Article

Automatic Counting of in situ Rice Seedlings from UAV Images Based on a Deep Fully Convolutional Neural Network

Jintao Wu ^{1,2,3} , Guijun Yang ^{1,2,*} , Xiaodong Yang ^{1,2}, Bo Xu ², Liang Han ^{1,2} and Yaohui Zhu ^{1,2}

¹ Key Laboratory of Quantitative Remote Sensing in Agriculture of Ministry of Agriculture, Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China; e16101001@stu.ahu.edu.cn (J.W.); yangxd@nercita.org.cn (X.Y.); tbp1600202045@student.cumtb.edu.cn (L.H.); yaohui_zhu@bjfu.edu.cn (Y.Z.)

² National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China; xub@nercita.org.cn

³ School of Computer Science and Technology, Anhui University, Hefei 230601, China

* Correspondence: yanggj@nercita.org.cn

Received: 1 February 2019; Accepted: 19 March 2019; Published: 22 March 2019

Abstract: The number of rice seedlings in the field is one of the main agronomic components for determining rice yield. This counting task, however, is still mainly performed using human vision rather than computer vision and is thus cumbersome and time-consuming. A fast and accurate alternative method of acquiring such data may contribute to monitoring the efficiency of crop management practices, to earlier estimations of rice yield, and as a phenotyping trait in breeding programs. In this paper, we propose an efficient method that uses computer vision to accurately count rice seedlings in a digital image. First, an unmanned aerial vehicle (UAV) equipped with red-green-blue (RGB) cameras was used to acquire field images at the seedling stage. Next, we use a regression network (Basic Network) inspired by a deep fully convolutional neural network to regress the density map and estimate the number of rice seedlings for a given UAV image. Finally, an improved version of the Basic Network, the Combined Network, is also proposed to further improve counting accuracy. To explore the efficacy of the proposed method, a novel rice seedling counting (RSC) dataset was built, which consisted of 40 images (where the number of seedlings varied between 3732 and 16,173) and corresponding manually-dotted annotations. The results demonstrated high average accuracy (higher than 93%) between counts according to the proposed method and manual (UAV image-based) rice seedling counts, and very good performance, with a high coefficient of determination (R^2) (around 0.94). In conclusion, the results indicate that the proposed method is an efficient alternative for large-scale counting of rice seedlings, and offers a new opportunity for yield estimation. The RSC dataset and source code are available online.

Keywords: rice seedlings; object counting; computer vision; deep learning; fully convolutional neural networks

1. Introduction

Rice is an important primary food that plays an essential role in providing nutrition to most of the world's population, particularly in Asia [1–3]. Therefore, methods for estimating rice yield have received significant research attention. The number of rice seedlings per unit area (rice seedling density) is an important agronomic component. It is not only closely associated with yield, but also plays an important role in the determination of survival rate. From a breeding perspective, researchers are interested in breeding varieties, and the survival rate of rice seedlings can provide a benchmark

for the selection of breeding materials. Thus, accurate determination of the number of rice seedlings is vital for estimating rice yield and is a key step in field phenotyping. In practice, counting rice seedlings still largely depends on manual human efforts. This is time-consuming and labor-intensive for researchers conducting large-scale field measurements. There is thus a pressing need to develop a fast, non-destructive, and reliable technique that can accurately calculate the number of rice seedlings in the field.

Advanced remote sensing has become a popular technique for obtaining large-scale and high-throughput crop information in plant phenotyping, due to its ability to capture multi-temporal remote sensing (RS) images of field crop growth on-demand [4]. In general, the satellite-based platform, the ground-based platform, and the unmanned aerial vehicle (UAV)-based platform are the three commonly used remote sensing platforms. Of these, the UAV-based platform can be used to conduct frequent flight experiments where and when needed, which allows for observation of high-resolution spatial patterns to capture a large number of multi-temporal high-definition images for crop monitoring [5]. Further, their low cost and high flexibility make UAV-based platforms popular in field research [6], and many novel applications for UAV image analysis have been proposed. These novel applications include not only vegetation monitoring [7–9], but also urban site analysis [10,11], disaster management, oil and gas pipeline monitoring, detection and mapping of archaeological sites [12], and object detection [13–16]. In this paper, we aimed to study the automatic counting of large-scale rice seedlings based on UAV images and provide a basis for predicting yields and survival rates.

There is no doubt that image-based techniques provide a feasible, low-cost, and efficient solution for crop counting, such as wheat ear counting [17,18], oilseed rape flower counting [7], and sorghum head counting [19]. However, because of the complexity of the field environment (e.g., illumination intensity, soil reflectance, and weeds, which alter colors, textures, and shapes in crop UAV images) and the diversity of image capture parameters (e.g., shooting angle and camera resolution, which can easily cause crops to be blurred in UAV images), accurate crop counting remains an enormous challenge. Recently, the so-called object detection technique [20–25] has made great progress in the field of computer vision. The purpose of object detection is to localize a single instance and output its corresponding location information. Therefore, it is straightforward to export the number of instances based on the number of positions output. An interesting question is “should we apply the object detection technique to the crop counting problem?” Unfortunately, crop UAV images generally contain high-density large-scale object instances. When processing these images, the two-stage strategy [20–23] and the one-stage strategy [24,25] are inefficient and imprecise, respectively. Therefore, this type of method is not a good choice for addressing our problem. In addition, for high-density crop counting problems, we are more concerned with the number of crops than with the specific location of each crop. In this paper, we show that it is better to model the task of counting rice seedlings as a typical counting problem, rather than a detection problem. Indeed, object detection is generally more difficult to solve than object counting.

