



Article

Research on Rapeseed Seedling Counting Based on an Improved Density Estimation Method

Qi Wang, Chunpeng Li, Lili Huang, Liqing Chen , Quan Zheng and Lichao Liu * 

College of Engineering, Anhui Agricultural University, Hefei 230036, China; 22720296@ahau.edu.cn (Q.W.); 21721247@stu.ahau.edu.cn (C.L.); hll@ahau.edu.cn (L.H.); lqchen@ahau.edu.cn (L.C.); zhengquan@ahau.edu.cn (Q.Z.)

* Correspondence: llchao@ahau.edu.cn

Abstract: The identification of seedling numbers is directly related to the acquisition of seedling information, such as survival rate and emergence rate. It indirectly affects detection efficiency and yield evaluation. Manual counting methods are time-consuming and laborious, and the accuracy is not high in complex backgrounds or high-density environments. It is challenging to achieve improved results using traditional target detection methods and improved methods. Therefore, this paper adopted the density estimation method and improved the population density counting network to obtain the rapeseed seedling counting network named BCNet. BCNet uses spatial attention and channel attention modules and enhances feature information and concatenation to improve the expressiveness of the entire feature map. In addition, BCNet uses a 1×1 convolutional layer for additional feature extraction and introduces the torch.abs function at the network output port. In this study, distribution experiments and seedling prediction were conducted. The results indicate that BCNet exhibits the smallest counting error compared to the CSRNet and the Bayesian algorithm. The MAE and MSE reach 3.40 and 4.99, respectively, with the highest counting accuracy. The distribution experiment and seedling prediction showed that, compared with the other density maps, the density response points corresponding to the characteristics of the seedling region were more prominent. The predicted number of the BCNet algorithm was closer to the actual number, verifying the feasibility of the improved method. This could provide a reference for the identification and counting of rapeseed seedlings.



Citation: Wang, Q.; Li, C.; Huang, L.; Chen, L.; Zheng, Q.; Liu, L. Research on Rapeseed Seedling Counting Based on an Improved Density Estimation Method. *Agriculture* **2024**, *14*, 783. <https://doi.org/10.3390/agriculture14050783>

Academic Editor: Jiehao Li

Received: 10 April 2024

Revised: 10 May 2024

Accepted: 11 May 2024

Published: 19 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: density estimation; rapeseed seedling; counting; density map

1. Introduction

The rape plant is a typical crop in China [1], and its high-quality production is closely related to the development of the agricultural economy. The seedling is an important stage in the growth period of rape, so seedling number identification is directly linked to obtaining essential information about seedlings, such as their sowing survival rate and emergence rate. This information indirectly impacts the detection efficiency and yield assessment [2]. At present, seedling counting mainly relies on manual observation. However, due to technical limitations of the seeding machine or improper manual operation, among other factors, uneven seed density can occur, leading to difficulties in manual observation. This method is time-consuming and consumes substantial manpower and material resources [3]. Therefore, it is necessary to adopt appropriate methods to replace manual observation. With the continuous progress of computer vision and the rapid development of agriculture, the relationship between agricultural production and computer vision is becoming increasingly close [4]. Based on the density estimation method [5], counting is based on learning the linear mapping between the target features and the corresponding density map, thereby integrating spatial information into the learning process [6]. Meanwhile, the counting accuracy is greatly improved by the powerful feature expression capability of convolutional neural networks [7].

At present, density estimation methods have been partially applied in the agricultural field and have obtained improved performance. Qi Yang et al. [8] proposed an effective cotton counting method based on feature fusion. The comparison test results with the target counting methods MCNN, CSRNet, TasselNet, and MobileCount showed that the MAE and RMSE of the algorithm are 63.46 and 81.33, respectively. Compared with the comparison method, the average MAE and RMSE decrease by 48.8% and 45.3%, respectively. Huang Ziyun et al. [9] utilized the density class classification method to count cotton bolls in the field. They combined the classification information with the features to generate a high-quality density map, which effectively improved the accuracy of counting cotton bolls in the field. Bao Wenxia et al. [10] first equalized and segmented the field wheat spike images. They trained the field wheat spike density map estimation model through transfer learning to achieve the estimation of the number of wheat spikes in the field during the irrigation period. Lu et al. [11] determined the total number of cornhusks in the image by dividing the image into blocks, calculating local counts from the block density map, and then merged and normalized them to obtain the total number of cornhusks in the image. However, the density estimation method has not been applied to rapeseed seedlings, and the common detection method for rapeseed seedlings is to use deep learning methods. The actual environment complexity of the research mentioned above is low. The crops can be differentiated, but the rapeseed seedlings background is complex. Additionally, there are high density regions caused by improper sowing operations. Occlusion, to a certain extent, can limit the counting ability of the density estimation method.

To sum up, within a high density environment, neither the traditional detection method nor the improved method yielded superior results in this study. Therefore, the density estimation method was employed, proving to be more effective in identifying rapeseed seedlings. Moreover, to address the challenge of counting rapeseed seedlings with severe occlusion in high-density regions, spatial attention and channel attention modules was incorporated to enhance and combine feature information from spatial and channel dimensions, respectively, to enhance the overall expression capability of the feature map. In order to extract more detailed features from the spliced attention features, a 1×1 convolution layer was also utilized for additional feature extraction. Finally, to more effectively constrain the distribution of the model parameters, the torch.abs function was implemented at the output layer of the network. This function was continuously trained and optimized to develop the rapeseed seedling counting model BCNet. The results showed that the counting error of BCNet was the smallest compared with the CSRNet and the Bayesian algorithm, and the MAE and MSE reached 3.40 and 4.99, respectively, indicating the highest counting accuracy. The density response points corresponding to the features in the seedling region of the improved density map are more prominent, and the predicted counts by the BCNet algorithm are closer to the actual counts.

2. Materials and Methods

2.1. Research Process

In this paper, a method based on density estimation is proposed to count rapeseed seedlings. During the training, the rapeseed seedling images in the training set are fed into the network for feature extraction to generate the predicted density map. At the same time, the manually labeled rapeseed seedling stage labeling data are used as supervised signals, and regression estimation is performed with the expected value of the predicted density map to calculate the loss. After training, the loss value gradually converges. During the test, the images of rapeseed seedling stage in the test set were input into the trained model to generate the predicted density map. Subsequently, the pixel values in the density map were summed to estimate the predicted quantity. Figure 1 shows the main flow of the algorithm.

