The different groups are compared in Table 4. All the images of the test set were arranged in an ascending order according to the number of seeds per image and then divided into ten groups. Among them, the Mean Absolute Error (MAE) and the Mean Squared Error (MSE) of groups 3, 7, and 9 were relatively small. Group 6 had a smaller MAE, but its MSE was relatively larger. This depicts that the proposed approach works better overall on images with 100 ± 5 seeds, but with more extreme outliers.

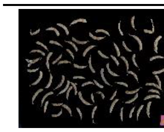
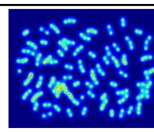
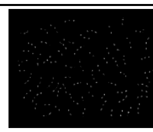
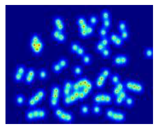
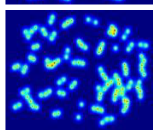
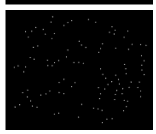
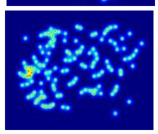
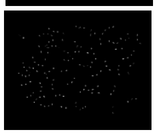
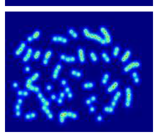
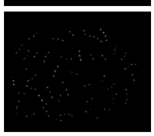
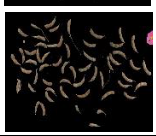
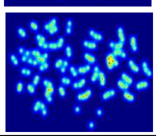
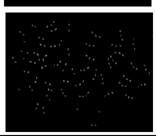
Table 5 depicts the experimental evaluation of the presented method, including the images in the testing set, the corresponding ground truth density map, estimated density map, ground truth seed number, and estimated seed number. The results reveal that the new technique is useful for counting the seed.

The estimated absolute counts and ground truth absolute counts of each group were also compared, as shown in figure 6. The absolute count in the vertical axis is the average seed number of images in each group. The plot shows that

TABLE 4. Mean absolute error (MAE), mean squared error (MSE), average ground-truth number of seeds per image (GT-count), and average estimated number of seeds per image (ET-count) of 10 groups.

Group	MAE	MSE	GT-count	ET-count
G1	12.38	15.72	8.47	20.26
G2	18.26	27.25	23.40	39.41
G3	7.29	9.51	49.93	49.22
G4	15.78	18.88	70.00	62.03
G5	15.75	18.21	89.00	86.90
G6	9.36	15.55	98.44	103.50
G7	6.48	9.22	110.63	110.30
G8	19.25	21.4	123.65	108.84
G9	10.89	12.68	135.29	125.42
G10	16.68	20.1	157.31	152.44

TABLE 5. Experimental assessment of the new method. The first column shows six original images in the testing set, the second column is the corresponding ground truth (GT) density map, the third column is the estimated (Et) density map, and the fourth column is ground truth seed number and estimated seed number.

Original image	GT-density map	Et-density map	Detail
			Image-100 GT-cou:176 Et-cou:179.23
			Image-129 GT-counts:81 Et-cou:81.86
			Image-135 GT-cou:112 Et-cou:111.25
			Image-139 GT-cou:123 Et-cou:127.17
			Image-140 GT-cou:109 Et-cou:107.38
			Image-143 GT-cou:130 Et-cou:126.07

proposed method provides the best seed counting estimate from an image with 40-120 seeds.

E. THE REAL-TIME AND COST OF OUR APPROACH

The proposed method was compared with the existing approach in terms of real-time performance. The existing approach used by Soybean breeder was manual counting. While collecting original images, we investigated the

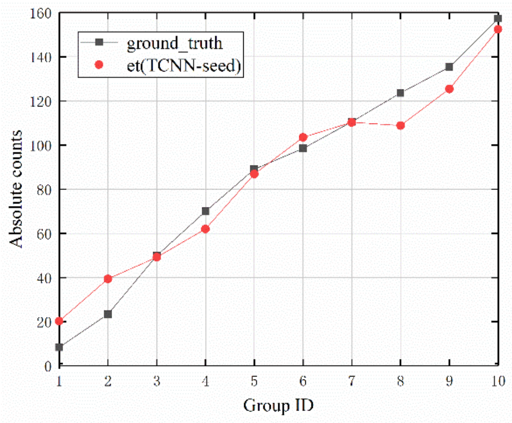


FIGURE 6. Comparison of the estimated absolute (et) counts and ground truth absolute counts within each group. Absolute count (shown in Y-axis) is the average seed number of images in each group.

