IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 28, 2020, accepted March 16, 2020, date of publication March 19, 2020, date of current version March 30, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.2982055*

# Crop Leaf Disease Image Super-Resolution and Identification With Dual Attention and Topology Fusion Generative Adversarial Network

**QIANG DAI[1], XI CHENG[2], YAN QIAO[1], AND YOUHUA ZHANG[1]**

[1]School of Information and Computer, Anhui Agricultural University, Hefei 230036, China
[2]School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Corresponding author: Yan Qiao (qiaoyan101@gmail.com)

This work was supported in part by the National Key Research and Development Project of China under Grant 2017YFD0301303, in part by the 2019 National Undergraduate Training Programs for Innovation and Entrepreneurship under Grant 201910364073, in part by the Open Foundation of Key Laboratory of Agricultural E-commerce under Grant AEC2018012, and in part by the Open Foundation of Anhui Key Laboratory of Intelligent Agricultural Technology And Equipment under Grant APKLSATE2019 × 012.

**ABSTRACT** For agricultural disease image identification, obtained images are typically unclear, which can lead to poor identification results in real production environments. The quality of an image has a significant impact on the identification accuracy of pre-trained image classifiers. To address this problem, we propose a generative adversarial network with dual-attention and topology-fusion mechanisms called DATFGAN. This network can effectively transform unclear images into clear and high-resolution images. Additionally, the weight sharing scheme in our proposed network can significantly reduce the number of parameters. Experimental results demonstrate that DATFGAN yields more visually pleasing results than state-of-the-art methods. Additionally, treated images are evaluated based on identification tasks. The results demonstrate that the proposed method significantly outperforms other methods and is sufficiently robust for practical use.

**INDEX TERMS** Crop leaf disease, attention, generative adversarial networks, super-resolution, identification.

## I. INTRODUCTION

Crop disease is a major factor limiting crop cultivation. Crop disease can lead to sharp drops in production, which can lead to huge losses in the agricultural economy. Therefore, early identification of crop disease is critical for the selection of optimal treatments and is an important prerequisite for reducing crop loss and pesticide use. All crops are susceptible to diseases and crop diseases negatively affect yield and quality. However, excessive chemical control can leave drug residues and lead to environmental pollution. Based on improved living standards, the demand for crop quality is greater than ever. Therefore, the early diagnosis and treatment of crop diseases are issues that must be resolved.

In recent years, agricultural disease identification is a hot research topic. Cheng *et al.* [1] used the fine-tuning method to classify and identify agricultural pests disease while using deep convolutional neural networks (DCNN), which can achieve a satisfactory recognition effect. Yue *et al.* [2] developed a super-resolution method for agricultural pests disease restoration and detection. A plant disease diagnosis system was proposed by Kawasaki *et al.* [3] to identify two leaf diseases in cucumber plants by using a CNN. Sun *et al.* [4] improved the traditional AlexNet [5] model by implementing CNN models combining batch normalization and global pooling to identify numerous leaf diseases. These studies demonstrate the feasibility and effectiveness of applying DCNNs in the field of leaf disease identification. However, images obtained from farms are typically unclear. Poor image quality significantly reduces the identification accuracy of pre-trained classifiers, which are typically trained on clear high-resolution datasets.

To improve the accuracy of agricultural disease image classification, low-resolution images must be super-resolved to increase spatial resolution and reconstruct the high-frequency details of sharp edges. In this paper, we propose a generative adversarial network (GAN) with dual-attention and topology-fusion mechanisms to transform low-resolution images captured at farms. The proposed network is called DATFGAN. To evaluate the proposed method, we compare

The associate editor coordinating the review of this manuscript and approving it for publication was Michele Nappi.

it to state-of-the-art methods in terms of classification accuracy when images are transformed. We present experiments using eight classic classification networks and crop leaf disease images with a total of 27 categories as a classification dataset. An example of a crop leaf disease image that was super-resolved using DATFGAN is presented in Figure 1.

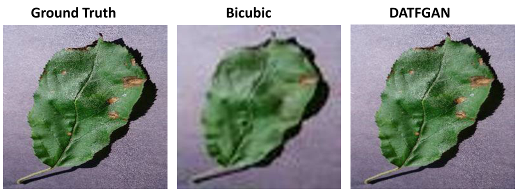



**FIGURE 1.** Super-resolved image generated by DATFGAN (right).

Experimental results demonstrated that classification accuracy can be improved if images are transformed using super-resolution methods. Compared to the state-of-the-art methods considered in this research, DATFGAN provides superior performance with an average accuracy improvement of 3%. Our main contributions can be summarized as follows.

1) We propose a novel image super-resolution method for agricultural disease images.

2) To the best of our knowledge, our method is the first to introduce GANs into agricultural disease image processing.

3) According to benchmark tests, DATFGAN outperforms state-of-the-art methods in terms of visual quality and classification accuracy.