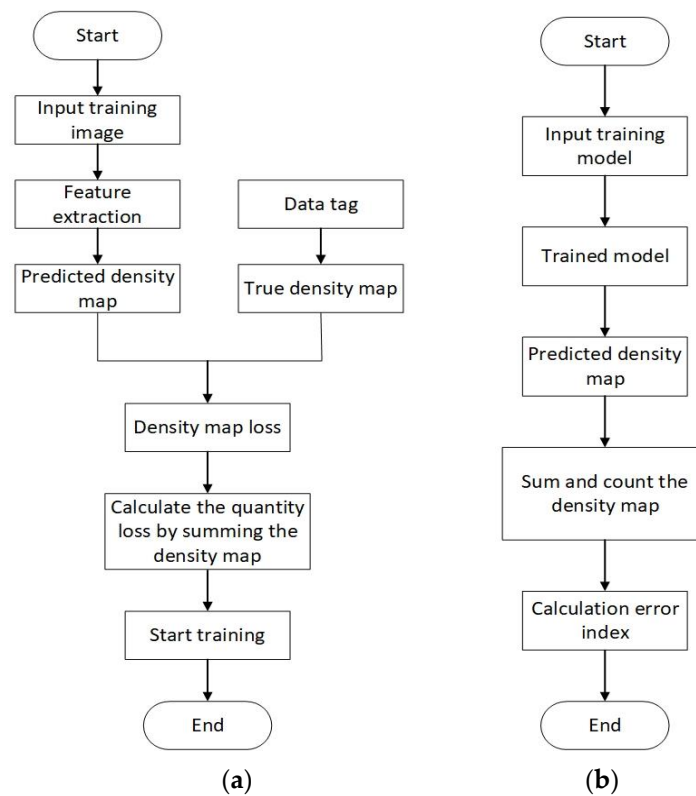


Figure 1. Flow chart of the oilseed rape plant counting algorithm with improved density estimation method. (a) Training flowchart. (b) Test flowchart.

2.2. Data Set Sample Collection and Production

2.2.1. Data Collection and Screening

The rapeseed seedling images were collected from the experimental field of rapeseed sowing in Hanshan County, Maanshan City, Anhui Province, the sowing method of rape was scattering, the sowing time was 20 October 2022, and the images were collected on 6 December 2022, using the DJI Royal 2 UAV as shown in Figure 2. A total of 600 images of the rapeseed seedling stage were carefully screened to eliminate fuzzy and invalid images.

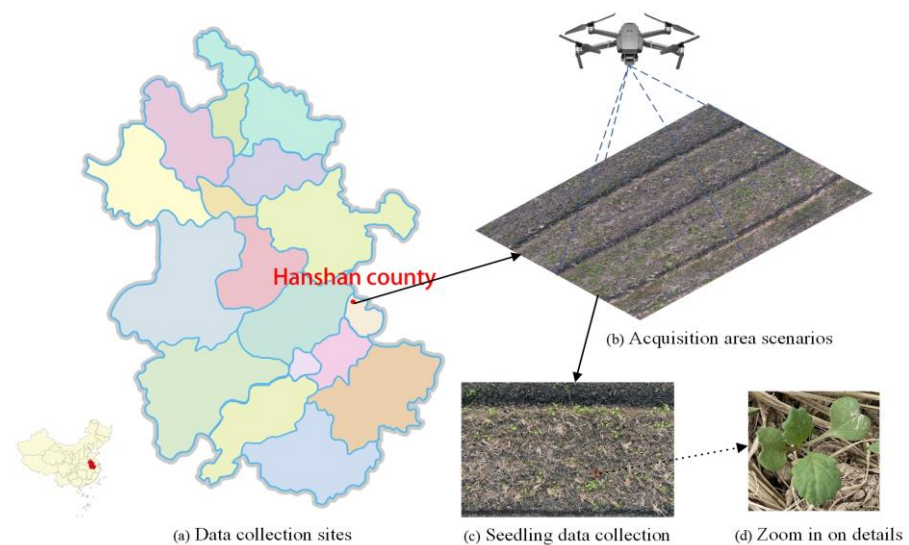


Figure 2. This is a figure. Schemes follow the same formatting. (a) Data collection sites. (b) Acquisition area scenarios. (c) Seedling data collection. (d) Zoom in on details.

2.2.2. Data Set Production

The rapeseed seedling dataset was labeled according to the population standard dataset format, using the labeling tool Stroller-spotter, which identified and labeled the locations of rapeseed plants. Each rapeseed seedling map, after labeling was completed, generated a corresponding mat label file. Each mat label file contains the total number of point labels and the coordinate information of each point label, representing the actual location of each rapeseed plant. The division of the training and test sets of the rapeseed seedling dataset was performed at the ratio of 3:1, with 450 training sets and 150 test sets. The schematic diagram of the labeling process is shown in Figure 3.

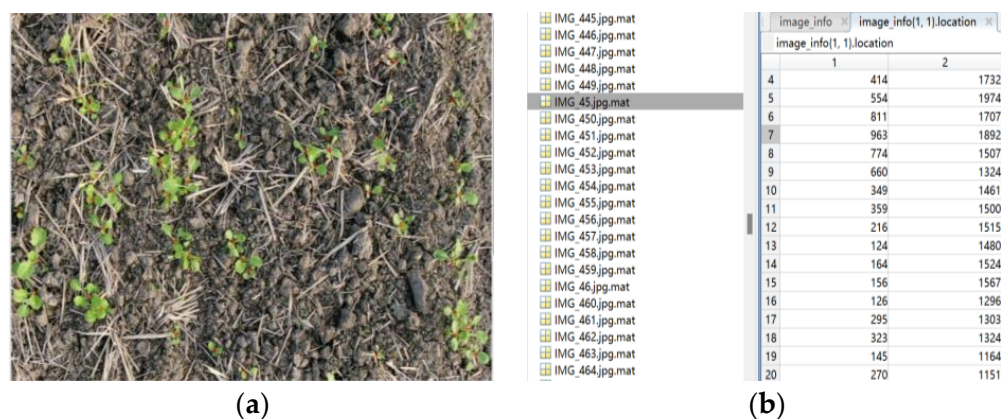


Figure 3. Schematic diagram of the labeling process. (a) Labeling schematic. (b) Mat tag file.

2.2.3. Information on Data Sets

Table 1 shows the pertinent information regarding the rapeseed plant dataset and the rape plants in all images. The minimum number of samples indicates the minimum number of single rapeseed plants in all images, and the maximum number of samples indicates the maximum number of single rape plants in any images. The overall distribution of data samples exhibits characteristics of diversification.

Table 1. Information about the rapeseed seedling dataset.

Image Number	Minimum Number of Samples	Maximum Sample Size	Average Sample Size	Total Sample Size
600	45	170	81	48,600

3. Design and Training of Recognition and Counting Models

3.1. Methods Based on Density Estimation

The CSRnet network [12] is primarily utilized for counting in crowded scenes. It can accurately count and generate high-quality density maps in highly crowded scenes. The CSRnet network model is mainly divided into front-end and back-end networks. VGG16 is utilized as the front-end network, and the size of the output image is 1/8th of the original input image. An increase in the number of convolutional layers will result in a smaller output image size, making it challenging to generate density maps. Therefore, this paper utilizes a dilated convolutional neural network as the back-end network. The increase in the number of convolutional layers will cause the output image to become smaller, which will increase the difficulty of generating density maps. Therefore, in this study, a novel convolutional neural network was utilized as the back-end network. This network expanded the perceptual domain while preserving the resolution to produce high-quality density maps.

3.1.1. VGG16

VGGNet [13] is a deep convolutional neural network built by the Visual Geometry Group (VGG) at the University of Oxford for the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) in 2014 [14]. VGGNet is still a popular model for extracting image features. The VGG16 model utilizes small convolutional kernels [15] stacked multiple times with 3×3 networks and a maximum pooled layer of 2×2 networks. The network structure of VGG16 is shown in Figure 4. It consists of 16 layers [16]. The overall network is divided into five segments, each comprising multiple 3×3 convolutional networks connected in series. A maximal pooling layer follows each segment. Finally, there are three fully connected layers and a softmax layer.

