MDPI

Article

Residual Self-Calibration and Self-Attention Aggregation Network for Crop Disease Recognition

Qiang Zhang [1], **Banyong Sun** [2], **Yaxiong Cheng** [1] **and Xijie Li** [2,*]

[1] School of Science, Wuhan University of Technology, Wuhan 430070, China;
zhangqiang977966@yeah.net (Q.Z.); drchern@yeah.net (Y.C.)
[2] Key Laboratory of Spectral Imaging Technology, Xi'an Institute of Optics and Precision Mechanics of CAS,
Xinxi Road No. 17, Xi'an 710119, China; hhhhaas@yeah.net
[*] Correspondence: lixijie@opt.ac.cn

Abstract: The correct diagnosis and recognition of crop diseases play an important role in ensuring crop yields and preventing food safety. The existing methods for crop disease recognition mainly focus on accuracy while ignoring the algorithm's robustness. In practice, the acquired images are often accompanied by various noises. These noises lead to a huge challenge for improving the robustness and accuracy of the recognition algorithm. In order to solve this problem, this paper proposes a residual self-calibration and self-attention aggregation network (RCAA-Net) for crop disease recognition in actual scenarios. The proposed RCAA-Net is composed of three main modules: (1) multi-scale residual module, (2) feedback self-calibration module, and (3) self-attention aggregation module. Specifically, the multi-scale residual module is designed to learn multi-scale features and provide both global and local information for the appearance of the disease to improve the performance of the model. The feedback self-calibration is proposed to improve the robustness of the model by suppressing the background noise in the original deep features. The self-attention aggregation module is introduced to further improve the robustness and accuracy of the model by capturing multi-scale information in different semantic spaces. The experimental results on the challenging 2018ai_challenger crop disease recognition dataset show that the proposed RCAA-Net achieves state-of-the-art performance on robustness and accuracy for crop disease recognition in actual scenarios.

Keywords: crop disease recognition; self-calibration; self-attention; residual

1. Introduction

The occurrence of crop diseases has a certain negative impact on agricultural production. If crop diseases are not discovered in time, it will increase the risk of food loss [1], especially for some major food crops, such as corn, rice, wheat, etc., which are key to meeting human living needs and promoting productivity development. Therefore, it is of great practical significance to explore an intelligent, low-cost, and highly accurate method to implement crop disease recognition. The feature extraction and pattern recognition in machine learning help to identify the type and severity of crop diseases. Automatic quality analysis of plant health status through the color, shape and size of plant leaf images is an accurate and reliable method to improve productivity [2,3].

Crop disease recognition based on traditional image processing methods is incomparable in recognition accuracy and robustness compared to methods based on deep neural networks that have emerged in recent years. Most of the current methods based on deep neural networks are trained on the public dataset PlantVillage [4] or simple background images to construct models for realizing crop disease image recognition. However, the type of method creates some problems. The public dataset PlantVillage has a simple background, and the characteristics of crop diseases are diverse. Since the acquisition of annotated images requires the participation of experts, the categories are often unbalanced,

and the direct migration of the model trained on PlantVillage is not very good. When the disease recognition method based on simple background images is applied to recognize the crop disease in the actual environment, it needs to fight against various noise interference factors. In addition, the actual recognition accuracy will be greatly reduced, which cannot meet the practical requirement.

Aiming at the characteristics of crop disease image recognition with a complex background, more interference, and diverse disease features, this paper takes both the recognition accuracy and robustness of the model into account and proposes a residual self-calibration and self-attention aggregation network (RCAA-Net) for crop disease recognition in actual scenarios. The main contributions of this paper are as follows:

- A residual self-calibration and self-attention aggregation network is proposed for crop disease recognition in actual scenarios. For the problem of crop disease recognition in actual scenarios, the proposed RCAA-Net can achieve a double improvement of accuracy and robustness.
- A feedback self-calibration module is proposed to further suppress the background noise in the original deep features by fine filtering and adjusting the network features again, thereby effectively improving the robustness of the model.
- A self-attention aggregation module is proposed to automatically focus on discriminative regions by capturing multi-scale information in different semantic spaces, thereby further improving the robustness and accuracy of the model.

The rest of this paper is arranged as follows. Section 2 summarizes the related work; Section 3 details the proposed RCAA-Net method; Section 4 introduces the experimental settings and results; Section 5 gives the conclusions.

2. Related Work

Crop disease image recognition is a comprehensive use of image processing, phytopathology, pattern recognition and other technical means to analyze disease image information to obtain the characteristic representation and classification model of the disease so as to accurately classifying the disease category. According to the current idea of disease image recognition, methods can be divided into the following two categories.

2.1. Traditional Image Processing Methods

Many previous works have considered the problem of image recognition and apply a special classifier to discern healthy and diseased images. Generally speaking, plant leaves are primary information for the recognition of crop diseases because most of the symptoms of diseases first appear on leaves. In the past few decades, the recognition and classification of major diseases have been widely used in plants, including K-Nearest Neighbor (KNN) [5], Support Vector Machine (SVM) [6], Fisher Linear Discriminant (FLD) [7], Artificial Neural Network (ANN) [8], Random Forest (RF) [9], etc. The disease recognition rate of classical methods largely depends on the lesion segmentation of various algorithms and hand-crafted features, such as seven invariant moments, scale-invariant feature transform (SIFT), Gabor transform, global-local singular values and sparse representation [10–12]. However, hand-crafted features require expensive resource conditions and professional knowledge, and at the same time, have a certain degree of subjectivity. Moreover, it is difficult to determine which disease recognition features are the best and most robust from the extracted target. In addition, most methods cannot effectively separate leaves and lesion images from the background under complex conditions, resulting in failure to predict the occurrence of disease. Therefore, due to the complexity of diseased leaf images, automatic recognition of crop disease is still a challenging task.