## II. RELATED WORKS

The reliable identification and detection of crop diseases and crop stress are a significant challenge in the agricultural industry [6], [7]. Zhang *et al.* proposed an apple leaf disease recognition method [8] based on image processing techniques and pattern recognition methods. In their experiments, an image dataset of diseased apple leaves containing 90 disease images was considered. Their approach achieved a recognition accuracy greater than 90%. In [9], Waghmare *et al.* focused on a grape plant leaf disease detection system. Their system takes a single leaf as an input and performs segmentation following background removal. Their work focused on major diseases that are commonly observed in grape plants, namely downy mildew and black rot. Their proposed approach achieved an accuracy of 96.6%. Bashish *et al.* developed a fast, automatic, cheap, and accurate image-based method [10] for the identification of leaf diseases. Their method consists of four main phases: a color transformation structure, image segmentation using the K-means clustering technique, calculation of texture features, and a pre-trained neural network. Experimental results demonstrated that their method could successfully detect and classify diseases with an accuracy of approximately 93%. In [11], Arivazhagan *et al.* proposed a software solution for the automatic detection and classification of plant leaf diseases. Their method consists of four main steps. First, a color transformation structure for an input RGB image is created. Second, green pixels are masked and removed using a specific threshold value followed by a segmentation process. Third, texture statistics are computed for useful segments. Fourth, extracted features are passed through a classifier. Their method successfully classified the examined diseases with an accuracy of 94%. However, traditional machine learning methods [12]–[14], [14]–[16] require complicated image preprocessing and classification steps, and cannot match the accuracy provided by deep learning methods based on CNNs.

Over the past few years, to improve crop management and health, many researchers have studied crop disease identification based on deep learning methods. Sladojevic *et al.* [17] proposed a novel approach to the development of plant disease recognition models based on leaf image classification using deep convolutional networks. Their model could recognize 13 different types of plant diseases and had the ability to distinguish plant leaves from their surroundings. Experimental results for their developed model demonstrated precision values between 91% and 98% for separate class tests with an average value of 96.3%. Amara *et al.* developed a deep-learning-based approach [18] that automates the process of classifying banana leaf diseases. They made use of the LeNet [19] architecture as a CNN to classify image datasets. Their preliminary results demonstrated the effectiveness of deep learning approaches, even under challenging conditions, such as illumination variation, complex backgrounds, and different resolutions, sizes, poses, and orientations of real images. Mohanty *et al.* [20] used a public dataset of 54,306 images of diseased and healthy plant leaves collected under controlled conditions to train a DCNN to identify 14 crop species and 26 diseases (or the absence thereof). Their trained model achieved an accuracy of 99.35% on a held-out test set. Using a dataset of cassava disease images captured in the field in Tanzania, Ramcharan *et al.* [21] applied transfer learning to train a DCNN to identify three diseases and two types of pest damage (or the lack thereof). Their best-trained model accuracies were 98% for brown leaf spot, 96% for red mite damage, 95% for green mite damage, 98% for cassava brown streak disease, and 96% for cassava mosaic disease. Ferentinos [22] trained models using an open database of 87,848 images containing 58 combinations of plants or diseases. Several model architectures were trained, including an AlexNet [5], VGG [23] and GoogLeNet [24], and the greatest accuracy achieved for plant disease identification was 99.53%.

According to these studies, CNNs are useful tools in the field of crop disease identification and satisfactory results have been obtained. However, the crop disease images obtained from farms are typically unclear low-resolution images, which have a significant impact on the improvement of crop disease image identification accuracy. Therefore, it is crucial to improve crop disease images using super-resolution methods. We propose a GAN with dual-attention and topology-fusion mechanisms to transform low-resolution images obtained from farms.