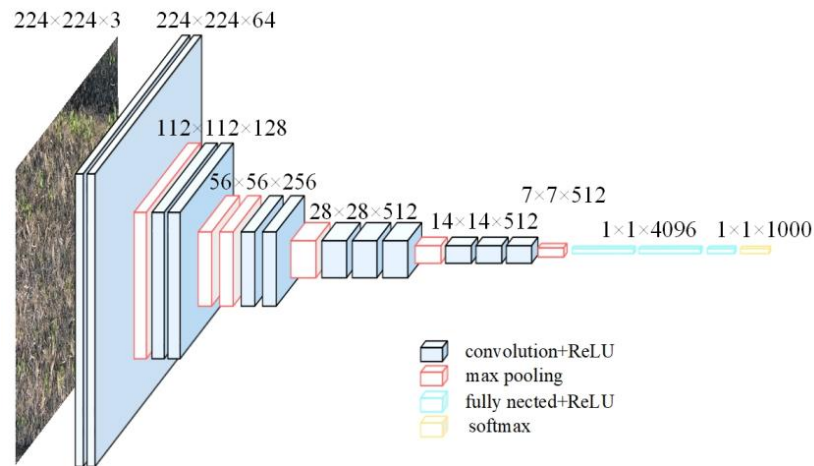


Figure 4. VGG16 network structure.

3.1.2. Dilated Convolution

Traditional convolution is limited by the size of the convolution kernel. For instance, when using a convolution kernel, it can only perceive an input region of a 3×3 size. To capture a wider range of contextual information, a larger receptive field needs to be considered in the density estimation model to capture the global information of the object.

Dilated convolution is a method for sampling data on feature graphs [17] that can expand the receptive field of the convolution kernel to better capture the long-distance dependencies in the image [18]. In addition, traditional convolution reduces the size of the feature map through pooling operations to expand the receptive field [19], which leads to the loss of spatial information. There are some small targets in the rapeseed seedling images, and the conventional convolution operation can cause the structural information of these small targets to be lost, thus affecting the accuracy of rapeseed seedling detection. Therefore, this paper utilizes dilated convolution in the back-end network to produce high-quality density maps. Dilated convolution operates by increasing the dilation rate parameter, constraining the convolution kernel to a specific scale and padding the convolution elements with zeros, thus preserving the structural information of small targets. The expression for calculating the convolution kernel size of dilated convolution is as follows:

$$H \times W = [h + (h - 1) \times (r - 1)] \times [w + (w - 1) \times (r1)] \quad (1)$$

In the above expression, r denotes the dilated rate coefficient in the dilated convolution, and H and W denote the size of the null convolution kernel in height and width, respectively. h and w represent the size of the original convolutional kernel in terms of height and width. The 3×3 convolution kernel is used as an example, and dilated convolution kernels with dilation rate coefficients (r) of 1, 2, and 4 are used, respectively. A diagram is shown in Figure 5.

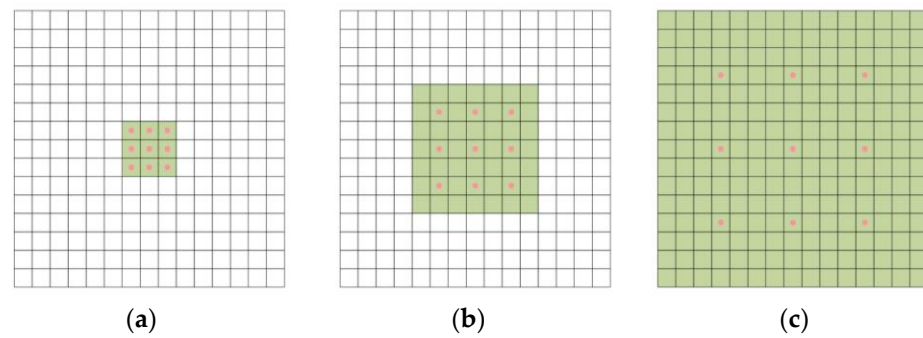


Figure 5. Schematic diagram of the perceptual field of the three cavitation factors. (a) $r = 1$ (b) $r = 2$ (c) $r = 4$.

3.2. The Rape Seedling Counting Model with Improved Density Estimation Method

3.2.1. The SCAM Attention Mechanism

The density distribution of rapeseed plants showed certain regularity as a result of the change in perspective during the rapeseed seedling stage scene. To address this situation, the Spatial Attention Module in SCAM (Spatial Context Attention Module) was utilized to represent large-scale contextual information and capture changes in density. The SCAM attention mechanism combines the SAM spatial and the CAM channel attention mechanisms, as shown in Figure 6. In order to distinguish foreground and background in the density map more effectively, the feature maps of all channels are weighted and summed, and the original channels are updated to enhance the effectiveness of foreground feature extraction. This design helps reduce the impact of viewpoint changes on density estimation and enhances the model’s capacity to adjust to various density distributions.

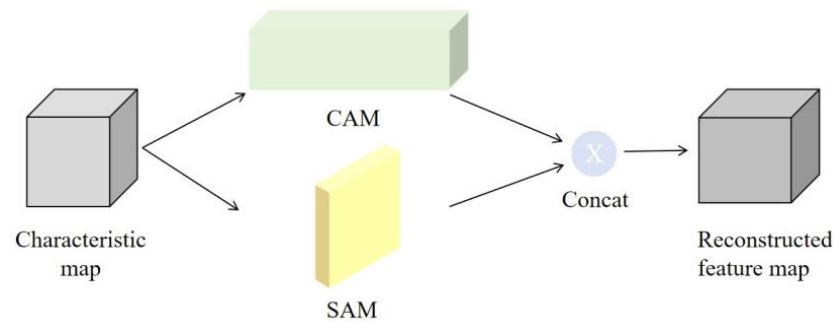


Figure 6. SCAM structure schematic.

SCAM performs feature information enhancement in channel and spatial dimensions through SAM and CAM, respectively. It generates reconstructed feature maps by Concat feature stitching of the enhanced features. The following is a detailed description of the SAM and CAM modules.

(1) The Space Attention Module (SAM)

Figure 7 illustrates the architecture of SAM. For the base feature F output of size $C \times H \times W$, it passes through three different 1×1 convolution layers. Three feature maps, S_1 , S_2 , and S_3 , are obtained through reshape or transpose operations. In order to generate spatial attention graphs, matrix multiplication and softmax classification operations are applied to S_1 and S_2 [20]. Finally, a spatial attention graph S_a of size $HW \times HW$ is obtained. The expression is as follows:

$$S_a^{mn} = \frac{\exp(S_1^i \cdot S_2^j)}{\sum_{i=1}^{HW} \exp(S_1^i \cdot S_2^j)} \tag{2}$$

where S_a^{mn} represents the influence of the m position on the n position. The more similar the feature maps of two locations are, the stronger the correlation between them. After obtaining the spatial attention graph S_a , matrix multiplication is applied again between S_a and S_3 , and the output is reshaped to $C \times H \times W$. For the final summation of F , the output is scaled by a learnable parameter factor. The output definition of SAM is shown in Equation (3):

$$S_{final}^j = \beta \sum_{i=1}^{HW} (S_a^j \cdot S_3^i) + F^j \tag{3}$$

where β is the learnable parameter. In practice, a convolutional layer with a kernel size of 1×1 is utilized to learn β .