2.2. Deep Neural Network Methods

In recent years, deep learning techniques, especially convolutional neural networks (CNN) [13–15], are rapidly becoming the preferred method to overcome the above-mentioned challenges [16–20]. Due to the scale invariance of the convolutional neural network, the

image problem it solves is not limited by the scale and shows outstanding ability in recognition and classification. For example, Mohanty et al. [21] trained a deep learning model to identify 14 crops and 26 crop diseases. Ma et al. [22] used deep CNN to identify the symptoms of cucumber downy mildew, anthracnose, powdery mildew and target leaf spot, with a recognition accuracy of 93.4%. Kawasaki et al. [23] proposed a CNN-based cucumber leaf disease recognition method, which achieved an accuracy of 94.9%. Similarly, this paper also uses CNN to extract plant leaf disease characteristics and proposes a lightweight convolutional network based on VGG-16. First of all, the original network introduces depthwise separable convolution (DSC) [24] and global average pooling (GAP) [25] to replace the standard convolution operation and perform the complete operation at the end of the network. The connection layer is replaced, and at the same time, batch normalization technology is applied to training the network and improving the data distribution in the middle layer and increasing the convergence speed [26]. The experimental results of the improved network on the plant leaf disease dataset PlantVillage show that the proposed lightweight convolutional network has a significant improvement in recognition accuracy and efficiency and is suitable for the task of plant leaf disease recognition, which has strong engineering practicality and high research value. Most of these methods are aimed at PlantVillage or simple background image recognition. When facing the recognition of complex background and various noise interference in the actual environment, the recognition accuracy will often be greatly reduced due to the complex background noise interference. Therefore, improving the accuracy, robustness and anti-interference ability of crop disease image recognition in the actual environment has become the key to crop disease recognition.

3. Methods

This paper aims to build a novel deep convolutional neural network with simple, accurate, robust and strong anti-interference ability to achieve high-precision recognition of crop disease in images. This section first introduces the framework of the proposed RCAA-Net. Then, the multi-scale residual module, feedback self-calibration module and self-attention aggregation module are elaborated, respectively. Finally, the network parameters of the proposed RCAA-Net method are reported, and the loss function is provided.

3.1. Overview

The proposed RCAA-Net method realizes the accurate recognition of crop diseases and, meanwhile, takes the anti-interference ability into account. The overall network structure of the RCAA-Net method is shown in Figure 1. For the disease image to be classified, this paper adopts 1 convolutional layer, 3 residual modules, 3 parallel feedback self-calibration modules, 1 self-attention aggregation module, 1 global average pooling layer and 1 Softmax layer to directly output the category probability of the cropped image.

In Figure 1, in order to effectively utilize features of different scales, the output of the three residual modules is adopted in the proposed RCAA-Net method. By synthesizing the features of the three scales, it can provide richer features for the subsequent network layer, improve the recognition accuracy of the model, and indirectly improve the robustness of the network. In order to finely filter the image features to improve their anti-interference ability, for each scale feature, we input a feedback self-calibration module to finely filter the image features and improve anti-interference abilities. In addition, the three scale features processed by the feedback self-calibration module are input to the self-attention aggregation module to capture multi-scale information in different semantic spaces to automatically focus on the discriminative regions, thereby further improving the robustness and accuracy of the model. The three main modules in the proposed RCAA-Net are described in detail as follows.
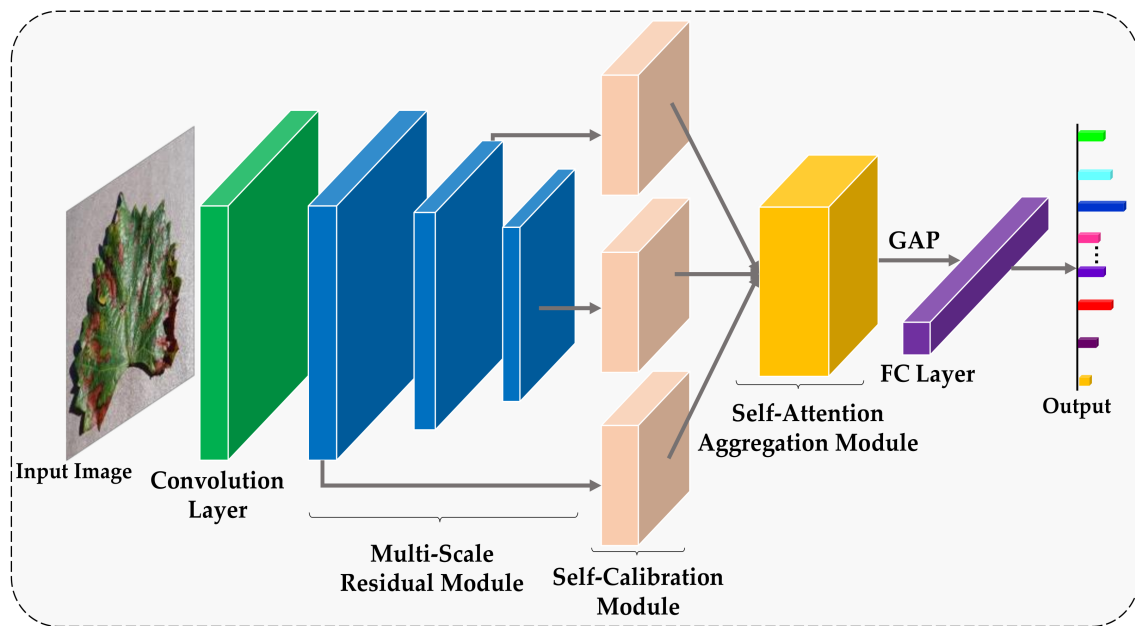
Int. J. Environ. Res. Public Health **2021**, *18*, 8404

4 of 14

Figure 1. The framework of the proposed RCAA-Net.