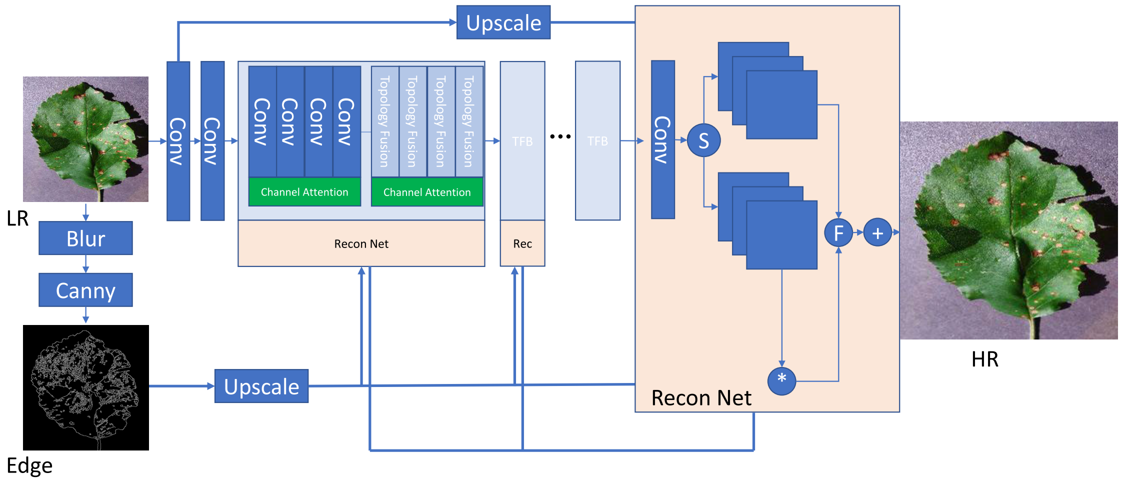



**FIGURE 2.** Generator network.

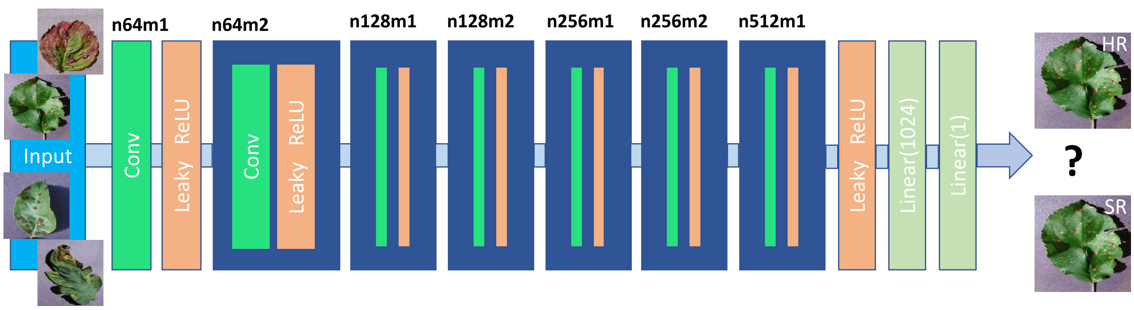



**FIGURE 3.** Discriminator network.

## III. PROPOSED METHOD

We detail our network architecture in Section III-A. We then introduce two attention mechanisms called channel attention and texture attention in Section III-B. Finally, we define adversarial training in Section III-C.

### A. NETWORK ARCHITECTURE

We divide this section into three parts to describe the overall network architecture (Section III-A1), parameter sharing (Section III-A2), and topology fusion (Section III-A3). First, we describe the overall architecture of DATFGAN. Second, we introduce parameter sharing operations that are used in the generator network. Third, discuss how to take advantage of both residual and dense connections.

#### 1) OVERALL ARCHITECTURE

DATFGAN can be divided into a generator and discriminator. Figure 2 presents the generator network of DATFGAN, which contains three basic components: a shallow feature extraction network, parameter-sharing attention-enhanced topology-fusion network, and reconstruction network. The shallow feature extraction network used for topology fusion contains two convolutional layers for extracting shallow features from the generator network. Low-resolution images are used as inputs for the generator network and divided into two branches. One branch feeds into an upscaling module after the first convolutional layer in the generator network. The other branch feeds into the topology fusion network to predict details following the second convolutional layer. The reconstruction network exploits global residual learning [25] and combines upscaled images with predicted details to generate high-resolution images.

The discriminator network is presented in Figure 3. The discriminator network is trained to solve a maximization problem. It contains seven convolutional layers with an increasing number of filter kernels (increasing by factors of 2 from 64 to 512 kernels, similar to the VGGNet [23]). Striding convolutions are used to reduce the image resolution each time the number of features is doubled. The resulting 512 feature maps are fed into a final LeakyReLu activation function and two linear layers to increase the probability of sample classification.

#### 2) PARAMETER SHARING

A convolution operation extracts local information and some statistical characteristics of local information may be the same as those of other local information, meaning the features learned through convolution operations can also be used for other information. Therefore, the same learning features can be reused for multiple positions in the image. In a CNN, a convolution kernel (filter) is used to extract a feature (one dimension of input data). If the input data have multiple features (dimensions), there will be many convolution kernels, leading to parameter explosion in the convolution layer. Additionally, each convolution kernel in a layer that extracts specific features while ignoring the local correlations between data.

In parameter sharing, each feature has translation invariance, meaning the same feature can appear in different positions in different data and the same convolution kernel can be used to extract this feature from multiple positions. Furthermore, based on the local correlations between data, by performing weight sharing, a convolutional layer can share its convolution kernel, which reduces the number of parameters in the convolutional layer. By increasing the number of layers in a deep neural network, each convolutional layer can use a different convolution kernel to extract as many features as possible. Therefore, we use parameter-sharing attention-enhanced topology-fusion networks in the generator network of DATFGAN to reduce the number of network parameters and probability of overfitting, and to make a deeper structure trainable.

#### 3) TOPOLOGY FUSION

ResNet [25] was proposed to solve the problem of degradation in deep learning. When the number of layers in a model increases, the error rate decreases. The degradation problem is closely related to optimization. When the structure of a model becomes increasingly complex, optimization becomes increasingly difficult, resulting in unsatisfactory learning results. The residual block in ResNet [25] was implemented using residual connections. The input and output of the block were added element-wise through the residual connections. This simple form of addition does not add any extra parameters or calculations to the network, but it can significantly increase the training speed of the model, thereby improving the overall effectiveness of training. When the number of layers in the model increases, this structure can also solve the degradation problem.

ResNet [25] can train deep CNNs by establishing residual connections between front and back layers, which aids with the back-propagation of gradients during training. The basic idea of DenseNet [26] is the same as ResNet [25], but dense connections are established between all previous layers and latter layers. DenseNet [26] achieves better performance than ResNet [25] with fewer parameters and lower computational cost. Compared to ResNet [25], DenseNet [26] uses a more aggressive and dense connection mechanism of connecting all layers, meaning DenseNet [26] performs direct concatenation of feature maps from different layers, which can improve feature reuse.