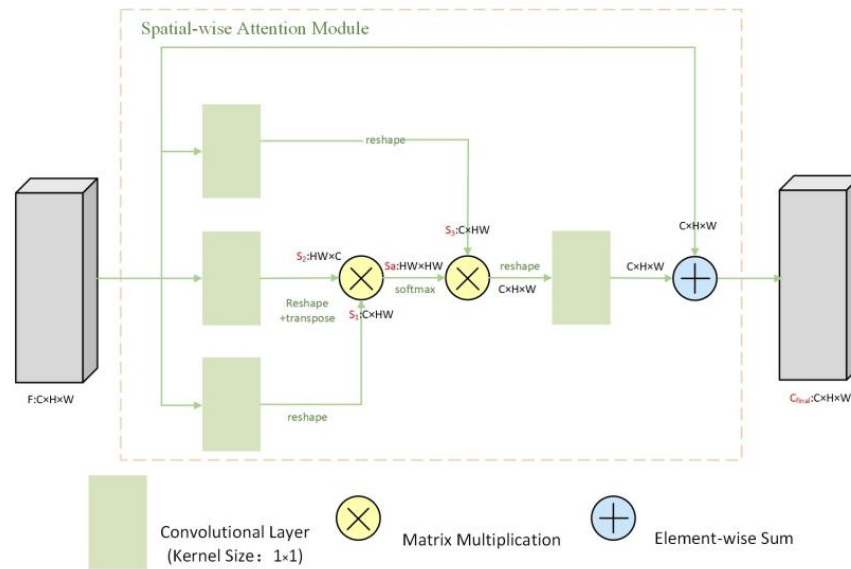


Figure 7. Detailed structure of spatial attention model (SAM).

In the detailed description of the entire spatial attention module, in the final output feature map, S_{final} is a weighted sum of the attention map and the original local feature map, containing global contextual features and self-attention information.

(2) The Channel Attention Module (CAM)

The departmental structure of CAM is shown in Figure 8. The CAM module consists of only one 1×1 convolutional layer to process the feature maps obtained from the backbone network, and the second matrix operation is the matrix operation of $C_a: C \times C$ and $C_3: C \times HW$. The main operation is similar to that of the spatial attention module. Specifically, the C_a of size $C \times C$ is defined as shown in Equation (4):

$$C_a^{ji} = \frac{\exp(C_1^i \cdot C_2^j)}{\sum_{i=1}^C \exp(C_1^i \cdot C_2^j)} \tag{4}$$

where C_a^{ji} denotes the effect of the i channel on the j channel. The C_{final} of size $C \times H \times W$ is calculated as follows:

$$C_{final}^j = \mu \sum_{i=1}^C (C_a^j \cdot C_3^i) + F^j \tag{5}$$

where μ is a learnable parameter. In practical applications, a convolutional layer with a kernel size of 1×1 is utilized to learn μ , which is the same as in Sec.

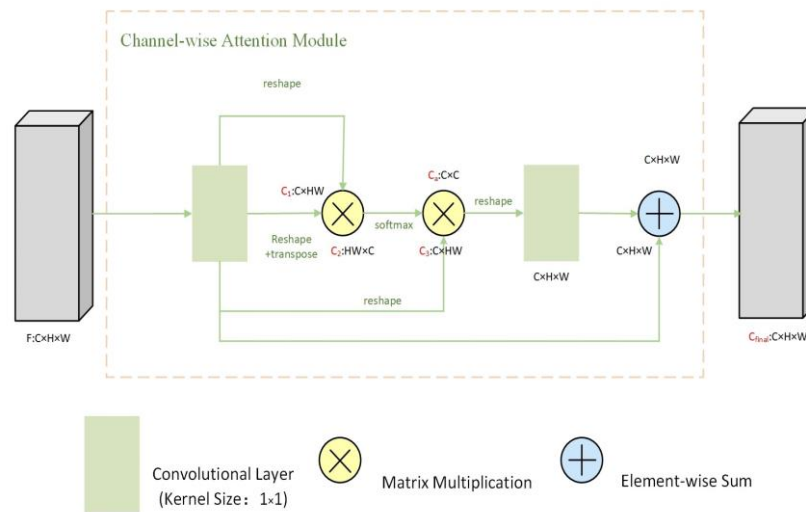


Figure 8. The detailed structure of the channel attention model (CAM).

3.2.2. Loss Function

In this study, the model was trained using a Bayesian loss function that utilizes the point labels in the rapeseed seedling sample data as supervised signals. In order to adapt to the supervised signal format, the expected operation is performed based on the estimated probability density map. This involves matching the discrete expectation of the probability density plot to the point labels to design the loss function and perform regression estimation of the expectation. This method helps to better align the supervised signal and enhance the performance of the model. The detailed process is shown in Figure 9.

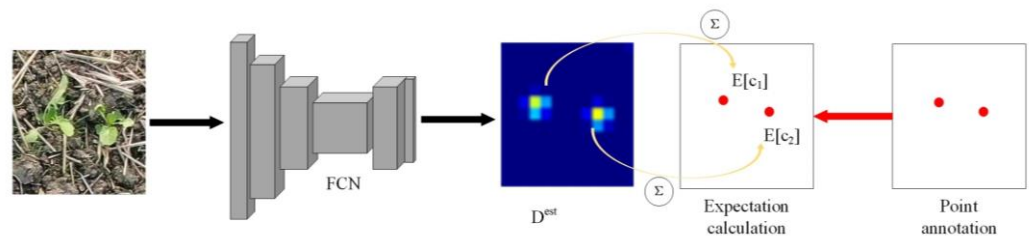


Figure 9. Detailed procedure for the action of Bayesian loss function.

In the Bayesian loss function, a two-dimensional Gaussian distribution is utilized to represent the likelihood probability distribution of the approximate target, enabling the location distribution of each rapeseed plant in the image. Assume that the two-dimensional spatial position of the rapeseed seedling in the density map is x , the spatial position of the rapeseed seedling labeled point is z , and the corresponding label of the rapeseed seedling labeled point is y . The probability that the n th rapeseed plant labeled point in a given point-labeled image is labeled y_n and position x_m can be expressed by Equation (6):

$$P(x_m | y_n) = N(x_m; z_n, \sigma^2 I_{2 \times 2}) \tag{6}$$

where x_m represents the two-dimensional spatial location of the x_m th rapeseed plant, where $m \in [1, M]$, M is the total number of density pixels that are finally output after the neural network model. Meanwhile, z_n denotes the spatial location of the n th rapeseed plant labeled as point, $n \in [1, N]$, while N is the total number of rapeseed plants in the point labeled image. $N(x_m; z_n, \sigma^2 I_{2 \times 2})$ denotes the two-dimensional Gaussian distribution computed at position x_m with mean x_m and covariance matrix $\sigma^2 I_{2 \times 2}$, where $I_{2 \times 2}$ is the second order unit matrix.