Over the past decade, significant research attention has been paid to computer vision to improve the precision of counting crowds [26–28], cells [29,30], cars [31,32], and animals [33]. However, few studies have focused on addressing crop-related computing tasks. Indeed, to the best of our knowledge, only two published articles have thus far considered counting problems related to crops. Rahnemoonfar and Sheppard [34] proposed a simulated deep convolutional neural network for automatic tomato counting. Lu et al. [35] proposed an effective deep convolutional neural networks (CNN)-based solution, TasselNet, to count maize tassels in wild scenarios. However, neither of these studies was conducted in a large-scale, high-density crop context and the former study only reported results for potted plants, which is very different from a field-based scenario. In contrast, our experiments used UAV images of large-scale rice seedlings that were accurately captured in an unconstrained in-field environment, leading to a more challenging situation which is representative of actual conditions on the ground. To address this challenge, we take inspiration from Boominathan et al. [26], and propose a deep fully convolutional neural network-based approach that utilizes computer vision techniques

to count the number of rice seedlings. The specific objectives of this study include: (1) exploring the potential of deep-learning modeling based on remote sensing to address in-field counting of rice seedlings; (2) verifying an improved method to estimate the number of rice seedlings and obtain more accurate results; and (3) providing a new paradigm for rice-seedling-like in-field counting problems.

To explore the efficacy of the proposed approach, a novel rice seedling counting (RSC) dataset was constructed and is released alongside this paper. The main goal of this dataset is to offer a basis for object counting researchers. Specifically, we have collected 40 high-resolution images from the UAV platform. The number of rice seedlings in the images varies between 3732 and 16,173. We also performed an augmentation on this dataset to generate more than 10,000 sub-images. Similar to the ground truth generation method in object counting problems [26,27,35], each rice seedling in the UAV image will be assigned a density that sums to one. During the prediction, the total number of rice seedlings can be reflected by summing over the whole density map. We believe that such a dataset can serve as a benchmark for assessing in situ seedling counting methods and will attract the attention of practitioners working in the field to address counting challenges.

2. Materials

2.1. Experimental Site and Imaging Devices

The case study field planted with rice was located in the Tangquan Farm comprehensive experimental base of the Chinese Academy of Agricultural Sciences (Tangquan Farm, Pukou District, Nanjing, China, 32°2'55.67" latitude North, 118°38'4.43" longitude East). The sowing date was 1 June 2018. UAV remote sensing images were acquired using an eight-rotor UAV (S1000, DJI Company, Guangdong, China) equipped with a red-green-blue (RGB) camera (QX-100 HD camera, Sony, Tokyo, Japan) with a resolution of 5472×3648 and a pixel size of $2.44 \times 2.44 \mu\text{m}^2$. Flight campaigns were conducted from 2:00 p.m. to 4:00 p.m. on 13 June 2018 at a flight altitude and flight speed of 20 m and 5 m/s, respectively. The forward overlap was 80% and the lateral overlap was 75%. The weather was sunny without significant wind, so the possibility of image distortion due to weather conditions was eliminated.

2.2. Rice Seedling Counting Dataset

We collected 40 rice seedling RGB images from the UAV platform, all of which were manually annotated in our RSC dataset. Due to the large number of seedlings in each image (the number of seedlings varied between 3732 and 16,173), eleven people were involved in image annotation and this required about 123 person-hours of work. These data provide the benchmark to determine the accuracy of our proposed seedling counting method. Indeed, dotting is regarded as a natural way for humans to count. It not only provides raw counts but also illustrates the spatial distribution of objects. Figure 1 shows an example of rice seedling annotations in the RSC dataset.

3. Methods

In this section, we introduce the details of our proposed Combined Network, including architecture, ground truth generation, and implementation details in terms of effectively addressing in-field counting of rice seedlings.

3.1. Combined Network for Counting Rice Seedlings

UAV images of rice seedlings are often captured from high viewpoints, resulting in highly dense scenarios (>10,000 seedlings). In such cases it is very difficult and time consuming to train an object detector to locate and count each rice seedling. In this work, we consider the problem of in-field counting of rice seedlings as an object counting task, and we follow the idea of counting by regressing the density map of a UAV image. The idea of regression counting is simple: given a UAV image I and a regression target T , the goal is to seek some kind of regression function F so that $T \approx F(I)$, where T

represents a density map and the counts can be acquired by integrating over the entire density map. In the following subsections, we first describe our proposed basic count regression network (Basic Network) to regress the density map and number of rice seedlings for a given UAV image. Then an improved version, the Combined Network, is proposed to further improve counting accuracy.

Figure 1. An example of the rice seedling annotations in the rice seedling counting (RSC) dataset: (A) raw unmanned aerial vehicle (UAV) image, (B) manually-annotated cropped image, and (C) exact positions of the rice seedlings.

3.1.1. Basic Network Architecture

An overview of the proposed Basic Network method is shown in Figure 2. Its input is the UAV image, and the output is a density map describing the distribution of rice seedlings. It is worth mentioning that our Basic Network estimates the distribution of rice seedlings required for seedling counting using an architectural design similar to the well-known VGG-16 [36] network. The VGG-16 is a powerful network originally trained for the purpose of object classification in computer vision. Recently, an increasing number of studies have applied VGG-16 networks to other research domains, such as saliency prediction [37] and object segmentation [38], as the base feature extractor. However, an inevitable disadvantage of the original VGG-16 is that the number of parameters that need to be calculated is very large. For example, the convolutional layers of VGG-16 require 30.69 billion floating-point operations for a single pass over a single image at 224×224 resolution [25]. To reduce the number of floating-point operations and improve the efficiency of the model for rice seedling counting, we simplified the original VGG-16 network. Specifically, the number of convolution kernels per layer was greatly reduced, the stride of the fourth max-pool layer was set to 1, the fifth pooling layer was replaced by a 1×1 convolutional layer, and each convolutional layer is followed by a rectified linear unit nonlinearity (ReLU). In addition, considering potential image classification problems, rice seedling density estimation requires per-pixel predictions, where a single discrete label is assigned to an entire image. We obtain these pixel-level predictions by removing the fully connected layers present in the original VGG-16 architecture, thereby rendering our network fully convolutional. These modifications enable our VGG-16-like network to make predictions at $1/8$ times the input resolution. The corresponding density map can be obtained using the bilinear interpolation operation outlined in [26].