3.2. Multi-Scale Residual Module

Residual network [27] has achieved satisfying results on IRSVRC. The residual network can not only speed up the network fitting and improve the recognition accuracy, but also has a certain anti-interference ability. Prior to this, residual networks for crop disease recognition had not attracted enough attention and research. In addition, multi-scale features can use different levels of semantic information at the same time, thereby avoiding the adverse mesoscale effects in crop disease recognition. To this end, we designed a multi-scale residual module to effectively solve the above problems and improve the performance of the crop disease recognition model.

The proposed multi-scale residual module consists of three consecutive residual modules, and the structure of each residual block is shown in Figure 2. As can be seen from Figure 2, each residual block consists of 3 dilation convolutional layers, which are respectively denoted as Conv1, Conv2 and Conv3. The detailed parameters of each convolutional layer are listed in Figure 2. Here, the adoption of dilation convolution is to increase the receptive field of feature points, thereby handling large-scale variance of the lesion area for crop disease without introducing additional computation [28]. The output after the input passes through Conv1, Conv2 and Conv3 are denoted as X_1, X_2 and X_3, respectively. We directly cascade X_1 and X_3 as the total output of the entire residual block. The specific cascade model can be expressed by:

$$R(X_1) = X_1 + X_3 = \Gamma(X_1) + X_1, \tag{1}$$

where $\Gamma(\cdot)$ represents the residual mapping function and $R(\cdot)$ represents the output of each residual block.

The residual block obtains more prominent fine information in the image by learning the residual mapping function. The residual block realizes that the low-level features extracted through Conv1 convolution and the high-level detailed features acquired through Conv1, Conv2 and Conv3 three-layer convolution are transmitted to the following network at the same time, and more refined feature extraction is continued. By inputting the output of the three residual blocks as multi-scale features into the subsequent network, the detailed description of the low-level features and the abstract representation of the high-level features in the convolutional neural network can be comprehensively used to

provide rich and detailed feature representation for the appearance of the disease. In this way, the recognition accuracy of the model can be effectively improved.
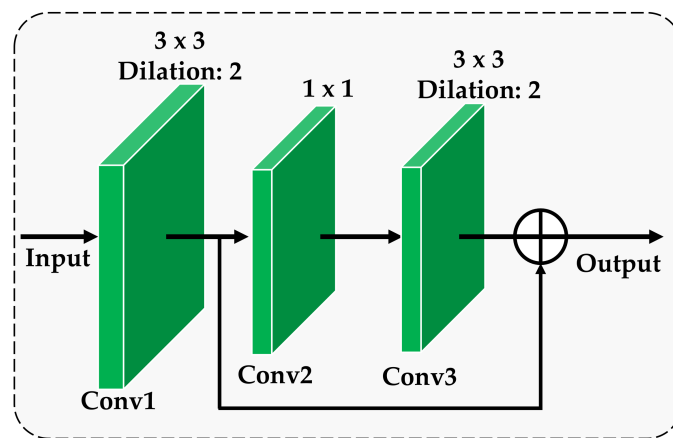

Figure 2. The architecture of the residual block.

3.3. Feedback Self-Calibration Module

In order to achieve high recognition accuracy and anti-interference ability for the images collected in the actual environment that may contain various noise factors, a feedback self-calibration module is designed, and its structure is shown in Figure 3. The feedback self-calibration module is to reload the convolutional layer, perform two deconvolution operations after loading, and then return the deconvolution result to the previous shallow layer. Subsequently, it is passed as output to the subsequent network layer model after repeated loading. The convolutional layer involved is a 3×3 convolution kernel with a step size of 1. We can clearly see the entire process of feedback to the self-calibration module from Figure 3. Let the input of the feedback self-calibration module be X_c, and the result after deconvolution be X_{dec}. Then, the feedback self-calibration module can be optimized under the constraints of the following equation:

$$\updownarrow_c = \|\Psi(X_c) - X_c\|_2 = \|X_{dec} - X_c\|_2, \tag{2}$$

where $\Psi(\bullet)$ represents the feedback self-calibration function and $\|\bullet\|_2$ represents the L2 norm. Through this constraint, the features after deconvolution can be used to feedback and adjust the original deep features, thereby suppressing the background noise in the original deep features and improving the robustness of the model.
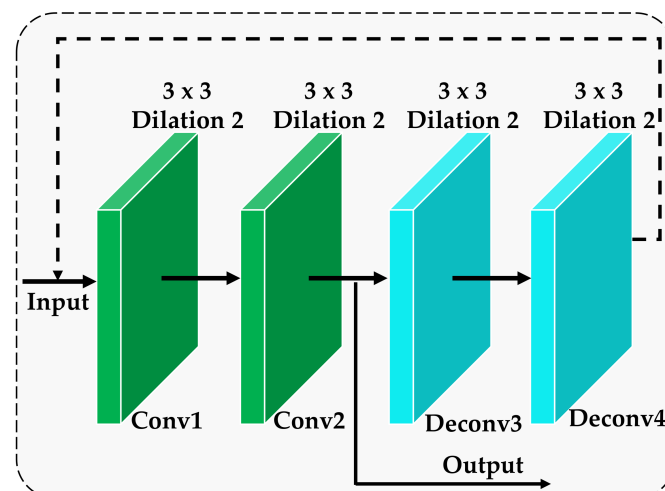

Figure 3. The architecture of the feedback self-calibration module.

Int. J. Environ. Res. Public Health **2021**, *18*, 8404

6 of 14

In summary, the purpose of introducing the feedback self-calibration module in this paper is to return the features of the deep convolutional layer in the network to the shallow convolutional layer so that the network features can be fine-filtered and readjusted. In this way, the background noise in the original deep features is further suppressed, effectively improving the robustness of the model.