The probability of the rapeseed seedling is closely related to its distance from the center marking point. The closer the distance from the center marker, the greater the

probability of the rapeseed seedling, and vice versa—the further the distance, the lower the probability. Equation (6) represents the probability distribution for a given location, according to which the posterior probability of the occurrence of a rapeseed seedling in each pixel of the density map can be calculated as shown in Equation (7):

$$P(x_m | y_n) = \frac{P(x_m | y_n) \cdot P(y_n)}{P(x_m)} = \frac{N(x_m; z_n, \sigma^2 1_{2 \times 2})}{\sum_{n=1}^N N(x_m; z_n, \sigma^2 1_{2 \times 2})} \tag{7}$$

Based on Equation (7), the expected value $E[c_n]$ for the emergence of the n th rapeseed seedling is calculated as

$$E[c_n] = \sum_{m=1}^M P(y_n | x_m) \cdot D^{est}(x_m) \tag{8}$$

Based on Equation (8), $D^{est}(x_m)$ represents the probability density of occurrence of rapeseed seedlings at different locations predicted by the network model.

Equations (6)–(8) represent the calculation of the probability of the occurrence of the rape seedling y_n on the density map, and its ideal expected count value of rape seedling y_n is 1, thus defining the Bayesian loss function l^{Bayes} as

$$l^{Bayes} = \sum_{n=1}^N F(1 - E[c_n]) \tag{9}$$

where $E[c_n]$ is the expected count value of the n th rapeseed seedling.

However, in the experiment, it was found that, while Equation (9) could accurately determine the boundaries of rape seedlings, it could not accurately predict the background pixels far from the marked positions of any rape seedlings. This is because such pixels are likely to be background pixels, whose computation results have a high a posteriori probability. Based on this, this study focused on the background as a specific target, introduced dynamic background dummy elements, and devised a new loss function l^{Bayes+} in the form of Equation (10):

$$l^{Bayes+} = \sum_{n=1}^N F(1 - E[c_n]) + F(0 - E[c_0]) \tag{10}$$

where $E[c_0]$ is the desired count value of the background.

The prediction process only requires inputting any rape seedling image into the network model. Finally, the number of predicted rape seedling images can be obtained by summing the network model output of rape seedling density estimation map $D^{est}(x_m)$. See Equation (11).

$$C = \sum_{n=1}^N E[c_n] = \sum_{n=1}^N \sum_{m=1}^M P(y_n | x_m) \cdot D^{est}(x_m) = \sum_{m=1}^M D^{est}(x_m) \tag{11}$$

3.3. Overall Network Structure

In this paper, the network structure is improved based on the crowd density estimation method, and the improved network structure is shown in Figure 10.

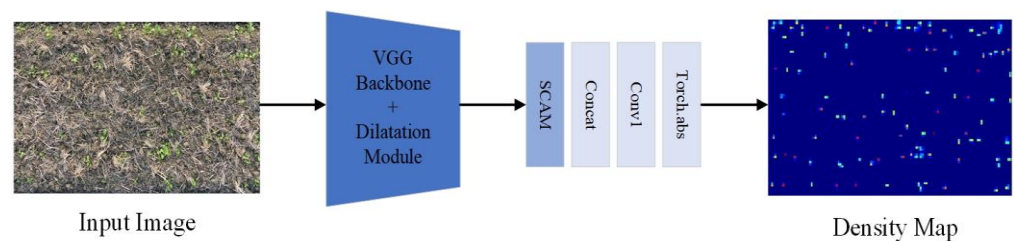


Figure 10. Improved network structure of the density estimation algorithm.

The overall network structure mainly includes five components: a front-end network, an expansion network, an attention network, a regression network, and an absolute quantity output. The front-end network consists of the initial four convolutional layers of VGG16 and the first three pooling layers, which mainly extract the basic features of the input image. The dilated convolution of the base features output from the front-end network is utilized in the expansion network to obtain more image information of rapeseed seedlings and extract more features of rapeseed seedlings. The attention network mainly enhances the feature information of rapeseed seedlings in both channel and spatial dimensions, attentively enhances the more useful features, and reduces the dense error estimation. Finally, more detailed feature extraction is carried out by an 1×1 convolutional layer regression network on the features after performing feature information enhancement, and the absolute number output module is utilized to restrict the distribution of the model parameters in order to generate the final predicted density maps.

3.4. Experimental Environment and Parameter Settings

The experimental environment was based on a Windows 10, Python 3.9, Pytorch [21] and Cuda built deep learning framework [22], the experimental hardware environment graphics card was a Gtx 1660Ti. The number of training iterations in each epoch is 1000 generations, the training batch size Batchsize is set to 1, the learning rate is 1×10^{-6} , and the optimizer is SGD.

4. Results and Discussion

4.1. Evaluation Indicators

The Mean Absolute Error (MAE) and Mean Square Error (MSE) are usually utilized to reflect the accuracy of counting methods in studies related to counting by density estimation methods. They are commonly employed to measure counting performance [23]. MAE is a common loss function used in regression models to reflect the distance between the estimated and true values. MSE is the most commonly used regression loss function to evaluate the performance of density plot estimation [24], which is defined as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^n |Z_i^{\text{tru}} - Z_i^{\text{est}}| \quad (12)$$

$$\text{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^n (Z_i^{\text{tru}} - Z_i^{\text{est}})^2} \quad (13)$$

where N is the number of test samples, Z_i^{tru} is the real number of rape seedlings in the i image, Z_i^{est} is the number of rape seedlings predicted by the model in the i image. The smaller the MAE and MSE, the smaller the counting error and the higher the counting reliability.

4.2. Model Training Analysis

In order to verify the effectiveness of the improved algorithm model in this paper, the changes in training parameters of the improved algorithm during the training process are provided. Figure 11 shows the convergence of related parameters such as the training process loss (Loss), mean squared error (MSE), and the mean absolute error (MAE). As the number of iterations increases, the model's counting error as well as the loss show a convergence tendency, indicating that the enhanced algorithm model can effectively ensure the convergence of the model.

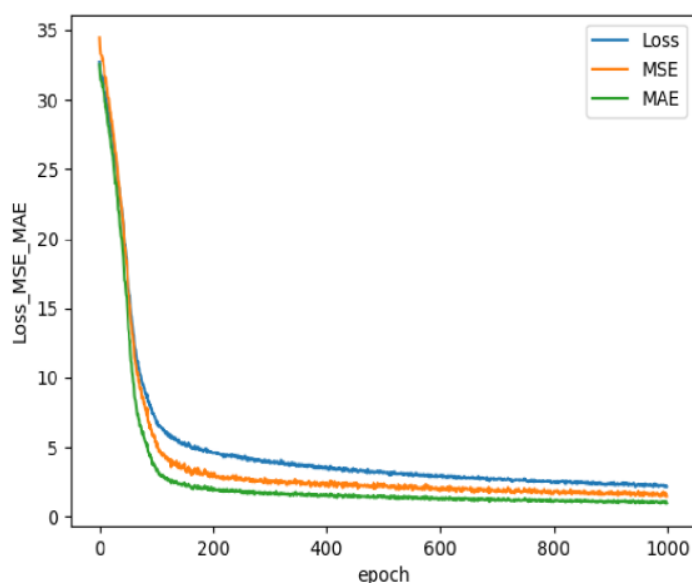


Figure 11. Model training iterative process.

4.3. Distribution Experiment

In order to verify the effectiveness of the algorithms in this chapter on the attention module SCAM, the absolute quantity output module abs, and the 1×1 convolutional regression network module, four sets of comparative experiments were conducted on the rapeseed seedling dataset.