Figure 2. The proposed Basic Network method for counting rice seedlings.

3.1.2. Combined Network Architecture

In the previous sub-section, we introduced the Basic Network for counting rice seedlings. For a static image of rice seedlings captured by UAV in situ, count accuracy cannot be improved further unless we manually crop the rice seedling area in the image. This is because there is less information about non-rice seedlings in the training set, which diminishes the ability of the trained Basic Network to distinguish between rice seedlings and non-rice seedlings. Therefore, vegetation resembling rice seedlings (shrubs and weeds, etc.) is likely to be recognized as rice seedlings by the Basic Network, with density values incorrectly assigned to these areas. Moreover, in the density map obtained using the Basic Network, some other areas such as houses, roads, etc. are also assigned low density values, which further reduces the accuracy of counting. To solve this problem, this sub-section introduces an improved version (Combined Network) with a Segmentation Network to accurately estimate the number of rice seedlings.

The framework of the Combined Network method involves two fully convolutional neural networks and one combination operation as outlined in Figure 3. The first network, the Segmentation Network, is responsible for segmenting rice seedling areas and non-rice seedling areas in each UAV image and processing the segmentation results into a binary image. The second network is the Basic Network described above, which estimates the distribution of rice seedlings in each UAV image and generates a density map. Finally, the combination operation generates a final density map based on the density map returned by the Basic Network and the binary image returned by the Segmentation Network.

Figure 3. The proposed Combined Network method for counting rice seedlings.

In our Segmentation Network, we aim to extract areas containing rice seedlings from the input UAV images. Since the segmentation does not need to be refined to individual rice seedlings, we designed the Segmentation Network as a shallow layer with a depth of only 3 convolutional layers and a kernel size of 7×7 , followed by average pooling. To equalize the spatial resolution of this network's prediction and that of its Basic Network, as with the latter, we use bilinear interpolation to upsample it to the size of the input image. To facilitate the discussion below, we refer to the results of the Segmentation Network prediction as a segmentation map, and the pixel values in the segmentation map are called intensity values.

Another key step is extracting rice seedling areas from the segmentation map. In fact, there is a significant difference between the intensity values of rice seedling areas in the segmentation map and the intensity values of other regions. In other words, we can select the appropriate range of intensity values to extract the rice seedling areas. We will show later, in discussing our experiments, the effect of selecting the threshold range of intensity values on the accuracy of rice seedling counts and give an empirical threshold range $[LT, UT]$, where LT , UT represent the lower and upper limits of the intensity

threshold, respectively, and $UT > LT$. Once the threshold range $[LT, UT]$ is obtained, we can use it to construct a binary map BM . Each element bm_{ij} in BM satisfies Equation (1).

$$bm_{ij} = \begin{cases} 1 & \text{if } sm_{ij} \in [LT, UT] \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where sm_{ij} represents the intensity value of the i th row and the j th column in the segmentation map SM obtained from the Segmentation Network.

To eliminate the inaccurate density values distributed in the density map and accurately predict rice seedling density and the number of rice seedlings, we combined the density map obtained from the Basic Network with the binary map derived from the Segmentation Network to generate a final density map. The combination operation can be described by Equation (2).

$$FDM = DM \otimes BM \quad (2)$$

where FDM represents the final density map, DM represents the density map obtained from the Basic Network, BM represents the binary map calculated using Equation (1), and $\otimes$ represents the dot multiplication operation. In this way, the density value of the rice seedling area in the DM remains unchanged in the FDM , and the density value of the non-rice seedling area is assigned a value of zero, so it does not feature in the subsequent calculation. This facilitates accurate predictions of rice seedling density. Furthermore, the number of rice seedlings can be obtained by accumulating all density values in the FDM . Since all the computations are executed using a graphics processing unit (GPU), the combination operation is very fast.

3.2. Ground Truth

In the Basic Network training process, training a fully convolutional neural network using the ground truth of rice seedling annotations (dots), would be onerous. Moreover, the position of rice seedling annotations is often ambiguous vis-à-vis the exact position of the in situ seedlings, and varies from annotator to annotator (bottom, center of the rice seedling etc.), making CNN training difficult. Inaccurate generation of ground truth will lead to calculation error in the loss function, thus affecting the performance of the final model. Thus, generation of the ground truth is an important step in the training process. In [26,27,35], the ground truth is generated by simply blurring each annotation using a two-dimensional Gaussian kernel normalized to sum to one. Inspired by this, herein we generate an $n \times n$ square matrix that satisfies the Gaussian distribution around each annotation coordinate and normalizes it to sum to one. This operation causes the sum of the density map to coincide with the total number of rice seedlings in a UAV image. Preparing the ground truth in this way makes it easier for CNN to learn the ground truth, because CNN no longer needs to obtain the exact point of rice seedling annotation. It also provides information about which areas contribute to the count, and by how much. This helps in training the CNN to correctly predict both rice seedling density as well as rice seedling count. Figure 4 shows an example of ground truth density given corresponding dotted annotations. Similar to [26,27,35], the summation of the ground truth density is a decimal. This is because when dots are close to the image boundary, their Gaussian probability will be partly outside the image, but this definition naturally takes a fraction of an object into account.

Figure 4. Example of a ground truth density map: manually-annotated cropped image (**left**) and corresponding ground truth rice seedling density map (**right**). Density map regression treats the two-dimensional density map as the regression target. Our proposed Basic Network regresses the global count computed from the global density map.