3.4. Self-Attention Aggregation Module

Research has found that attention can selectively focus on important information in the data. This paper takes this as inspiration and draws on the Transformer model [29]. The multi-head self-attention mechanism (MHA) is adopted to extract dependency relationships in different semantic spaces. The architecture is shown in Figure 4. Multi-head self-attention is based on the principle of scaled dot-product attention, and its calculation formula is as follows:

$$\hbar(\mathbf{Q}, \mathbf{V}, \mathbf{K}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \tag{3}$$

where $\hbar(\cdot, \cdot, \cdot)$ stands for scaled dot-product attention operation, $\mathbf{Q}, \mathbf{V}, \mathbf{K}$ are the query, value, and key matrix for calculating self-attention, respectively. $\mathbf{Q}\mathbf{K}^T$ is the attention matrix, weighting the $\mathbf{V}$ matrix. d_k represents the dimension of the key. $\sqrt{d_k}$ turns the attention matrix into a standard normal distribution so that the result is more stable and a balanced gradient can be obtained when backpropagating.
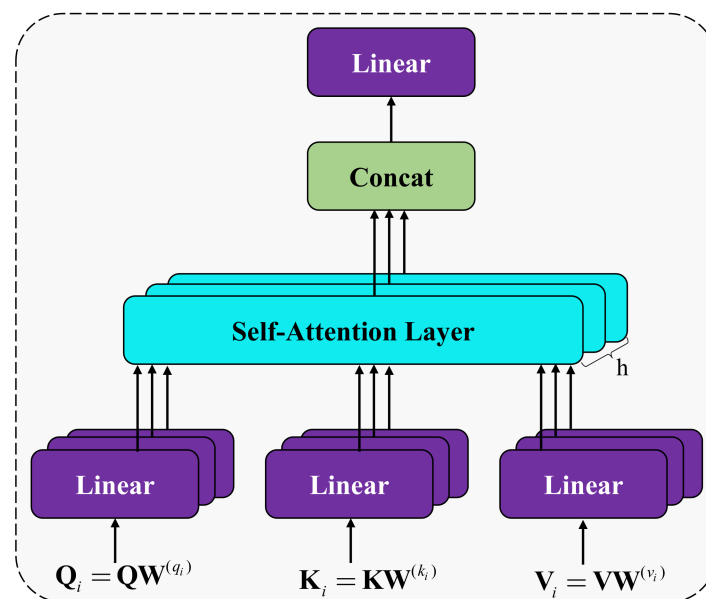

Figure 4. The architecture of the self-attention aggregation module.

Based on the scaled dot-product attention calculated by Equation (1), the semantic features are integrated from the subspace containing different semantic information.

Furthermore, the value of MultiHead is obtained through the following two steps.

(1) Firstly, the $\mathbf{Q}, \mathbf{V}$ and $\mathbf{K}$ matrices are mapped into multiple subspaces:

$$\begin{cases} \mathbf{Q}_i = \mathbf{Q}\mathbf{W}^{(q_i)} \\ \mathbf{K}_i = \mathbf{K}\mathbf{W}^{(k_i)} \\ \mathbf{V}_i = \mathbf{V}\mathbf{W}^{(v_i)} \end{cases}, \tag{4}$$

where $\mathbf{Q}_i$, $\mathbf{K}_i$, and $\mathbf{V}_i$ are the query, key and value matrix of each subspace. $\mathbf{W}^{(q_i)}$, $\mathbf{W}^{(k_i)}$ and $\mathbf{W}^{(v_i)}$ are conversion matrices.

(2) Secondly, the scaled dot-product attention in each subspace is calculated in parallel, and then the results are concatenated to obtain the context matrix after linear transformation:

$$\text{head}_i = \hbar(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) \tag{5}$$

$$\text{MultiHead} = \text{Concat}(\text{head}_1, \text{head}_2, \ldots, \text{head}_h)\mathbf{W}^T \tag{6}$$

where head_i is the scaled dot-product attention of each subspace, and MultiHead is the final result.

3.5. Network Parameters and Loss Function

In this paper, RCAA-Net is a simple, accurate and highly robust convolutional neural network. Table 1 lists the main parameters in this method. Among them, the parameters of the multi-scale residual module and the feedback self-calibration module are shown in Figures 2 and 3, respectively. In order to reduce the number of network parameters, we only use two types of kernels, 1×1 and 3×3, which also helps to avoid overfitting due to the small image set.

Table 1. The detailed parameters of the proposed RCAA-Net.

Layers	Output Size
Convolution layer	$256 \times 256 \times 16$
Residual block1	$128 \times 128 \times 32$
Residual block2	$64 \times 64 \times 64$
Residual block3	$32 \times 32 \times 128$
Feedback self-calibration module 1	$128 \times 128 \times 128$
Feedback self-calibration module 2	$64 \times 64 \times 128$
Feedback self-calibration module 3	$32 \times 32 \times 128$
Self-attention aggregation module	1×128
FC layer	1×256
Softmax layer	1×61

In order to realize the proposed RCAA-Net for end-to-end training, the objective function of this paper adopts Softmax, and its formula is as shown in Equation (7).

$$\updownarrow_{cls} = -\frac{1}{L}\left[\sum_{l=1}^{L}\sum_{k=1}^{K} q(y_l == k) \log \frac{e^{\theta_k^T X_t}}{\sum_{p=1}^{K} e^{\theta_p^T X_t}}\right] \tag{7}$$

where $(X_1, y_1), (X_2, y_2), \ldots, (X_L, y_L)$ is the training set, X_l is the l-th training sample, and $y_l \in 1, 2, 3, \ldots, K$ is the label corresponding to X_l. θ_k^T and θ_p^T denote the transposition of θ_k and θ_p, respectively. L and K denote the number of training samples and the number of categories, respectively. $q(\cdot)$ is the guiding function.

By combining Equations (2) and (7), the final loss function is obtained as follows:

$$\updownarrow_{final} = \updownarrow_c + \updownarrow_{cls} = \|X_{dec} - X_c\|_2 - \frac{1}{L}\left[\sum_{l=1}^{L}\sum_{k=1}^{K} q(y_l == k) \log \frac{e^{\theta_k^T X_t}}{\sum_{p=1}^{K} e^{\theta_p^T X_t}}\right] \tag{8}$$

By minimizing the final loss in Equation (8), the proposed RCAA-Net is trained end-to-end.