Experiment 1 is represented in Table 2 as A1: VGG16+Dilated+Bayesian Loss. VGG16 was utilized as the front-end network, the dilated null convolution served as the expansion network for feature extraction, and supervised training was performed using the probability density map loss function Bayesian Loss. This result was used as a comparison benchmark for verifying the effectiveness between different modules.

Table 2. Partial data sample display.

Experimental Setup	MAE	MSE
A1: VGG16+Dilated+Bayesian Loss	4.12	5.76
A2: VGG16+Dilated+Bayesian Loss+SCAM	3.68	5.29
A3: VGG16+Dilated+Bayesian Loss+SCAM+abs	3.60	5.20
A4: VGG16+Dilated+Bayesian Loss+SCAM+Conv1 \times 1	3.52	5.09
A5: VGG16+Dilated+Bayesian Loss+SCAM+Conv1 \times 1+abs	3.40	4.99

Experiment 2 is represented in Table 2 as A2: VGG16+Dilated+Bayesian Loss+SCAM. The attention module SCAM was added based on the basis of Experiment 1, so that, after the image entered the front-end network and the expansion network, the global feature information was enhanced using the attention module from the channel and spatial dimensions, respectively, and the attentional enhancement of the more useful features to reduce error estimation due to occlusion. Finally, supervised training was performed by the loss function Bayesian Loss.

Experiment 3 is represented in Table 2 as A3: VGG16+Dilated+Bayesian Loss+S-CAM+abs. The absolute number of the abs output module was added on the basis of 2, so that, after being enhanced with network features, it was constrained to the distribution of the model parameters by the absolute number of the abs output module, thus obtaining a better generalization performance and reducing the counting error.

Experiment 4 is represented in Table 2 as A4: VGG16+Dilated+Bayesian Loss+S-CAM+Conv1. A 1×1 convolutional regression network module was added to Experiment 2 for further feature extraction at each pixel point of the feature-enhanced model.

Experiment 5 is represented in Table 2 as A5: VGG16+Dilated+Bayesian Loss+S-CAM+abs+ Conv11. All three proposed modules were added to Experiment 1 to verify the effect of the final improved model. The counting results of the step-by-step experiments are shown in Table 2.

Comparison of Experiment 1 and Experiment 2 shows that the MAE of the model counts decreased by 0.44 as well as the mean square error after using the attentional mechanism to enhance the feature information of the image features in both channel and spatial dimensions.

Comparing Experiment 2 and Experiment 3, it can be seen that constraining and optimizing the parameters of the feature-enhanced model by the abs absolute quantity output module does play a role in reducing the counting error.

Comparing Experiment 2 and Experiment 4, it can be seen that more accurate pixel counting also leads to a significant improvement in the counting error in the feature maps after the Concat feature splicing of the spatial and channel attention modules through a 1×1 convolutional regression network, which reduces the average absolute error by 0.16 in this metric alone, with a corresponding decrease in the MSE.

Comparing the effects of Experiment 4 with the other parameters, the 1×1 convolutional regression network with the abs absolute number output module also made an improvement in some of the redundant information after feature enhancement, reducing the model's counting error to 3.39.

In addition, the results of the step-by-step experiments described above are visualized in Figure 12. Figure 12a is the original figure, and Figure 12b–f correspond to Experiment 1–5, respectively. The brighter region in the density map of rapeseed indicates a higher density of the plant, and it is evident from the figures that Figure 12f shows more prominent density-responsive points in the overall characterization map of the rape seedling region compared to the other four density maps.

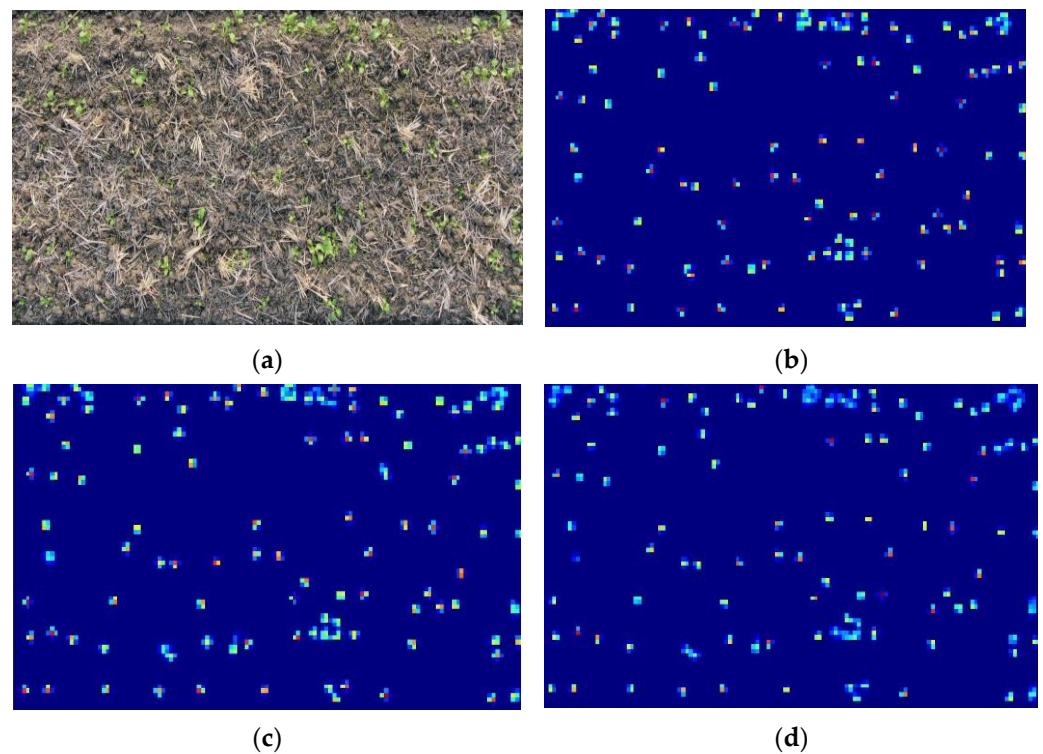


Figure 12. *Cont.*

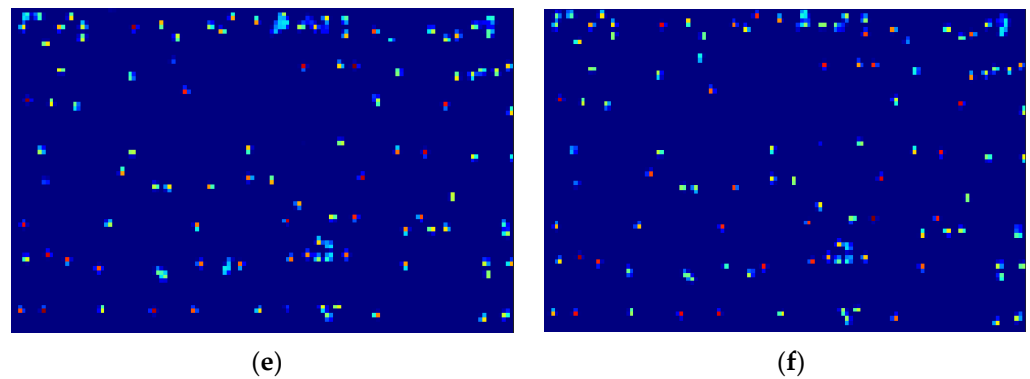


Figure 12. Visualization of test results of distribution test. (a) Original figure. (b) A1. (c) A2. (d) A3. (e) A4. (f) A5.