situation of five soybean breeding workers who completed counting seeds within three days. According to statistics, it takes about one minute for each person to count the number of seeds on a soybean plant. We took one picture of all the pods of a soybean plant. Thus the test set had a total of 157 images, equivalent to 157 soybean plants. The time required to manually count the seeds of 157 plants would be two hours and thirty-seven minutes, assuming each plant seed counting took one minute. However, using the proposed system it would take only 1'59''73 (less than two minutes) to test 157 images. Thus, compared with manual counting, this method saved about 2 hours and 35 minutes. This proves that the proposed method has the capability to count seeds efficiently.

Electronic seed counters are very expensive. The price of an ordinary electronic automatic seed counter is several thousand yuan. Based on the current level of hardware and software development, the proposed method can easily be turned into a simple software or even a small program. The cost of this will be very low compared with the electronic seed counting devices.

VI. CONCLUSION

The image used to count soybean seeds is generally a seed image without other impurities after the threshing process. However, in this study, the soybean seeds were counted from the pod image based on the convolutional neural network, which has never been reported before. To better evaluate performances of soybean seed counting methods, we first introduced a new large-scale seed counting dataset, named Soybean-pod. It contained 500 annotated pod images, with a total of 32126 seeds with their centers annotated. For soybean seed counting, this is the largest annotated dataset so far. Then a Two-column Convolution Neural Network was proposed, which estimated seed number accurately in a single image. The proposed estimation model proved to be faster than manual seed counting and automatic seed counters. Experimental data showed that our method provided the best estimate for an image with 40-120 seeds. The results indicate that deep learning techniques can be easily adapted to precision tasks in plant phenology, helping in the improvement of current and future breeding programs. The future focus will be on using the deep learning method to achieve image-based counting of the number of pods per soybean plant and the number of threshed seeds for plant breeding and yield estimation tasks in other crops as well.

REFERENCES

[1] A. Singh, B. Ganapathysubramanian, A. K. Singh, and S. Sarkar, "Machine learning for high-throughput stress phenotyping in plants," *Trends Plant Sci.*, vol. 21, no. 2, pp. 110–124, Feb. 2016.

[2] F. Fiorani and U. Schurr, "Future scenarios for plant phenotyping," *Annu. Rev. Plant Biol.*, vol. 64, pp. 267–291, Apr. 2013.

[3] M. E. Ghanem, H. Marrou, and T. R. Sinclair, "Physiological phenotyping of plants for crop improvement," *Trends Plant Sci.*, vol. 20, no. 3, pp. 139–144, Mar. 2015.

[4] N. Fahlgren, M. A. Gehan, and I. Baxter, "Lights, camera, action: High-throughput plant phenotyping is ready for a close-up," *Current Opinion Plant Biol.*, vol. 24, pp. 93–99, Apr. 2015.

[5] G. A. Slafer, R. Savin, and V. O. Sadras, "Coarse and fine regulation of wheat yield components in response to genotype and environment," *Field Crops Res.*, vol. 157, pp. 71–83, Feb. 2014.

[6] W. R. Fehr *et al.*, *Principles of Cultivar Development: Theory and Technique*, vol. 1. New York, NY, USA: Macmillan, 1987.

[7] J. Li, M. Thomson, and S. R. McCouch, "Fine mapping of a grain-weight quantitative trait locus in the pericentromeric region of rice chromosome 3," *Genetics*, vol. 168, no. 4, pp. 2187–2195, 2004.

[8] T. Liu *et al.*, "A shadow-based method to calculate the percentage of filled rice grains," *Biosyst. Eng.*, vol. 150, pp. 79–88, Oct. 2016.