4. Experiment

4.1. Experiment Setup

4.1.1. Dataset

The dataset employed in this paper comes from the crop disease detection competition in 2018ai_challenger. The dataset contains 31,718 training images, 4540 verification images, and 4514 testing images, covering different diseases in apples, corn, grapes, citrus, peaches, peppers, potatoes, strawberries, tomatoes and others. Some examples of the dataset are shown in Figure 5. The images in this dataset contain various noises and environmental factors such as angles and lighting. Therefore, the dataset can truly reflect the current status of data resources during crop disease image recognition in the actual environment and is sufficient to verify the accuracy and robustness of the method in this paper for crop disease recognition in actual scenarios.

Figure 5. Some examples of the dataset.

4.1.2. Implementation Details

We verify the RCAA-Net method on the 2018ai_challenger crop disease recognition dataset. The model is trained on a machine with NVIDIA GPU 1080i with 300 epochs. Generally, after 50 iterations of training, RCAA-Net can output satisfactory accuracy. In this paper, the Adam optimization algorithm is used to optimize the loss function of Equation (8), and the initial learning rate is 3×10^{-3}. The batch size is set to 128. In the test phase, in order to prove the anti-interference ability of the proposed RCAA-Net network, we add different levels of Gaussian and salt and pepper noise to the test images to verify the recognition accuracy of the network and evaluate the robustness of the network.

4.1.3. Comparison Methods

In order to verify the effectiveness and superiority of the proposed RCAA-Net, we conducted experiments on the 2018ai_challenger crop disease recognition dataset. Specifically, detailed experiments were conducted to verify the proposed RCAA-Net in terms of accuracy and robustness.

The comparison methods used in the experiment include LeafSnap SVM (RBF) method [30], LeafSnap NN method, HCF SVM (RBF) method [31], HCF-Scale Robust SVM (RBF) method [31], combined linear SVM method [31] and SIFT linear SVM method [32].

Among them, the LeafSnap NN method uses neural networks to classify and recognize gist features. HCF SVM (RBF) classifier leverages SVM (RBF) to classify hand-designed features. Here, the SVM (RBF) method is to apply the radial basis kernel function SVM

classifier to classify the leaf gist features [31]. The SVM classifier was implemented by libsvm [33]. The HCF-Scale Robust SVM (RBF) method extracts the features except for the leaf contour length, area and skeleton length from the HCF features and uses the SVM (RBF) classifier for classifying. The features in the combined linear SVM method include the features extracted by the convolutional neural network ConvNet [34] and the features extracted by the HCF-Scale Robust method. Among them, ConvNet includes 5 convolutional layers, 3 maximum pooling layers and 2 fully connected layers. The SIFT linear SVM method is to extract SIFT features and use a simple linear SVM classification method based on sparse coding linear space pyramid matching SPM kernel for classification and recognition.

In addition, to make a fair comparison, we adopt the proposed RCAA-Net by adopting the input image size 256×256, which matches the input images size of the other comparison methods. The adapted method is noted as RCAA-Net (adaptive).

In order to verify the effectiveness of each module, we design different models. Specifically, a model that does not include residual connections, feedback self-calibration and self-attention aggregation is used as a baseline. On this basis, we have added residual connections to form a comparison method named RES. The feedback self-calibration module is added based on the baseline to form the self-calibration method. The self-attention aggregation module is added on the basis of the baseline to form the self-attention method. The effectiveness of each module is illustrated by comparing each method.

4.2. Comparison with State-of-the-Art Methods

In order to verify the accuracy of the RCAA-Net method in this paper, this paper trains the model on the 2018ai_challenger crop disease recognition dataset. The accuracy change of the model during training is shown in Figure 6. It can be seen that the RCAA-Net proposed in the end-to-end training manner can quickly converge and achieve an ideal recognition accuracy. This shows the efficiency of the proposed RCAA-Net.
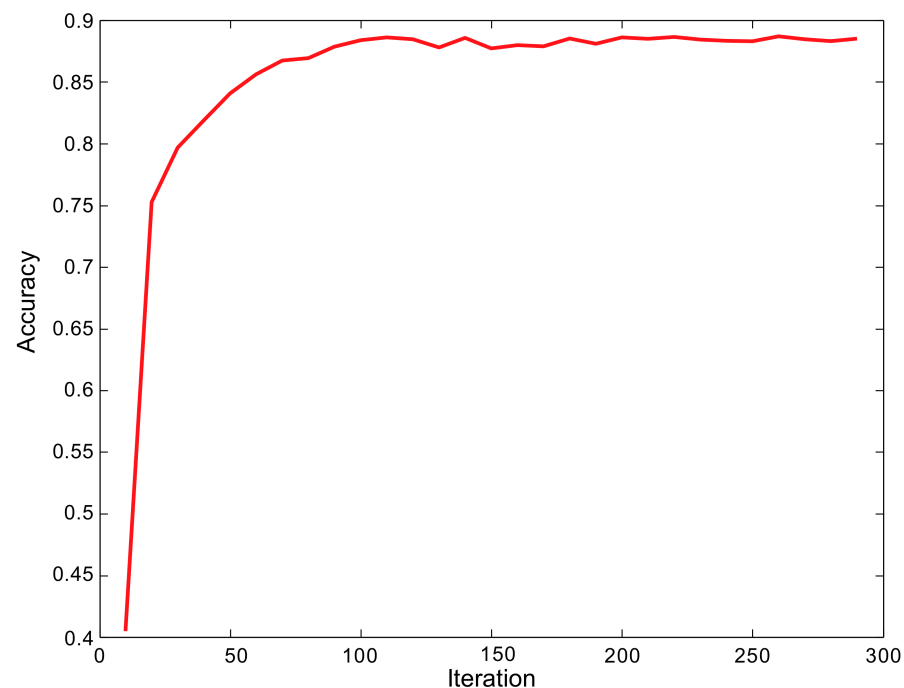

Figure 6. The accuracy curve during the training phase.