Instead of the Basic Network training process, we opted to train the Segmentation Network using UAV images that contained a large amount of non-rice seedling areas. We sampled about 1000 such sub-images from the RSC dataset and annotated for the rice seedling area in each image. To more easily separate the rice seedling area from other areas and capture the specific location of the rice seedling area, we used the binary map as the ground truth of the Segmentation Network. The generated rule is defined in Equation (3), where D represents the seedling annotation area and f_{ij} represents the value in the i th row and the j th column of the ground truth. Figure 5 shows an example of ground truth for a given rice seedling area (white) and a non-rice seedling area (black).

$$f_{ij} = \begin{cases} 255 & \text{if pixel } p_{ij} \text{ is in } D \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

Figure 5. Example of the area of rice seedling annotations: (A) cropped UAV image, (B) manually-annotated rice seedling area image, and (C) its corresponding ground truth binary map. Binary map regression treats the two-dimensional binary map as the regression target in our proposed Segmentation Network.

3.3. Learning and Implementation

Regression network learning should be driven by a loss function. In this paper, both the Basic Network and Segmentation Network involved in the Combined Network are trained by back propagating the squared loss function computed with respect to the ground truth. The squared loss function is defined in Equation (4):

$$L = \sum_i^W \sum_j^H (v_{ij}^{pred} - v_{ij})^2 \quad (4)$$

where W and H denote the width and height of the training image, respectively, and v_{ij}^{pred} and v_{ij} represent the density value (intensity value) in the i th row and the j th column of the density map (segmentation map) and ground truth density (ground truth binary map), respectively. $v_{ij}^{pred} - v_{ij}$ is the residual that measures the difference between the predicted density value (intensity value) and the ground truth density value (ground truth binary value) for the input image.

We use the RSC dataset to train our Basic Network. The RSC dataset contains 40 RGB UAV images with a resolution of 5472×3648 pixels and rice seedling annotations. In a manner similar to recent works [26,27], we evaluate the performance of our Basic Network using multi-fold cross validation, which randomly divides the dataset into four splits with each split containing ten images. In each fold of the cross validation, we consider three splits (30 images) as the training dataset and the remaining split (ten images) as the test dataset to explore its performance. Note that the test dataset in each fold is also used to evaluate the performance of the Combined Network. As CNNs require a large amount of training data, we crop 512×512 sub-images with a stride of 256 from each of the 30 training images for extensive augmentation of the training dataset. This augmentation yields an average of 8000 training sub-images in each fold of the cross validation. Moreover, we also perform a random shuffling of these sub-images. The training model (Basic Network) per fold is saved to disk, which can be called using the TensorFlow [39] Python interface in the testing process.

During Segmented Network training, we cropped approximately 1000 sub-images of size 512×512 pixels from the RSC dataset. These sub-images contain rice seedling areas and non-rice seedling areas. We use 90% of the sub-images for training and the remainder for validation. The trained Segmented Network model is also saved to disk.

In the testing stage, our Combined Network amalgamates the predictions of the Basic and Segmentation networks and generates a final density map that describes the distribution of rice seedlings. However, we observed that it is difficult for CNN to directly process high-resolution UAV images in the test dataset. To overcome this, we perform a similar cropping operation on each image in each test dataset. The crop size is still set to 512×512 . To avoid sub-image overlap affecting counting accuracy, we set the stride to 512. In addition, for sub-images smaller than 512×512 , we supplement them with black pixels. Then, the number of seedlings per raw UAV image in the test dataset can be calculated by Equation (5):

$$N_i = \sum_{k=1}^n p_k \quad (5)$$

where i is the index number of the UAV image in the test dataset, n represents the number of sub-images cropped by the i -th UAV image, and p_k denotes the number of seedlings obtained by accumulating the density value of the k -th sub-image.

We implemented all operations in the Combined Network described above in the TensorFlow deep learning framework [39]. Of these, the Basic Network and Segmentation Network involved in the Combined Network are trained using stochastic gradient descent (SGD) optimization with learning rates of $1e-5$ for 400 and $1e-4$ for 50 iterations, respectively. In each SGD iteration, we uniformly sample 100 sub-images to construct a mini-batch. The streamlined Combined Network architecture does not require the calculation of a large number of parameters, so the average training time per fold

is about 1.2 hours, and the test can be completed in 1s (Python 3.5.4, IDE: Pycharm 2017, OS: Ubuntu 18.04 64-bit, CPU: Intel i7-6800K 3.40GHz, GPU: Nvidia GeForce GTX 1080Ti, RAM: 16 GB).

4. Experimental Evaluation

To explore the efficacy of our proposed method for counting rice seedlings in situ, a series of experiments was conducted on the challenging RSC dataset comprised of UAV images. Following the multi-fold cross validation setup in [26,27], UAV images were randomly allocated to one of four splits. In each fold of the cross validation, we repeated the experiment five times. The average mean absolute error, accuracy, R^2 value, and root mean squared error were used as evaluation metrics.

4.1. Evaluation Metrics

The mean absolute error (MAE), accuracy (Acc), R^2 value, and the root mean squared error (RMSE) were used as evaluation metrics to assess counting performance (Equations (6)–(9)):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |t_i - c_i| \quad (6)$$

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

$$r^2 = 1 - \frac{\sum_{i=1}^n (t_i - c_i)^2}{\sum_{i=1}^n (t_i - \bar{t}_i)^2} \quad (8)$$

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

where t_i , $\bar{t}_i$, and c_i represent the ground truth count for the i -th image, the mean ground truth count, and the predicted count for the i -th image (computed by accumulating over the whole density map), respectively; n denotes the number of UAV images in the test set; MAE and Acc quantify prediction accuracy; R^2 and RMSE assess the performance of our trained model. The lower the value of MAE and RMSE is, the better the counting performance, while the higher the value of Acc and R^2 is, the better the counting performance.

4.2. Evaluating the Efficacy of the Proposed Basic Network

In this section, we evaluate the counting capability of the proposed Basic Network method. Table 1 shows the Acc and MAE obtained from each fold. The results shown do not include any post-processing methods. According to the results, we observed that the average counting accuracy of our Basic Network exceeded 80%, and the average MAE of the counting was approximately 1600. Note that the number of rice seedlings in most of the images in the test dataset exceeded 10,000. Therefore, the average MAE value for this experiment was within an acceptable range. However, we also observed that results derived from the Basic Network fluctuated substantively in different folds. This is because the image data in each fold is different, for example, some test images in fold-1 and fold-3 contain a large number of non-rice areas. Unfortunately, the Basic Network does not have functionality for segmenting rice seedling areas and non-rice seedling areas, which negatively impacts on the accuracy of this method. Based on this, we made some improvements to the Basic Network which are described in the next section.