[9] N. B. Mclaughlin, J. Giesbrecht, and D. F. Bligh, "Design and performance of an electronic seed counter," *Can. J. Plant Sci.*, vol. 56, no. 2, pp. 351–355, 1976.

[10] Q. Zhong, P. Zhou, Q. Yao, and K. Mao, "A novel segmentation algorithm for clustered slender-particles," *Comput. Electron. Agricult.*, vol. 69, no. 2, pp. 118–127, Dec. 2009.

[11] H. Scharr, H. Dee, A. P. French, and S. A. Tsaftaris, "Special issue on computer vision and image analysis in plant phenotyping," *Mach. Vis. Appl.*, vol. 27, no. 5, pp. 607–609, 2016.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[13] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agricult.*, vol. 147, pp. 70–90, Apr. 2018.

[14] S. Aich and I. Stavness, "Leaf counting with deep convolutional and deconvolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2080–2089.

[15] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in *Proc. CVPR*, Jun. 2016, pp. 589–597.

[16] F. Zheng, L. Zhang, Z. Gong, K. Du, J. Ma, and Z. Sun, "A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network," *Comput. Electron. Agricult.*, vol. 154, pp. 18–24, Nov. 2018.

[17] L. C. Uzal *et al.*, "Seed-per-pod estimation for plant breeding using deep learning," *Comput. Electron. Agricult.*, vol. 150, pp. 196–204, Jul. 2018.

[18] W. S. Reid, D. J. Buckley, and W. Mason, "A photoelectric seed counting detector," *J. Agricult. Eng. Res.*, vol. 21, no. 2, pp. 213–215, Jun. 1976.

[19] A. D. Severini, L. Borras, and A. G. Cirilo, "Counting maize kernels through digital image analysis," *Crop Sci.*, vol. 51, no. 6, pp. 2796–2800, Nov. 2011.

[20] Z. Mussadiq, B. Laszlo, L. Helyes, and C. Gyuricza, "Evaluation and comparison of open source program solutions for automatic seed counting on digital images," *Comput. Electron. Agricult.*, vol. 117, pp. 194–199, Sep. 2015.

[21] R. Makanza *et al.*, "High-throughput method for ear phenotyping and kernel weight estimation in maize using ear digital imaging," *Plant Methods*, vol. 14, no. 1, p. 49, Dec. 2018.

[22] U.-O. Dorj, M. Lee, and S.-S. Yun, "An yield estimation in citrus orchards via fruit detection and counting using image processing," *Comput. Electron. Agricult.*, vol. 140, pp. 103–112, Aug. 2017.
[23] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 743–761, Apr. 2012.
[24] G. J. Brostow and R. Cipolla, "Unsupervised Bayesian detection of independent motion in crowds," in *Proc. CVPR*, Jun. 2006, pp. 594–601.
[25] V. Rabaud and S. Belongie, "Counting crowded moving objects," in *Proc. CVPR*, Jun. 2006, pp. 705–711.
[26] Y. Li, X. Zhang, and D. Chen, "CSRNet: Dilated convolutional neural networks for understanding the highly congested scenes," in *Proc. CVPR*, Jun. 2018, pp. 1091–1100.
[27] L. Boominathan, S. S. Kruthiventi, and R. V. Babu, "CrowdNet: A deep convolutional network for dense crowd counting," in *Proc. 24th ACM Int. Conf. Multimedia*, Oct. 2016, pp. 640–644.
[28] L. Zhang, M. Shi, and Q. Chen, "Crowd counting via scale-adaptive convolutional neural network," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Mar. 2018, pp. 1113–1121.
[29] S. Aich and I. Stavness. (2018). "Object counting with small datasets of large images." [Online]. Available: https://arxiv.org/abs/1805.11123
[30] D. B. Sam, S. Surya, and R. V. Babu, "Switching convolutional neural network for crowd counting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4031–4039.
[31] M. Marsden, K. McGuinness, S. Little, and N. E. O'Connor. (2016). "Fully convolutional crowd counting on highly congested scenes." [Online]. Available: https://arxiv.org/abs/1612.00220
[32] S. Aich and I. Stavness. (2018). "Improving object counting with heatmap regulation." [Online]. Available: https://arxiv.org/abs/1803.05494

YUE LI is currently pursuing the master's degree with the College of Information and Electrical Engineering, China Agricultural University, Beijing, China. Her research focuses on object counting and classification based on image using convolutional neural networks for crop phenotypes, with objects primarily related to crops such as soybean seeds.