In order to more accurately illustrate the superiority of the proposed RCAA-Net method, we have compared the recognition accuracy with the existing methods. The experimental comparison results are shown in Table 2.

Int. J. Environ. Res. Public Health **2021**, *18*, 8404

10 of 14

Table 2. The comparison results with the state-of-the-art methods.

Methods	Accuracy	Image Size
RCAA-Net	0.892	224×224
RCAA-Net (adaptive)	0.895	256×256
LeafSnap SVM (RBF)	0.407	256×256
LeafSnap NN	0.569	256×256
HCF SVM (RBF)	0.676	256×256
HCF-Scale Robust SVM (RBF)	0.625	256×256
Combined linear SVM	0.871	256×256
SIFT linear SVM	0.548	256×256

It can be seen from Table 2 that the RCAA-Net method in this paper is significantly better than other methods except for the combined linear SVM method. Compared with the combined linear SVM method that utilizes more complex networks, the proposed RCAA-Net method still has higher recognition accuracy, which further illustrates the efficiency of the proposed RCAA-Net method in crop disease recognition. The proposed RCAA-Net method only adopts a simple network to achieve end-to-end recognition. The size of the convolutional layer and the number of parameters are small, which effectively reduces the difficulty of model training. In this paper, the network operation parameters are only 45.68% of the ConvNet parameters in the combined linear SVM method. In addition, when identifying crop diseases in the actual environment, due to the limited number of labeled image sets and fewer parameters, the over-fitting problem caused by insufficient training data can be better alleviated by the proposed method. We argue that the main reason for the state-of-the-art performance includes two aspects. On the one hand, we develop a self-attention aggregation module to automatically focus on the discriminative regions by capturing multi-scale information in different semantic spaces, which can effectively make the proposed RCAA-Net more accurate. On the other hand, we develop a feedback self-calibration module for further suppressing the background noise in the original deep features by fine filtering and adjusting the network features, thereby effectively improving the effectiveness of the proposed RCAA-Net. Note that when we adopt more small-input images (224×224), the accuracy of the proposed method had almost no change, but the computational burden was further decreased.

4.3. The Discussion under Different Noise Conditions

In order to prove the effectiveness of various modules in the proposed RCAA-Net method, we perform the proposed RCAA-Net method, baseline model, RES model, self-calibration model and self-attention model on the 2018ai_challenger crop disease recognition dataset. The experimental results are shown in Table 3.

Table 3. The experimental ablation results.

Methods	Accuracy
RCAA-Net	0.892
Baseline	0.617
RES	0.705
Self-calibration	0.684
Self-attention	0.751

From Table 3, it can be seen that on the 2018ai_challenger crop disease recognition dataset, when the general convolutional network CNN is utilized, the recognition accuracy rate was 0.617. When the residual connection, the feedback self-calibration module and self-attention aggregation module were added separately, the accuracy rates were increased to 0.705, 0.684 and 0.751, respectively. Compared with the baseline, RES, self-calibration and self-attention methods, the RCAA-Net method in this paper achieves the highest recognition accuracy rate of 0.892. This fully shows that when combined with the

Int. J. Environ. Res. Public Health **2021**, *18*, 8404

11 of 14

residual connection, the self-calibration module and the self-attention aggregation module can be fed back to bring higher recognition accuracy, which has important enlightening significance and reference value for crop disease recognition and recognition problems in other fields.

Furthermore, in order to verify the robustness of the proposed RCAA-Net method in this paper, we add Gaussian noise, salt and pepper noise, and Gaussian and salt and pepper noise at the same time to the testing set. The proposed RCAA-Net method is evaluated to show the robustness of the model by adding different levels of noise interference. The noise level interval added in the testing set of this experiment is 0.005.

Figures 7–9 respectively show the comparison results of recognition accuracy when adding Gaussian noise, salt and pepper noise, and both Gaussian and salt and pepper noise in the test set of the 2018ai_challenger crop disease recognition dataset. In Figures 7–9, the abscissa represents the added noise level, and the ordinate represents the recognition accuracy obtained by the test. It is obvious from Figures 7–9 that the recognition accuracy of the blue curve (corresponding to the RCAA-Net method) is higher than other methods. In this work, we can also find the performances of the proposed RCAA-Net dropdown more rapidly when the noise levels increase, which is maybe due to the adverse interaction of various modules when combining them. In the future work, we will explore a more effective combination manner among different modules to further improve the performance for crop disease recognition in actual scenarios. Furthermore, when the noise is added, the recognition accuracy will decrease because the addition of noise will affect the network model extracts and effective features of the image lesions, which in turn affects the accurate recognition of the model. In the same way, the recognition accuracy will decrease as the noise level increases. This is because as the level increases, the number of effective pixels in the lesion area that can be extracted will gradually decrease, which makes it difficult for the model to obtain an accurate prediction category.
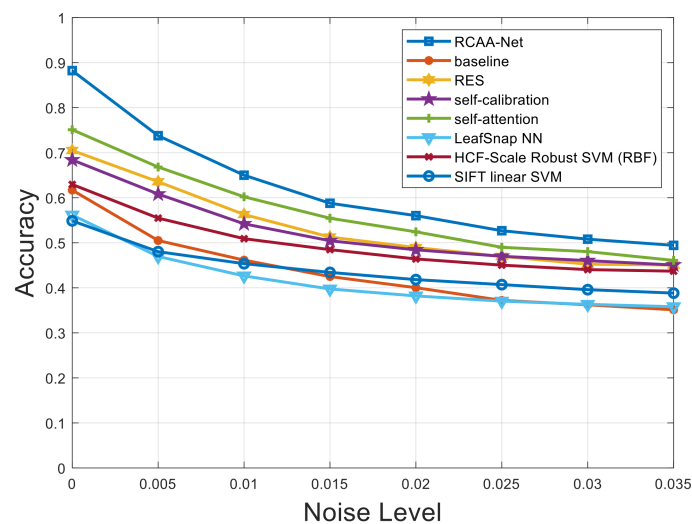

Figure 7. Recognition accuracy results when adding different levels of Gaussian noise.