In Figure 6 we present the predicted counts for each image in the dataset along with their ground truth counts. For almost all images, the Basic Network tended to overestimate the count. This overestimation could possibly be a function of the presence of other ‘objects’ such as shrubs or roads which are incorrectly assigned a small amount of density values by the Basic Network.

Table 1. Performance of the Basic Network on the RSC dataset.

	ACC	MAE
Fold-1	78.74%	1838.965
Fold-2	83.51%	1500.078
Fold-3	74.05%	2161.408
Fold-4	91.47%	923.5362
Average	81.94%	1605.997

Figure 6. Predicted count obtained from the Basic Network model versus the ground truth count for each of the 40 images in the RSC dataset.

Next, a linear regression analysis was executed between the number of detected rice seedlings according to manual counting (ground truth) and automatic counting derived from the Basic Network for all images in the RSC dataset. The Basic Network model (Figure 7) performed well in terms of estimating the number of rice seedlings. The relationship between manual and automatic rice seedling counts was positive and strong for all images with R^2 and RMSE of 0.91 and 1916.2, respectively. This indicates that the Basic Network model can reasonably estimate the number of in-field rice seedlings.

Figure 7. Manual counting versus automatic counting using our Basic Network model. The 1:1 line is dashed red. R^2 and RMSE represent the coefficient of determination and the root mean square error, respectively.

4.3. Evaluating the Efficacy of the Proposed Combined Network

Although reasonable count estimates can be achieved by the Basic Network, its accuracy was unduly influenced by non-rice seedling areas such as shrubs, roads, etc. To solve this problem, we designed a Combined Network which involved adding a Segmentation Network to the Basic Network. In the process of counting, the Combined Network filters incorrect density values distributed in the non-rice seedling areas through the Segmentation Network, which improves counting accuracy. The advantages of the proposed Combined Network for rice seedling counting are detailed in this section.

We first studied the impact of different intensity value threshold ranges $[LT, UT]$ mentioned in Section 3.1.2 on the performance of the Combined Network (Figure 8). In this experiment, LT and UT ranged from 50 to 300 in steps of 10, and Acc was used as the evaluation metric to assess counting performance. We have refrained from reporting trivial results, i.e., $LT \geq UT$. We observed that the counting performance for each test sequence was different, which suggests that the Combined Network is sensitive to specification of the intensity value threshold range. According to the results, we also observed that the Combined Network has good counting performance when $LT < 100$ and when $UT > 200$. This is perhaps because the intensity values of rice seedling areas output by the Segmentation Network are distributed between 100 and 200. Intensity values beyond this range may signify other objects, for example, our experimental results showed that areas with intensity values above 300 denoted roads. Without loss of generality, in the following experiment the LT and UT were fixed at 90 and 265, respectively. Figure 9 shows an example of the results (segmentation map) from the Segmentation Network, and a binary map within the fixed intensity value threshold range ($LT = 90$, $UT = 265$). The results show that the Segmentation Network can reasonably approximate the rice seedling area. The binary map was obtained by Equation (1) and used to filter out those density values that were incorrectly assigned by the Basic Network.

Figure 8. Performance of the Combined Network model under different intensity value threshold ranges. The X axis is the lower limit of the intensity value threshold range, the Y axis is the upper limit of the intensity value threshold range, and the Z axis is the accuracy of our proposed Combined Network for counting rice seedlings.

The proposed Combined Network in this study was evaluated on our RSC dataset and compared to the counting performance of our Basic Network (Table 2). We observed that the Combined Network significantly outperformed the Basic Network. Specifically, the average accuracy of the Combined Network was 11% higher than that of the previous Basic Network—achieving an Acc of 93.35%. Furthermore, the average MAE of the Combined Network was less than 700, which is much smaller than the average MAE obtained by the Basic Network. In each fold of the cross validation, the counting performance of the Combined Network was more stable compared to the Basic Network; the Acc always fell between 91% and 95%, and the MAE was maintained at between 600 and 800. This indicates that the Combined Network is more effective than the Basic Network for the purposes of accurately estimating the number of in-field rice seedlings.

Figure 9. Segmentation map (B) predicted by the Segmentation Network on a cropped UAV image (A) and the corresponding binary map (C) calculated by Equation (1). The threshold range is fixed at [90, 265].

Table 2. Comparing the performance of the Basic Network and Combined Network for counting rice seedlings in the RSC dataset.

	ACC		MAE	
	Basic Network	Combined Network	Basic Network	Combined Network
Fold-1	78.74%	91.91%	1838.965	772.3573
Fold-2	83.51%	94.07%	1500.078	623.1984
Fold-3	74.05%	92.76%	2161.408	754.4838
Fold-4	91.47%	94.68%	923.5362	638.9999
Average	81.94%	93.35%	1605.997	697.2598

Figure 10 shows the results of three counts for each image in our RSC dataset, including the predicted counts obtained by the Combined Network, the predicted counts obtained by the Basic Network, and the ground truth counts which served as the baseline for the comparison. In most cases, the count predicted by the Combined Network was closer to the ground truth count than the count predicted by the Basic Network. However, we observed that in rare cases, the predicted count of the Combined Network was not as accurate as the Basic Network. For example, for the fifth image with a ground truth count of 14,994, the predicted count from the Combined Network was 14,244 with a MAE of 770, while the predicted count from the Basic Network was 15,541 with a MAE of only 547. Even so, the accuracy of the Combined Network in predicting the number of seedlings in the fifth image was still over 94%. Therefore, we suggest that the effectiveness of the Combined Network will be acceptable in most, if not all, real-world applications.