JINGDUN JIA is currently a Researcher with the China Rural Technology Development Center. He is also a Committee Member of the Policy Advisory Board for the Australian Center for International Agricultural Research and an Adjunct Professor with China Agricultural University. He has long been engaged in development strategy, plan, and policy for science and technology management. He also involved in agricultural and rural development and regional development strategy. He has conducted in-depth research on rural scientific and technological innovation, agricultural biotechnology and food industry, biological energy and biomass industry, nutrition and health, and intelligent agricultural scientific and technological innovation.

LI ZHANG is currently pursuing the Ph.D. degree with the College of Information and Electrical Engineering, China Agricultural University, Beijing China. Her researches focus on deep learning for classification, object recognition, tracking, detection, segmentation for the vision system of agricultural robot, and image/video processing techniques, including enhancement and denoising.

ABDUL MATEEN KHATTAK received the Ph.D. degree in horticulture and landscape from The University of Reading, U.K., in 1999. He was a Research Scientist in different agriculture research organizations before joining the University of Agriculture Peshawar Pakistan, where he is currently a Professor with the Department of Horticulture. He has conducted academic and applied research on different aspects of tropical fruits, vegetables, and ornamental plants. He was also with Alberta Agriculture and Forestry, Canada, as a Research Associate, and the Organic Agriculture Centre of Canada as a Research and Extension Coordinator (for Alberta province), where he helped in developing organic standards for greenhouse production and energy saving technologies for Alberta greenhouses. He is a Professor with considerable experience in teaching and research. He is also a Visiting Professor with the College of Information and Electrical Engineering, China Agricultural University, Beijing. He has published 55 research articles in scientific journals of international repute. Some of his research areas include greenhouse production, medicinal, aromatic and ornamental plants, light quality, supplemental lighting and temperature effects on greenhouse crops, aquaponics, and organic production. He has also attended and presented in several international scientific conferences.

SHI SUN received the B.S. degree in agronomy from Yangzhou University, in 1995, and the Ph.D. degree in crop genetics breeding from Nanjing Agricultural University. He has been involved in soybean breeding for more than 20 years, since 1995. In the past decades, he has released 30 soybean cultivars and has published 50 refereed journal articles. His research focuses on conventional and molecular breeding of soybean, especially for the disease-resistant breeding.

WANLIN GAO received the B.S., M.S., and Ph.D. degrees from China Agricultural University, in 1990, 2000, and 2010, respectively.

He is currently the Dean of the College of Information and Electrical Engineering, China Agricultural University. He has been the Principal Investigator (PI) of over 20 national plans and projects. He has published 90 academic papers in domestic and foreign journals, and among them, over 40 are cited by SCI/EI/ISTP. He has written two teaching materials, which are supported by the National Key Technology R&D Program of China during the 11th Five-Year Plan Period, and five monographs. Moreover, he owns 101 software copyrights, 11 patents for inventions, and eight patents for new practical inventions. His major research area is the informationizion of new rural areas, intelligence agriculture, and the service for rural comprehensive information. He is also the member of the Science and Technology Committee of the Ministry of Agriculture, the member of the Agriculture and Forestry Committee of Computer Basic Education in Colleges and Universities, and a Senior Member of the Society of Chinese Agricultural Engineering.

MINJUAN WANG received the Ph.D. degree from the School of Biological Science and Medical Engineering, Beihang University, under the supervision of Prof. H. Liu, in 2017. She was a Visiting Scholar with the School of Environmental Science, Ontario Agriculture College, University of Guelph, from 2015 to 2017. She is currently the Postdoctoral Fellow with the School of Information and Electrical Engineering, China Agricultural University. Her researches mainly focus on bioinformatics and the Internet of Things key technologies.