To take advantage of both residual and dense connections, we combined both connections types in a single layer. Compared to residual networks, our proposed generator can preserve more information from previous states, providing our network with contiguous memory. Compared to dense networks, our proposed structure can reduce the channel growth rate by half. This significantly reduces the number of network parameters and makes deeper structures trainable. Additionally, this topology can enhance the flow of information and gradients. Figure 4 presents the inner structure of the proposed mixed-link connections. The operator M in Figure 4 denotes a mixed-link operation, which yields a fusion of residual and dense connections between the current layer and previous layer. Mixed-link operations can be calculated using Formulas 1–3.

$$F_{i-1}^1, F_{i-1}^2 = Slice(F_{i-1}), \tag{1}$$

In Formula 1, the input channels are sliced into two equal part. Slice(.) denotes a slice operation in Formula (2). Assuming that the feature map of an input $F_{i-1}$ has N channels, the output feature maps $F_{i-1}^1$ and $F_{i-1}^2$ will contain N/2 channels after the slicing operation.

$$F_i^1, F_i^2 = Slice(W(F_{i-1}) + b), \tag{2}$$

In Formula 2, the output of one layer or unit is sliced into two equal parts in the channel dimension. In this formula, W denotes the weight of a convolutional layer and b denotes the bias.

$$F_{i+1} = C(C(F_i^1 + F_{i-1}^2, F_i^2), F_{i-1}^1) \tag{3}$$

In Formula 3, C(.) denotes a concatenation operation, $F_{i-1}^1$ and $F_{i-1}^2$ are the sliced parts of the features from the previous layer, and $F_i^1$ and $F_i^2$ are the sliced parts from the current layer. The addition of $F_i^1$ and $F_{i-1}^2$ makes the topology residual and the concatenation of $F_i^1 + F_{i-1}^2$, $F_i^2$, and $F_{i-1}^1$ makes the topology dense. These formulas make our network partially residual [25] and partially dense [26].

At the end of each block, we use a transition convolution to reshape features to their original size, as shown in Formula 4, where $W_t$ denotes the weight of a $1 \times 1$ convolution for block-feature fusion, which can reduce the number of channels. $F_{j-1}$ denotes the features of the preceding mixed-link block and $F_j$ denotes the output features of the current mixed-link block. Based on this mixed-link mechanism, our network can synchronously generate both residual and dense connections, which decreases parameter growth and improves network performance.

$$F_j = W_t(F_{j-1}) + b, \tag{4}$$

### B. DUAL ATTENTION

We propose channel attention in Section III-B1 and texture attention in Section III-B2. Both of these mechanisms are used for improving the effectiveness of transforming images.

#### 1) CHANNEL ATTENTION

Channel attention is used in our topology fusion network to model the interdependencies of convolution channels, which

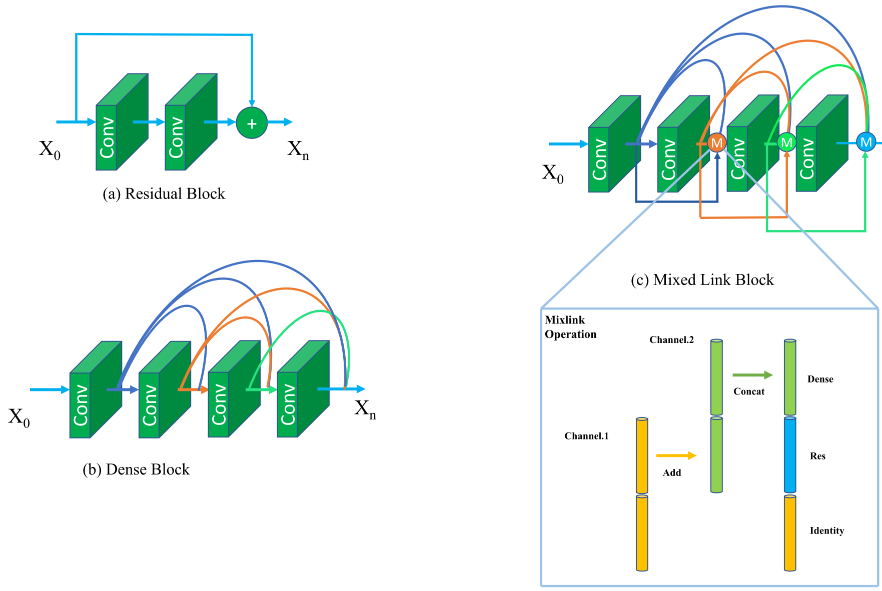



**FIGURE 4.** Topology fusion.

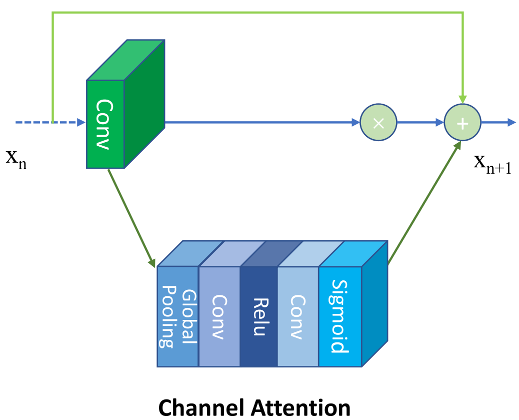



**FIGURE 5.** Channel attention.

can be learned autonomously to boost important channels and suppress useless channels. This mechanism behaves like a filter to recalibrate the information and gradient flow among networks. As shown in Figure 5, the channel attention module consists of a global average pooling layer, which squeezes features spatially to extract global information from channels. Next, two $1 \times 1$ convolutions generate a bottleneck. Finally, a Sigmoid layer is used to normalize information and the outputs are used to reweight the original outputs to generate self-trained channel-wise attention. Channel attention operates according to Formulas 5 and 6:

$$S(F) = \frac{1}{HW} \sum_{i}^{H} \sum_{j}^{W} F(i,j), \tag{5}$$

where $S(.)$ is a squeeze operation that pools the features in each channel into a global mean, and $H$ and $W$ denote the height and width of the input feature map, respectively.

$$A(F) = \delta(W_u \sigma(W_d S(F))) * F, \tag{6}$$

where $A(.)$ denotes the channel attention function, $\sigma$ denotes the ReLU function, and $W_u$ and $W_d$ denote two $1 \times 1$ convolutions. $W_d$ first reduces the channels to 1/16th of their original size, then $W_u$ expands the tensor to the original shape, which forms a bottleneck. $\delta$ denotes the sigmoid function, which normalizes the weights for each channel to values between zero and one. We use these weights to boost useful information and suppress useless information.