4.4. Comparison with Other Algorithms

In order to further verify the effectiveness of the algorithm discussed in this chapter, this paper also selected two other density estimation algorithms, CSRNet and Bayesian, for experimental comparison of rape seedlings, and the algorithm based on improved density estimation was named BCNet. The comparison results are shown in Table 3.

Table 3. Comparison of counting results of different algorithms.

Algorithm	MAE	MSE
CSRNet	8.32	12.48
Bayesian	3.71	5.20
BCNet	3.40	4.99

From Table 3, it can be seen that, compared to the other two algorithms, BCNet has the lowest counting error, with an MAE and MSE of 3.40 and 4.99 for, respectively, indicating the highest counting accuracy.

In order to better compare the detection results across different algorithms, some of the visualized detection results are given in Figure 13, while the actual counts are included to compare the error between the predicted and actual counts. In the visualization results of the three density estimation algorithms, the brighter the region in the density map, the higher the density of the rapeseed seedlings. Due to the similar color and irregular shape of the characteristics of the target region of rapeseed seedlings, applying density estimation algorithms may cause the probability value of the density response point of a single shaded plant in a dense region to be less than 1, and therefore also result in a situation where the total number of probability predictions is lower than the actual number. By comparing the disparity between the predicted counts and the number of actual counts, it is evident that the BCNet algorithm's predicted counts more accurately reflect the actual counts.

4.5. Model Counting Performance Analysis

Based on the BCNet algorithm with an improved density estimation method, this paper predicted the counts of 150 rapeseed seedling test sets during the testing process and compared the predicted number with the labeled true numbers, and the results are shown in Figure 14. It is evident from the figure that the two curves of the actual number and the predicted number are very close to each other, and the predicted number for the majority of rape seedling test pictures has only a small error with the actual number.

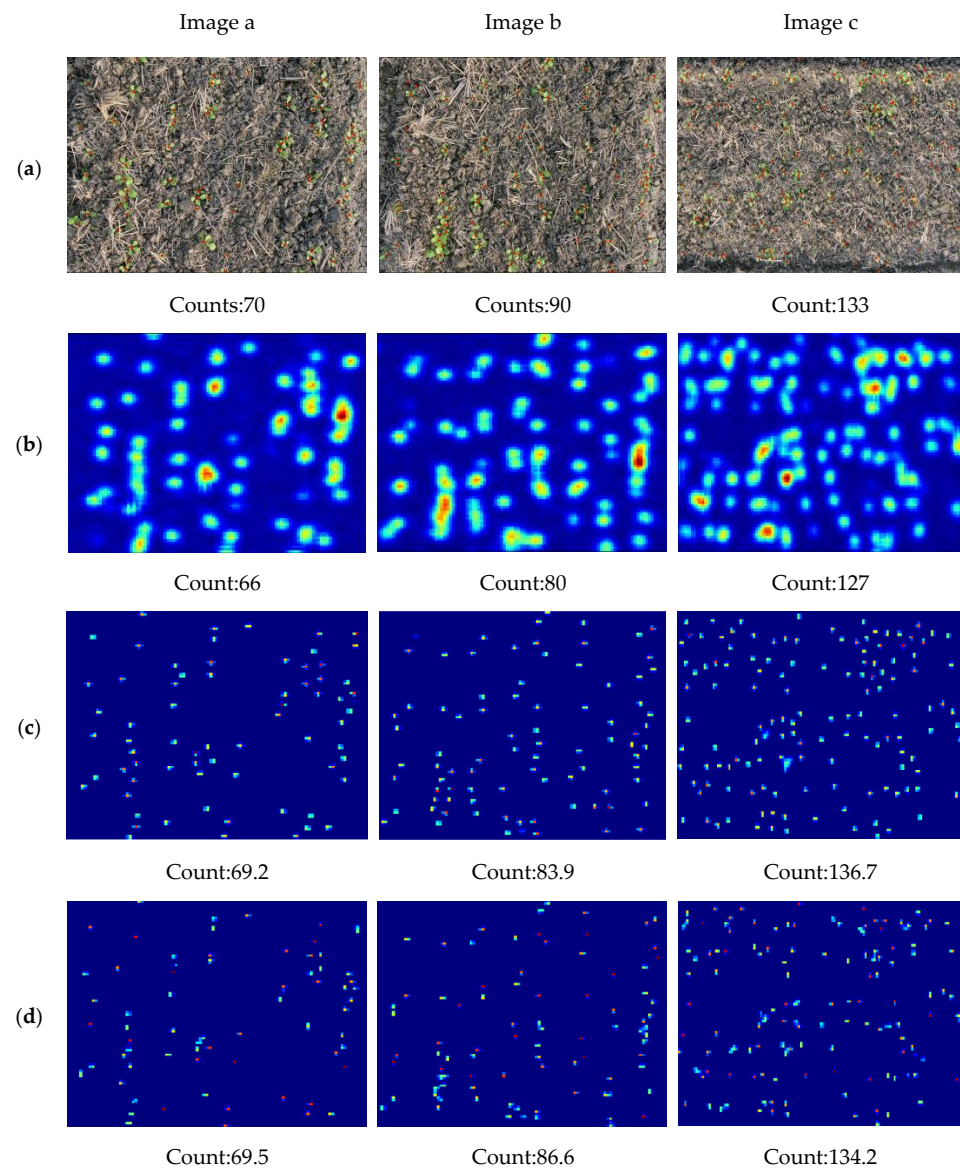


Figure 13. Visual detection results for different algorithms. (a) TrueCount. (b) CSRNet. (c) Bayesian. (d) BCNet.

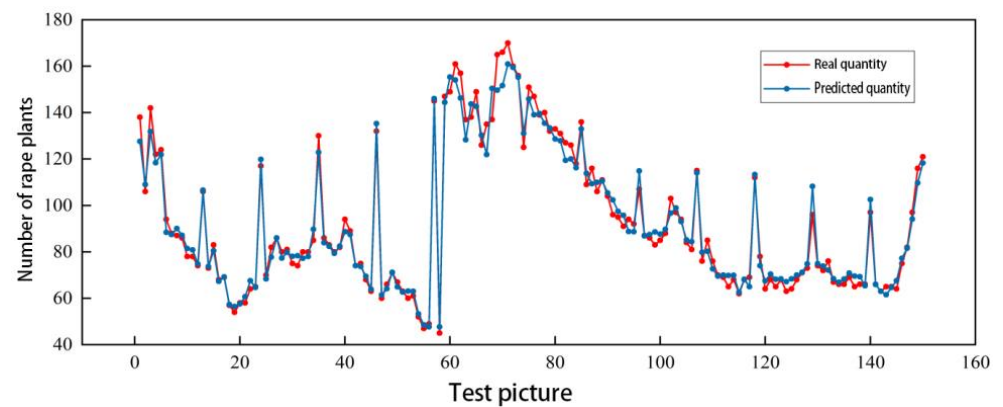


Figure 14. Comparison of the real number of rape plants in the test set with the predicted number.