Figure 10. Predicted count obtained from the Combined Network and Basic Network models versus the ground truth count for each of the 40 images in the RSC dataset.

Figure 11 depicts the regression relationship between the ground truth count and the automatic count derived from the Combined Network. Compared to the Basic Network (Figure 7), the Combined Network model performed better at estimating the number of rice seedlings with R^2 and RMSE of 0.94 and 906.8813, respectively. We can also observe from Figure 11 that the black regression line converges closely with the 1:1 dashed red line, which testifies once again that our Combined Network model can effectively solve the problem of in-field counting of rice seedlings.

Figure 11. Manual counting versus automatic counting with our Combined Network model. The 1:1 line is dashed red. R^2 and RMSE represent the coefficient of determination and the root mean square error, respectively.

In sum, for a given image taken by a UAV, we can estimate the number of rice seedlings therein to a high degree of accuracy. Nevertheless, we have not explored the distribution of errors based on our proposed methods. An interesting question is “compared with the ground truth count, is our method overestimating or underestimating the number of seedlings?” or “in a single estimate, how many seedlings are overestimated (underestimated)?” Figure 12 presents histograms of count errors, where the height of each of blue (red) bar represents the frequency of occurrence of the proposed Basic Network (Combined Network) within a particular error range. The results suggest that overestimations occurred significantly more frequently in the Basic Network compared to the Combined Network, and it was common for the former network to overestimate by more than 1000 seedlings. The count error distribution in the Combined Network was relatively uniform, and substantive underestimation or overestimation (relative count difference < -1000 or > 1000) did not occur frequently. These results

further confirm that the Combined Network model, which amalgamates the Basic and Segmentation Networks, achieves comparable or better results compared to the Basic Network model in isolation.

Figure 12. Distribution of count errors for the Basic Network and Combined Network.

5. Discussion

5.1. Performance of the Segmentation Network

Herein, the Segmentation Network was crucial for automatic counting of large-area rice seedlings. As shown in Table 2 and Figure 10, the Segmentation Network played a significant role in improving the accuracy of seedling number estimates. This is because adding a Segmentation Network can reduce the occurrence of incorrectly assigned density values in the final density map. However, the Segmentation Network was highly sensitive to the intensity value threshold range, which is consistent with the results shown in Figure 8. Based on the experimental results, we put forward an empirical threshold range $[LT, UT]$ for segmentation of rice and non-rice seedling areas. In most cases, the segmented results were similar to the ground truth segmentation maps. However, there were also some circumstances where the Segmentation Network did not yield an accurate segmentation in the empirical threshold range $[LT, UT]$. The results in Figure 13 show two cases of failure in this respect: (1) when dark areas such as shadows appeared in the image, the Segmentation Network erroneously recognizes these areas as rice areas; (2) Recognizing weeds in images was also beyond the ability of the Segmentation Network. To navigate these issues, one may consider employing additional training data that contains non-rice seedling areas. Alternatively, since failures of the Segmentation Network to extract rich features could be a function of its simple and shallow structure, it may be possible to use more complex network structures such as VGG-16 [36], GoogleNet [40] and ResNets [41] to extract rice regions more accurately. Fortunately, only a small proportion of the sub-image data in our RSC dataset could not be segmented by the Segmentation Network, thus the impact on final count estimates was trivial. Overall, we believe that our Segmentation Network is an effective yet simple auxiliary method for improving the accuracy of rice seedling counts in the RSC dataset.

Figure 13. Two examples of erroneous predictions obtained from the Segmentation Network: (A) cropped UAV images, where the upper image contains roads, shadows, and rice seedling areas, while the lower image contains weeds and rice seedling areas; (B) the corresponding ground truth binary maps; (C) segmentation maps predicted by the Segmentation Network; (D) the corresponding binary maps generated by Equation (1) based on the empirically determined intensity value threshold range.

5.2. Analysis of Predicted Distribution Maps

Rice seedling density or count is frequently identified as the main agronomical component of yield and it appears to be the most relevant pursuant of future increases in grain yield. In this study, we consider the problem of in-field counting of rice seedlings as an object counting task. An effective deep CNN-based solution was presented to effectively count rice seedlings. Extensive experiments showed that our proposed approach was particularly suitable for the following scenarios: (1) large-scale counting exercises; (2) extremely crowded scenes; (3) counting problems where ground truth density maps cannot be precisely defined. However, our method does not provide specific location information for each rice seedling in the UAV images. Figure 14 shows an example of a predicted rice seedling distribution map (C) extracted from the final density map using the k-NN algorithm and its corresponding ground truth seedling distribution map (B). It clearly shows that our method can accurately estimate the distribution and number of rice seedlings, but that estimates of the location of each rice seedling are not accurate. The main reason for this is that we did not count the number of instances of an object in a given image from the perspective of training an object detector, but rather considered the counting problem from the perspective of training an object counter. Of course, we believe that the approaches considered from the perspective of the former, such as Fast R-CNN [22], Faster R-CNN [23], YOLO [24], YOLO9000 [25], etc., can also be applied to object counting problems. However, how to balance the efficiency and effectiveness of these approaches is a key challenge for high-density object counting in UAV images and is fruitful terrain for future research.

Figure 14. Examples of predicted rice seedling distribution maps extracted from the final density map: (A) cropped UAV images where rice seedlings are unevenly (upper image) and evenly (lower image) distributed; (B) ground truth rice seedling distribution map where the points represent rice seedlings; (C) predicted rice seedling distribution map extracted from the final density map using the k-NN algorithm.

5.3. Comparison with Other Techniques

Table 3 presents a comparison of our results with the Count Crops tool [42] in ENVI-5.5 software. This is a tool which aims at locating crops and estimating their number in high-resolution, single-band imagery. Note that we compared our method to the two scenarios: (1) without pre-processing of the image to remove all non-rice seedling regions (abbreviated in Table 3 as “without pre-processing”); (2) pre-processing of the image to remove all non-rice seedling regions (abbreviated in Table 3 as “pre-processing”). Additionally, the Count Crops tool in ENVI-5.5 software only supports single-channel images, so we counted each RGB image three times, i.e., on the R channel, G channel, and B channel.