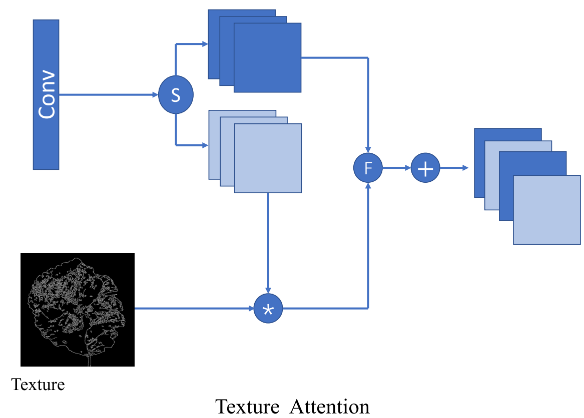



**FIGURE 6.** Texture attention.

### 2) TEXTURE ATTENTION

As shown in Figure 6, texture is a very important feature in plant images and is very useful for image super-resolution tasks. Additionally, the high-frequency details of an image are typically located around edges, meaning it is important to assign attention with guidance from edges. Therefore, we use texture attention in our reconstruction network.

We utilize edges as global spatial attention components for image reconstruction according to Formula 7-8, where $W_{exp}$ represents expanding the original number of channels. In this process, the number of global features is doubled. Half of the channels are weighted using global information and the other half retains local information. The two halves are summed and averaged to fuse global and local information.

$$F_i^1, F_i^2 = Slice(W_{exp}(F_{i-1})), \tag{7}$$

$$F_{i+1} = Up(Canny(F_0)) * F_i^1 + F_i^2, \tag{8}$$

As shown in Formula 8, Up denotes an up-sampling operation and Canny denotes an operator for extracting edge features. $F_0$ denotes the initial input features. The resulting edge features following up-sampling are multiplied by half of the initial input features using large-scale pixel maps to guide smaller maps. The result of this operation is then added to the other half of the initial input features to perform fusion.

Figure 7 presents edge features obtained from an RGB image processed by the Canny operator. We also present colored edge features for the sake of clarity.

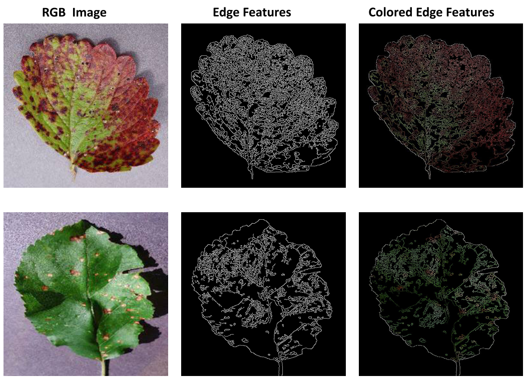



**FIGURE 7.** Edge features.

### C. ADVERSARIAL TRAINING

Adversarial training is based on the concept of confrontation. Two modules can achieve similar teaching and learning goals through confrontational learning. Adversarial training is a method for smoothing the loss function of a landscape and is designed to make the output of a model smoother. We did not expect this method to improve the generalization ability of our model, but there are potential simplifications that can be applied to real data to make prediction functions operate more smoothly and simply.

This method operates based on the premise that similar inputs should yield similar outputs. If a small perturbation is introduced into an input, the corresponding output should also change very little. Therefore, the most effective way for implementing this type of regularization is to force different inputs to yield the same output. One can then attempt to identify the disturbance that generates the greatest loss to construct a confrontation sample and minimize the cross entropy between the output and ground truth.

Instead of directly optimizing the mean squared error between input images and targets, we introduced adversarial training to generate more visually pleasing images. Adversarial loss is defined by the following formula:

$$L_{GAN} = \mathbb{E}[D(G(I_{LR}))] - \mathbb{E}[D(I_{HR})], \tag{9}$$

where D(.) denotes the discriminator of DATFGAN and G(.) denotes the generator. $I_{LR}$ denotes generated pseudo-high-resolution images and $I_{HR}$ denotes real-world high-resolution images.

After incorporating adversarial loss, the total loss can be represented as follows:

$$L = \alpha L_{GAN} + L_{content}, \tag{10}$$

where L is the total loss and $L_{GAN}$ is the adversarial loss. $L_{content}$ denotes the total perceptual loss for the target content. We set $\alpha$ to 0.01 in this work.

## IV. EXPERIMENTS

Our experiments consisted of four stages: experimental setup (Section IV-A), dataset management (Section IV-B), training DATFGAN (SectionIV-C), and comparing DATFGAN to state-of-the art methods (SectionIV-D). In the first stage, we defined hardware and software environments. In the second stage, we collected data for training DATFGAN and performing classification. In the third stage, we trained DATFGAN using the collected data. In the final stage, we transformed images using different super-resolution methods and compared the results in terms of image classification accuracy.