It can be seen from Figure 7 that when Gaussian noise with a level of 0.005 is added, the recognition accuracy of all methods gets different degrees of attenuation. Under the conditions of different levels of Gaussian noise, in addition to the RCAA-Net method, the self-attention method has the highest recognition accuracy. When other levels of noise are added, it can also be seen that the RCAA-Net method has the highest recognition accuracy. It can be seen that the anti-interference ability of the RCAA-Net method against Gaussian noise is stronger than other comparison methods. Similar conclusions can be drawn from the test results of adding different noises. Therefore, when adding different levels of noise to the testing set of the 2018ai_challenger crop disease recognition dataset, the proposed RCAA-Net in this paper achieves high accuracy and strong robustness.
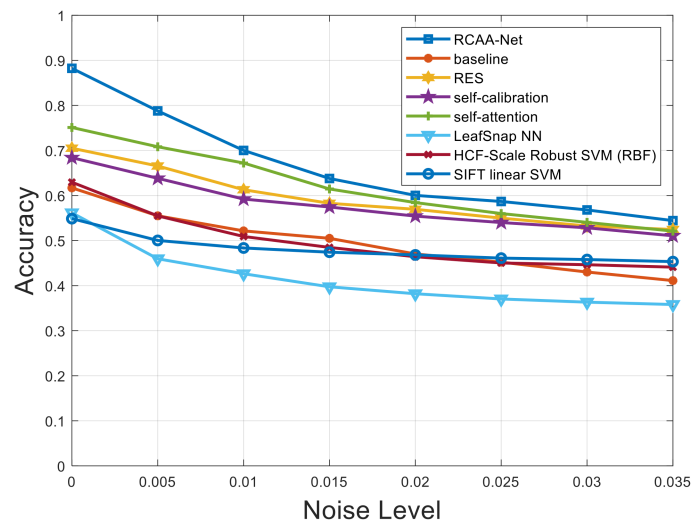
Int. J. Environ. Res. Public Health **2021**, *18*, 8404

12 of 14

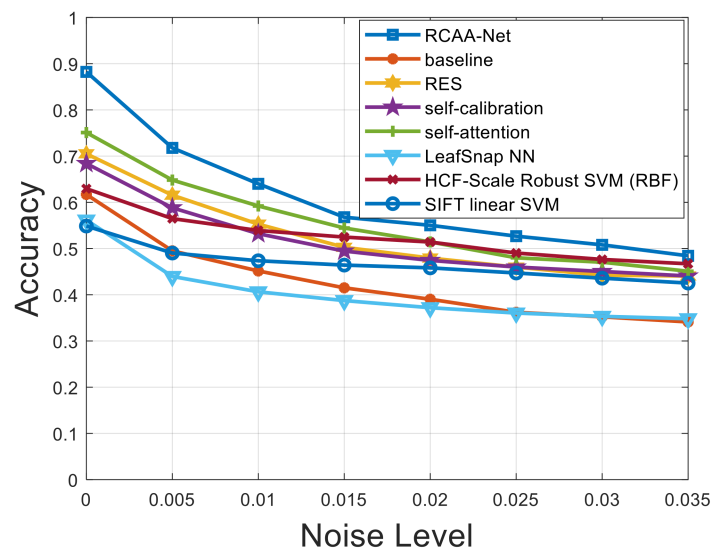**Figure 8.** Recognition accuracy results when adding different levels of salt and pepper noise.

Figure 9. Recognition accuracy results when adding different levels of Gaussian noise + salt and pepper noise.

5. Conclusions

In order to improve the accuracy and robustness of crop disease recognition, this paper introduces a residual self-calibration and self-attention aggregation network (RCAA-Net) for crop disease recognition in actual scenarios. On the one hand, we develop a self-attention aggregation module to automatically focus on the discriminative regions by capturing multi-scale information in different semantic spaces, which can effectively make the proposed RCAA-Net more accurate. On the other hand, we develop a feedback self-calibration module for further suppressing the background noise in the original deep features by fine filtering and adjusting the network features; thereby, effectively improving the effectiveness of the proposed RCAA-Net. Subsequently, in order to verify the proposed RCAA-Net method, this paper carried out corresponding experiments on the 2018ai_challenger crop disease recognition dataset. After a large number of experimental verifications, the proposed RCAA-Net method had higher accuracy and robustness on the same testing dataset. In the next step, we will consider two aspects to further improve the method for crop disease recognition in actual scenarios. Firstly, we plan to add a saliency detection module to the network model to better locate the significant lesion area in the data, further optimize the model structure, and improve the accuracy and robustness of

network recognition. Secondly, we will explore a more effective combination manner among different modules to further improve the performance for crop disease recognition in actual scenarios.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated for this study are available online at https://challenger.ai/competition/pdr2018 (accessed on 5 March 2021).