Table 3. Comparing the performance of the Combined Network and the ENVI-5.5 Crop Science platform for counting rice seedlings in the RSC dataset.

Methods		MAE	Acc	Times (s)
Count Crops tool (without pre-processing)	Channel R	1371.675	87.29%	40.93
	Channel G	2362.125	77.54%	41.26
	Channel B	1595.35	84.63%	41.04
Count Crops tool (pre-processing)	Channel R	1001.325	90.71%	27.57
	Channel G	1901.925	81.76%	27.49
	Channel B	1204.375	88.16%	27.73
Our proposed Combined Network method		697.2598	93.35%	0.93

As can be seen from Table 3, the Count Crops tool had good counting performance, and the counting accuracy reached 90% under certain conditions. However, the main inconvenience of the

Count Crops tool is that it requires manual intervention, such as setting parameters and extracting regions of interest. In contrast, our proposed combined network approach is not only superior in calculating MAE and accuracy, but also much faster without manual intervention. However, the method we proposed also has shortcomings, and requires a lot of manual annotation and iterative training in the early stage. As we all know, this is a very tedious work. As a summary of our evaluations, we suggest the following good practices for rice-seedling-like in-field counting problems: (1) due to time requirements, use of the Count Crops tool for crop counting is not a good choice when the number of images is large; (2) when the number of images is insufficient, use of our proposed method for crop counting is not suitable because too few training samples can easily lead to low precision of the prediction model.

6. Conclusions and Future Work

In this paper, we studied the problem of counting rice seedlings in the field and formulated the problem as an object counting task. A tailored RSC dataset with 40 field UAV images containing a large number of rice seedlings and corresponding manually-dotted annotations was constructed. An effective deep CNN-based image processing method was also presented to estimate total rice seedling counts and rice seedling densities from high-resolution images captured in situ by UAVs. A series of experiments using the RSC dataset was conducted to explore the efficacy of our proposed method. The results demonstrated high average accuracy (higher than 93%) between manual and automated measurements, and very good performance, with a high coefficient of determination (R^2) (around 0.94), which would potentially facilitate rice breeding or rice phenotyping in future.

Several interesting directions could be explored by future research in this domain. First, efforts may be made to deploy our method into an embedded system on a UAV for online yield estimation and precision agriculture applications. Second, since training data are always the key to good performance, especially the diversity of such data, it could be interesting to continue to enrich the RSC dataset as a precursor to further modeling work. Third, the precise location of rice seedlings and the feasibility of improving counting performance from the perspective of object detection requires exploring, because the precise location of crops is a key step in precision agriculture. Finally, the problem of object counting in large-scale and high-density environments is still an open issue, with plenty of scope for exploring the feasibility of improving counting performance in this context.

Author Contributions: Conceptualization, J.W. and G.Y.; Data curation, B.X.; Formal analysis, J.W.; Funding acquisition, G.Y. and X.Y.; Investigation, B.X. and Y.Z.; Methodology, J.W.; Project administration, G.Y.; Resources, L.H.; Software, X.Y.; Supervision, G.Y.; Validation, B.X.; Visualization, L.H.; Writing—original draft, J.W.; Writing—review & editing, G.Y.

Funding: This study was supported by the National Key Research and Development Program of China (2017YFE0122500 and 2016YFD020060306), the Natural Science Foundation of China (41771469 and 41571323), the Beijing Natural Science Foundation (6182011), and the Beijing Academy of Agriculture and Forestry Sciences (KJCX20170423).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, Q. Inaugural article: Strategies for developing green super rice. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 16402–16409. [[CrossRef](#)] [[PubMed](#)]
2. Lu, H.; Cao, Z.; Xiao, Y.; Fang, Z.; Zhu, Y.; Xian, K. Fine-grained maize tassel trait characterization with multi-view representations. *Comput. Electron. Agric.* **2015**, *118*, 143–158. [[CrossRef](#)]
3. Chen, J.; Gao, H.; Zheng, X.M.; Jin, M.; Weng, J.F.; Ma, J.; Ren, Y.; Zhou, K.; Wang, Q.; Wang, J.; et al. An evolutionarily conserved gene, FUWA, plays a role in determining panicle architecture, grain shape and grain weight in rice. *Plant J.* **2015**, *83*, 427–438. [[CrossRef](#)] [[PubMed](#)]
4. Li, W.; Niu, Z.; Chen, H.; Li, D.; Wu, M.; Zhao, W. Remote estimation of canopy height and aboveground biomass of maize using high-resolution stereo images from a low-cost unmanned aerial vehicle system. *Ecol. Indic.* **2016**, *67*, 637–648. [[CrossRef](#)]