### A. EXPERIMENTAL SETUP

We trained the proposed network using a computer equipped with the hardware and software listed in Table 1. Pytorch was used as a framework for constructing the network and CUDA was adopted for acceleration.

**TABLE 1.** Experiment setup.

| Hardware | Software |
|---|---|
| CPU: 8 Cores | Windows10 |
| RAM: 32 GB DDR4 | CUDA10.0 + CUDNN7.0 |
| GPU: NVIDIA RTX2080Ti (11GB GDDR6) | Pytorch1.0.1 + Python 3.7 |

### B. DATASETS

We used the DIV2K dataset [27] for pre-training the proposed super-resolution model. We used bicubic interpolation to down-sample images and we added additive Gaussian noise to the low-resolution images to create clear and unclear image pairs. We also used 1350 crop leaf disease images from the Plant Disease Recognition Competition of the AI Challenger 2018. These images include 27 different categories and each category contains 50 images. We refer to

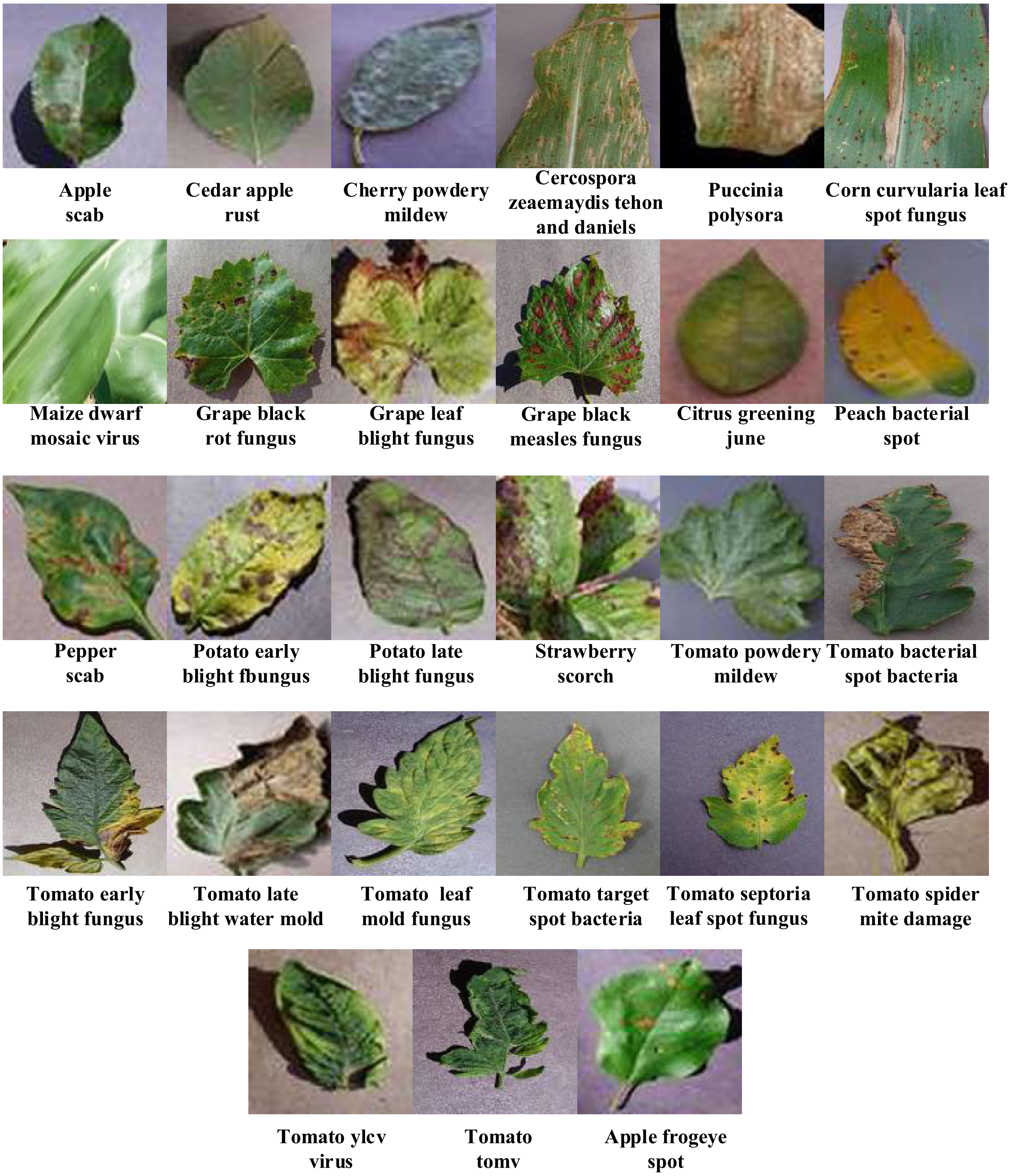



**FIGURE 8.** CLDI dataset.

this dataset as the CLDI dataset. The CLDI dataset contains both crop leaf disease images of different species and different disease images for the same species, which increases the difficulty of classification and reduces the potential for bias.

We selected 40 images from each category to train the classification network models and 10 images to test the classification models. All images were preprocessed for better training and testing results. We randomly rotated and flipped the images and used batch normalization for data augmentation. The CLDI dataset is presented in Figure 8. Additionally, Table 2 lists the categories and numbers of images in the CLDI dataset.