5. Conclusions

1. The density estimation method utilized spatial attention, a channel attention module, feature information enhancement, and splicing, respectively, which improved the representation of the entire feature map, using a 1×1 convolutional layer for further feature extraction and introducing the torch.abs function at the output port of the network. This improved model is named BCNet.
2. Distribution experiments and result visualization were performed, and the density response points corresponding to the features of the seedling region were more prominent in the improved density map compared to the other four density maps. Compared with the CSRNet and the Bayesian algorithms, BCNet has the lowest counting error, with MAE and MSE reaching 3.40 and 4.99, respectively. BCNet exhibits the highest counting accuracy, as the predicted count closely aligns with the actual count.
3. Under complex backgrounds or high density conditions, the BCNet algorithm was employed to predict the count of rapeseed seedlings, and the two curves of the actual and predicted numbers of rapeseed seedlings obtained were very close to each other, which verified the feasibility of the method in this paper. It can provide a reference for seedling identification and counting methods of rapeseed, and provide technical support for achieving precise seedling interplanting and seedling replenishment.

Author Contributions: Conceptualization, Q.W. and C.L.; methodology, C.L.; validation, Q.W. and C.L.; investigation, Q.W.; resources, L.L.; data curation, Q.W. and C.L.; writing—original draft preparation, Q.W. and C.L.; writing—review and editing, Q.W. and L.L.; supervision, Q.Z. and L.L.; project administration, L.H. and L.C.; funding acquisition, L.C. and L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 52305241) and the National Key Research and Development Program of China (Grant No. 2022YFD2301402).

Institutional Review Board Statement: Not applicable

Data Availability Statement: Due to the sensitivity and confidentiality of the data, this study does not provide the original data when publishing the paper. For data acquisition, please contact the corresponding author.

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

References

1. Liu, C.; Feng, Z.C.; Xiao, T.H.; Ma, X.M.; Zhou, G.S.; Huang, F.H.; Li, G.N.; Wang, H.Z. Development status, potential and countermeasures of rapeseed industry in China. *Chin. J. Oil Crop Sci.* **2019**, *41*, 485–489.
2. Dai, J.G.; Xue, J.L.; Zhao, Q.Z.; Wang, Q.; Chen, B.; Zhang, G.S.; Jiang, N. Extracting cotton seedling information from UAV visible light remote sensing images. *Trans. CSAE* **2020**, *36*, 63–71. [[CrossRef](#)]
3. Liu, L.C.; Liang, J.; Wang, J.Q.; Hu, P.Y.; Wan, L.; Zheng, Q. An Improved YOLOv5-Based Approach to Soybean Phenotype Information Perception. *Comput. Electr. Eng.* **2023**, *106*, 108582. [[CrossRef](#)]
4. Liu, L.C.; Bi, Q.P.; Liang, J.; Li, Z.D.; Wang, W.W.; Zheng, Q. Farmland Soil Block Identification and Distribution Statistics Based on Deep Learning. *Agriculture* **2022**, *12*, 2038. [[CrossRef](#)]
5. Peng, J.; Rezaei, E.E.; Zhu, W.; Wang, D.; Li, H.; Yang, B.; Sun, Z. Plant Density Estimation Using UAV Imagery and Deep Learning. *Remote Sens.* **2022**, *14*, 5923. [[CrossRef](#)]
6. de Arruda, M.d.S.; Osco, L.P.; Acosta, P.R.; Gonçalves, D.N.; Junior, J.M.; Ramos, A.P.M.; Matsubara, E.T.; Luo, Z.; Li, J.; de Andrade Silva, J. Counting and locating high-density objects using convolutional neural network. *Expert Syst. Appl.* **2022**, *195*, 116555. [[CrossRef](#)]
7. Lempitsky, V.; Zisserman, A. Learning to count objects in images. In *Advances in Neural Information Processing Systems 23 (NIPS 2010)*; Curran Associates Inc.: Red Hook, NY, USA, 2010.
8. Qi, Y.; Li, Y.N.; Sun, M. Cotton seedling counting algorithm based on feature fusion. *Trans. CSAE* **2022**, *38*, 180–186.
9. Huang, Z.Y.; Li, Y.N.; Wang, H.H. Cotton boll counting algorithm in the field based on density class classification. *J. Comput.-Aided Des. Comput. Graph.* **2020**, *32*. [[CrossRef](#)]

10. Bao, W.X.; Zhuang, X.; Hu, G.S.; Huang, L.S.; Liang, D.; Lin, Z. Estimation and Counting of Wheat Ear Density in the Field Based on Deep Convolutional Neural Network. *Trans. CSAE* **2020**, *36*, 186–193+323.
11. 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. [[CrossRef](#)]
12. Li, Y.; Zhang, X.; Chen, D. CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
13. Karen, S. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
14. Sujatha, R.; Chatterjee, J.M.; Jhanjhi, N.; Brohi, S.N. Performance of deep learning vs machine learning in plant leaf disease detection. *Microprocess. Microsyst.* **2021**, *80*, 103615. [[CrossRef](#)]
15. Tao, T.; Wei, X. A hybrid CNN–SVM classifier for weed recognition in winter rape field. *Plant Methods* **2022**, *18*, 29. [[CrossRef](#)] [[PubMed](#)]
16. Manataki, M.; Papadopoulos, N.; Schetakis, N.; Di Iorio, A. Exploring Deep Learning Models on GPR Data: A Comparative Study of AlexNet and VGG on a Dataset from Archaeological Sites. *Remote Sens.* **2023**, *15*, 3193. [[CrossRef](#)]
17. Li, Y.W.; XU, J.J.; Liu, D.X.; Yu, Y. Field Road Scene Recognition in Hilly and Mountainous Areas Based on Improved Dilated Convolutional Neural Network. *Trans. CSAE* **2019**, *35*, 150–159.
18. Zhu, L.; Wu, R.; Fu, G.; Zhang, S.; Yang, C.; Chen, T.; Huang, P. Segmenting banana images using the lightweight UNet of multi-scale serial dilated convolution. *Trans. CSAE* **2022**, *38*, 194–201. [[CrossRef](#)]
19. Wang, L.Y. Research on Image Crowd Counting Based on Convolutional Neural Network. Ph.D. Thesis, University of Science and Technology of China, Hefei, China, 2020.
20. Tan, S.H.; Chuah, J.H.; Chow, C.O.; Kanesan, J. Spatially Recalibrated Convolutional Neural Network for Vehicle Type Recognition. *IEEE Access* **2023**, *11*, 142525–142537. [[CrossRef](#)]
21. Kim, T.; Lee, D.H.; Kim, W.S.; Zhang, B.T. Domain adapted broiler density map estimation using negative-patch data augmentation. *Biosyst. Eng.* **2023**, *231*, 165–177. [[CrossRef](#)]
22. Sozzi, M.; Cantalamessa, S.; Cogato, A.; Kayad, A.; Marinello, F. Automatic bunch detection in white grape varieties using YOLOv3, YOLOv4, and YOLOv5 deep learning algorithms. *Agronomy* **2022**, *12*, 319. [[CrossRef](#)]
23. Wu, J.; Yang, G.; Yang, X.; Xu, B.; Han, L.; Zhu, Y. Automatic counting of in situ rice seedlings from UAV images based on a deep fully convolutional neural network. *Remote Sens.* **2019**, *11*, 691. [[CrossRef](#)]
24. Feng, W.; Wang, K.; Zhou, S. An efficient neural network for pig counting and localization by density map estimation. *IEEE Access* **2023**, *11*, 81079–81091. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.