5. Verger, A.; Vigneau, N.; Chéron, C.; Gilliot, J.M.; Comar, A.; Baret, F. Green area index from an unmanned aerial system over wheat and rapeseed crops. *Remote Sens. Environ.* **2014**, *152*, 654–664. [[CrossRef](#)]
6. Gao, J.; Liao, W.; Nuytens, D.; Lootens, P.; Vangeyer, J.; Pižurica, A.; He, Y.; Pieters, J.G. Fusion of pixel and object-based features for weed mapping using unmanned aerial vehicle imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *67*, 43–53. [[CrossRef](#)]
7. Wan, L.; Li, Y.; Cen, H.; Zhu, J.; Yin, W.; Wu, W.; Zhu, H.; Sun, D.; Zhou, W.; He, Y. Combining UAV-Based Vegetation Indices and Image Classification to Estimate Flower Number in Oilseed Rape. *Remote Sens.* **2018**, *10*, 1484. [[CrossRef](#)]
8. Berni, J.; Zarco-Tejada, P.J.; Suarez, L.; Fereres, E. Thermal and Narrowband Multispectral Remote Sensing for Vegetation Monitoring from an Unmanned Aerial Vehicle. *Inst. Electr. Electron. Eng.* **2009**, *47*, 722–738. [[CrossRef](#)]
9. Uto, K.; Seki, H.; Saito, G.; Kosugi, Y. Characterization of Rice Paddies by a UAV-Mounted Miniature Hyperspectral Sensor System. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 851–860. [[CrossRef](#)]
10. Püschel, H.; Sauerbier, M.; Eisenbeiss, H. A 3D Model of Castle Landenberg (CH) from Combined Photogrammetric Processing of Terrestrial and UAV-based Images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2008**, *37*, 93–98.
11. Moranduzzo, T.; Mekhalfi, M.L.; Melgani, F. LBP-based multiclass classification method for UAV imagery. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 2362–2365.
12. Lin, A.Y.-M.; Novo, A.; Har-Noy, S.; Ricklin, N.D.; Stamatiou, K. Combining GeoEye-1 Satellite Remote Sensing, UAV Aerial Imaging, and Geophysical Surveys in Anomaly Detection Applied to Archaeology. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *4*, 870–876. [[CrossRef](#)]
13. Moranduzzo, T.; Melgani, F. A SIFT-SVM method for detecting cars in UAV images. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 6868–6871.
14. Moranduzzo, T.; Melgani, F. Automatic Car Counting Method for Unmanned Aerial Vehicle Images. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 1635–1647. [[CrossRef](#)]
15. Moranduzzo, T.; Melgani, F. Detecting Cars in UAV Images with a Catalog-Based Approach. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6356–6367. [[CrossRef](#)]
16. Moranduzzo, T.; Melgani, F.; Bazi, Y.; Alajlan, N. A fast object detector based on high-order gradients and Gaussian process regression for UAV images. *Remote Sens.* **2015**, *36*, 37–41.
17. Fernandez-Gallego, J.A.; Kefauver, S.C.; Gutiérrez, N.A.; Nieto-Taladriz, M.T.; Araus, J.L. Wheat ear counting in-field conditions: High throughput and low-cost approach using RGB images. *Plant Methods* **2018**, *14*, 22–34. [[CrossRef](#)] [[PubMed](#)]
18. Zhou, C.; Liang, D.; Yang, X.; Yang, H.; Yue, J.; Yang, G. Wheat Ears Counting in Field Conditions Based on Multi-Feature Optimization and TWSVM. *Front. Plant Sci.* **2018**, *9*, 1024–1040. [[CrossRef](#)] [[PubMed](#)]
19. Guo, W.; Zheng, B.; Potgieter, A.B.; Diot, J.; Watanabe, K.; Noshita, K.; Jordan, D.R.; Wang, X.; Watson, J.; Ninomiya, S.; et al. Aerial Imagery Analysis—Quantifying Appearance and Number of Sorghum Heads for Applications in Breeding and Agronomy. *Front. Plant Sci.* **2018**, *9*, 1544–1553. [[CrossRef](#)] [[PubMed](#)]
20. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 24–27 June 2014; pp. 580–587.
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 1904–1916. [[CrossRef](#)] [[PubMed](#)]
22. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 13–16 December 2015; pp. 1440–1448.
23. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *39*, 91–99. [[CrossRef](#)] [[PubMed](#)]
24. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788.
25. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 22–25 July 2017; pp. 6517–6525.

26. Boominathan, L.; Kruthiventi, S.S.; Babu, R.V. Crowdnet: A deep convolutional network for dense crowd counting. In Proceedings of the 2016 ACM on Multimedia Conference (ACMMM), Amsterdam, The Netherlands, 15–19 October 2016; pp. 640–644.
27. Zhang, C.; Li, H.; Wang, X.; Yang, X. Cross-scene crowd counting via deep convolutional neural networks. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 8–10 June 2015; pp. 2980–2988.
28. Sam, D.B.; Surya, S.; Babu, R.V. Switching convolutional neural network for crowd counting. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 22–25 July 2017; pp. 7263–7271.
29. Lempitsky, V.; Zisserman, A. Learning to count objects in images. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Vancouver, BC, Canada, 6–11 December 2010; pp. 1324–1332.
30. Xie, W.; Noble, J.A.; Zisserman, A. Microscopy cell counting and detection with fully convolutional regression networks. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **2018**, *6*, 283–292. [[CrossRef](#)]
31. Onoro-Rubio, D.; López-Sastre, R.J. Towards perspective-free object counting with deep learning. In Proceedings of the 13th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016; pp. 615–629.
32. Arteta, C.; Lempitsky, V.; Noble, J.A.; Zisserman, A. Interactive object counting. In Proceedings of the 13th European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; pp. 504–518.
33. Arteta, C.; Lempitsky, V.; Zisserman, A. Counting in the wild. In Proceedings of the 13th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016; pp. 615–629.
34. Rahnemounfar, M.; Sheppard, C. Deep count: Fruit counting based on deep simulated learning. *Sensors* **2017**, *17*, 905. [[CrossRef](#)] [[PubMed](#)]
35. Lu, H.; Cao, Z.; Xiao, Y.; Zhuang, B.; Shen, C. TasselNet: Counting maize tassels in the wild via local counts regression network. *Plant Methods* **2017**, *13*, 79–96. [[CrossRef](#)] [[PubMed](#)]
36. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 2015 International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015; pp. 1556–1570.
37. Pan, J.; Sayrol, E.; Giro-i-Nieto, X.; McGuinness, K.; O'Connor, N.E. Shallow and deep convolutional networks for saliency prediction. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 598–606.
38. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 8–10 June 2015; pp. 3431–3440.
39. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
40. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 8–10 June 2015; pp. 1–9.
41. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
42. ENVI Crop Science 1.1. Crop Counting and Metrics Tutorial. 2018. Available online: <http://www.harrisgeospatial.com/portals/0/pdfs/envi/TutorialCropCountingMetrics.pdf> (accessed on 9 March 2019).