### C. TRAINING DETAILS AND PARAMETERS FOR DATFGAN

We trained DATFGAN on an NVIDIA RTX2080Ti GPU using the DIV2K dataset [27] and used bicubic interpolation to down-sample the images. We also added additive Gaussian noise to the low-resolution images to create clear and unclear image pairs. We randomly rotated and flipped the images for data augmentation. For optimization, we used the RMSProp optimizer to minimize the loss

**TABLE 2.** CLDI dataset.

| Name | Amount |
|---|---|
| Apple Scab | 50 |
| Potato Late Blight Fungus | 50 |
| Cedar Apple Rust | 50 |
| Strawberry Scorch | 50 |
| Cherry Powdery Mildew | 50 |
| Tomato Powdery Mildew | 50 |
| Cercospora Zeaemaydis Tehon and Daniels | 50 |
| Tomato Bacterial Spot Bacteria | 50 |
| Puccinia Polysora | 50 |
| Tomato Early Blight Fungus | 50 |
| Corn Curvularia Leaf Spot Fungus | 50 |
| Tomato Late Blight Water Mold | 50 |
| Maize Dwarf Mosaic Virus | 50 |
| Tomato Leaf Mold Fungus | 50 |
| Grape Black Rot Fungus | 50 |
| Tomato Target Spot Bacteria | 50 |
| Grape Leaf Blight Fungus | 50 |
| Tomato Septoria Leaf Spot Fungus | 50 |
| Grape Black Measles Fungus | 50 |
| Tomato Spider Mite Damage | 50 |
| Citrus Greening June | 50 |
| Tomato YLCV Virus | 50 |
| Peach Bacterial Spot | 50 |
| Tomato Tomv | 50 |
| Pepper Scab | 50 |
| Apple Frogeye Spot | 50 |
| Potato Early Blight Fungus | 50 |
| Total | 1350 |

function and trained the model for 200 epochs. We set the initial learning rate to 0.0001 and reduced the rate after every 60 epochs. The momentum was set to 0.9 and the weight decay was set to 0.0001. We used 64 images as the mini-batch size to feed into the model. We also pre-trained our discriminative model using a VGG19 model trained in Pytorch to perform initialization and avoid undesired local optima.

### D. COMPARISON TO STATE-OF-THE ART METHODS

We divided this stage into two phases of visual result inspection (Section IV-D1) and image classification (Section IV-D2). In Section IV-D1, we present super-resolution results and the ground truth images of crop leaf disease images. In Section IV-D2, we describe the training details and experimental results of images classification.

#### 1) VISUAL RESULT INSPECTION

We compared our final models to state-of-the-art peak signal-to-noise ratio (PSNR)-oriented super-resolution methods, including Biubic, SRResNet [28], EDSR [29], SRDenseNet [30], VDSR [31] and LapSRN [32], using the CLDI dataset.

Because there is no effective standard metric for perceptual quality, we present representative qualitative results in Figure 9. PSNRs and structural similarity indexes are also provided for reference. In Figure 9, one can see that DATFGAN outperforms previous approaches in terms of both sharpness and details. For example, DATFGAN can produce sharper and more natural crop leaf disease image textures compared to state-of-the-art PSNR-oriented super-resolution methods, which tend to generate blurry results with unnatural and noisy textures. Furthermore, previous PSNR-oriented super-resolution methods sometimes introduce unpleasant artifacts. DATFGAN eliminates such artifacts and produces natural results. As a result, DATFGAN can reconstruct the detailed appearances of lesions more accurately than state-of-the-art PSNR-oriented super-resolution methods and improve classification accuracy.

#### 2) IMAGE CLASSIFICATION

We selected AlexNet [5], VGG-16 [23], Inception-v3 [33], ResNet-101 [25], Resnext50 [34], DenseNet-121 [26], MobileNet V2 [35], and ShuffleNet V2 [36] as classification networks. During the process of training these classification networks, we retained most of the weights in the original models and only trained softmax layers. We used 1080 images from the CLDI dataset to train each model and 270 images from the CLDI dataset to test each model. Adam was used as an optimizer and cross entropy was used as a loss function. Additional training details could be found in Table 3.

**TABLE 3.** Training details.

| Method | Learning rate | Batch size | Epochs |
|---|---|---|---|
| Alexnet | 0.0001 | 20 | 50 |
| VGG16 | 0.0001 | 20 | 50 |
| Inception-v3 | 0.0001 | 20 | 65 |
| Resnet101 | 0.0005 | 20 | 50 |
| Resnext50 | 0.0005 | 20 | 60 |
| Densenet121 | 0.0005 | 20 | 50 |
| MobileNet V2 | 0.0001 | 20 | 60 |
| ShuffleNet V2 | 0.0001 | 20 | 70 |

Figure 10 presents the classification accuracies for images transformed by different super-resolution methods and raw images. In Figure 10, one can see that classification accuracy can be improved if images are transformed using super-resolution methods. Compared to the state-of-the-art methods tested in this experiment, DATFGAN performs better and improves classification accuracy to a greater extent than the other super-resolution methods, particularly for the ResNet-101 [25] and DenseNet-121 [26] classifiers. Crop leaf disease image classification results are listed in Table 4.