References

1. Wu, H.; Kang, Z.; Li, X.; Li, Y.; Li, Y.; Wang, S.; Liu, D. Identification of Wheat Leaf Rust Resistance Genes in Chinese Wheat Cultivars and the Improved Germplasms. *Plant Dis.* **2020**, *104*, 2669–2680. [CrossRef] [PubMed]
2. Boulent, J.; Foucher, S.; Théau, J.; St-Charles, P.L. Convolutional neural networks for the automatic identification of plant diseases. *Front. Plant Sci.* **2019**, *10*, 941. [CrossRef] [PubMed]
3. Wang, X.; Liu, J.; Zhu, X. Early real-time detection algorithm of tomato diseases and pests in the natural environment. *Plant Methods* **2021**, *17*, 1–17. [CrossRef] [PubMed]
4. Hughes, D.; Salathé, M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv* **2015**, arXiv:1511.08060.
5. Guettari, N.; Capelle-Laizé, A.S.; Carré, P. Blind image steganalysis based on evidential k-nearest neighbors. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 2742–2746.
6. Deepa, S.; Umarani, R. Steganalysis on Images using SVM with Selected Hybrid Features of Gini Index Feature Selection Algorithm. *Int. J. Adv. Res. Comput. Sci.* **2017**, *8*, 1503–1509.
7. Ramezani, M.; Ghaemmaghami, S. Towards genetic feature selection in image steganalysis. In Proceedings of the 2010 7th IEEE Consumer Communications and Networking Conference, Las Vegas, NV, USA, 9–12 January 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 1–4.
8. Sheikhan, M.; Pezhmanpour, M.; Moin, M. Improved contourlet-based steganalysis using binary ppaper swarm optimization and radial basis neural networks. *Neural Comput. Appl.* **2012**, *21*, 1717–1728. [CrossRef]
9. Kodovsky, J.; Fridrich, J.; Holub, V. Ensemble classifiers for steganalysis of digital media. *IEEE Trans. Inf. Forensics Secur.* **2011**, *7*, 432–444. [CrossRef]
10. Guo, Y.; Hastie, T.; Tibshirani, R. Regularized linear discriminant analysis and its application in microarrays. *Biostatistics* **2007**, *8*, 86–100. [CrossRef] [PubMed]
11. Zhang, S.; Wang, Z. Cucumber disease recognition based on global-local singular value decomposition. *Neurocomputing* **2016**, *205*, 341–348. [CrossRef]
12. Zhang, S.; Wu, X.; You, Z.; Zhang, L. Leaf image based cucumber disease recognition using sparse representation classification. *Comput. Electron. Agric.* **2017**, *134*, 135–141. [CrossRef]
13. Ning, H.; Zhao, B.; Yuan, Y. Semantics-Consistent Representation Learning for Remote Sensing Image-Voice Retrieval. *IEEE Trans. Geosci. Remote. Sens.* **2021**. [CrossRef]
14. Almabadi, E.S.; Bauman, A.; Akhter, R.; Gugusheff, J.; Van Buskirk, J.; Sankey, M.; Eberhard, J. The Effect of a Personalized Oral Health Education Program on Periodontal Health in an At-Risk Population: A Randomized Controlled Trial. *Int. J. Environ. Res. Public Health* **2021**, *18*, 846. [CrossRef] [PubMed]
15. Alsoghair, M.; Almazyad, M.; Alburaykan, T.; Alsultan, A.; Alnughaymishi, A.; Almazyad, S.; Alsuhaibani, M. Medical Students and COVID-19: Knowledge, Preventive Behaviors, and Risk Perception. *Int. J. Environ. Res. Public Health* **2021**, *18*, 842. [CrossRef] [PubMed]

Int. J. Environ. Res. Public Health **2021**, *18*, 8404

14 of 14

16. Duan, C.; Xiao, N. Parallax-based second-order mixed attention for stereo image super-resolution. *IET Comput. Vis.* **2021**. [CrossRef]

17. Xie, X.; Yang, T.; Zhang, Y.; Liang, B.; Liu, L. Accurate localization of moving objects in dynamic environment for small unmanned aerial vehicle platform using global averaging. *IET Comput. Vis.* **2021**. [CrossRef]

18. Hu, J.; Kong, H.; Fan, L.; Zhou, J. Enhancing feature fusion with spatial aggregation and channel fusion for semantic segmentation. *IET Comput. Vis.* **2021**. [CrossRef]

19. Kong, J.; Shen, H.; Huang, K. DualPathGAN: Facial reenacted emotion synthesis. *IET Comput. Vis.* **2021**. [CrossRef]

20. Sohrabi Nasrabadi, M.; Safabakhsh, R. 3D object recognition with a linear time-varying system of overlay layers. *IET Comput. Vis.* **2021**. [CrossRef]

21. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using deep learning for image-based plant disease detection. *Front. Plant Sci.* **2016**, *7*, 1419. [CrossRef] [PubMed]

22. Ma, J.; Du, K.; Zheng, F.; Zhang, L.; Gong, Z.; Sun, Z. A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network. *Comput. Electron. Agric.* **2018**, *154*, 18–24. [CrossRef]

23. Kawasaki, Y.; Uga, H.; Kagiwada, S.; Iyatomi, H. Basic study of automated diagnosis of viral plant diseases using convolutional neural networks. In Proceedings of the International Symposium on Visual Computing, Monte Carlo, Monaco, 14–16 December 2015; Springer: Cham, Switzerland, 2015; pp. 638–645.

24. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.

25. Lin, M.; Chen, Q.; Yan, S. Network in network. *arXiv* **2013**, arXiv:1312.4400.

26. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *PMLR* **2015**, *37*, 448–456.

27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

28. Ding, H.; Jiang, X.; Shuai, B.; Liu, A.Q.; Wang, G. Semantic correlation promoted shape-variant context for segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 8885–8894.

29. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Polosukhin, I. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762.

30. Kumar, N.; Belhumeur, P.; Biswas, A.; Jacobs, D.; Kress, W.; Lopez, I.; Soares, J. Leafsnap: A computer vision system for automatic plant species identification. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 502–516.

31. Hall, D.; McCool, C.; Dayoub, F.; Sunderhauf, N.; Upcroft, B. Evaluation of features for leaf classification in challenging conditions. In Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 5–9 January 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 797–804.

32. Yang, J.; Yu, K.; Gong, Y.; Huang, T. Linear spatial pyramid matching using sparse coding for image classification. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 1794–1801.

33. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1–27. [CrossRef]

34. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]