## V. DISCUSSION

This paper proposed a super-resolution method for increasing the spatial resolution of crop leaf disease images. According to our experiments, the classification accuracy for images transformed by super-resolution methods is greater than that for low-resolution images obtained from farms. This is because images transformed by super-resolution method, can convey much more information, such as details regarding lesions, compared to low-resolution images. Our experiments on CLDI dataset classification clearly demonstrate this phenomenon because using the low-resolution images resulted in lower accuracy. These results indicate

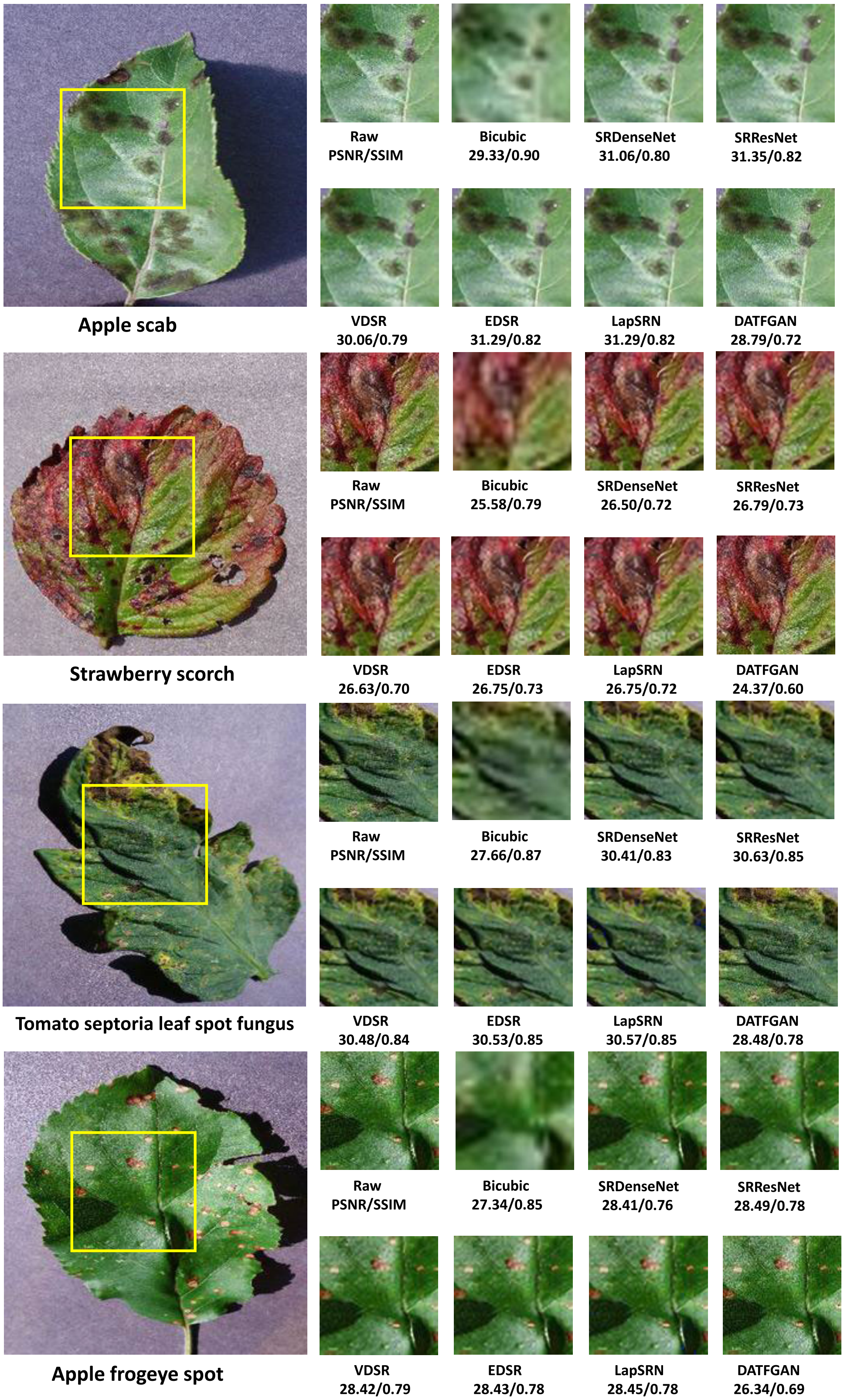



**FIGURE 9.** Visual results.

that the super-resolution methods successfully reconstructed the detailed appearances of lesions and facilitated the identification of diseases. We compared DATFGAN to state-of-the-art methods and determined that DATFGAN improves classification accuracy more than the other methods.

**TABLE 4.** Classification accuracy.

| Method | Accuracy(%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Raw | SRResNet | EDSR | LapSRN | VDSR | SRDenseNet | DATFGAN(Ours) |
| Alexnet | 85.92 | 86.35 | 86.48 | 86.29 | 87.03 | 86.29 | 88.14 |
| VGG16 | 85.55 | 86.26 | 86.26 | 85.92 | 86.66 | 86.29 | 88.51 |
| Inception-v3 | 87.78 | 88.53 | 88.44 | 88.57 | 88.74 | 88.41 | 89.67 |
| Resnet101 | 88.88 | 89.62 | 89.70 | 89.62 | 90.00 | 89.62 | 91.48 |
| Resnext50 | 87.77 | 88.41 | 88.61 | 88.82 | 88.41 | 88.61 | 89.63 |
| Densenet121 | 88.51 | 89.25 | 89.60 | 89.62 | 89.25 | 89.25 | 92.59 |
| MobileNet V2 | 90.74 | 91.34 | 91.37 | 91.37 | 91.44 | 91.77 | 92.73 |
| ShuffleNet V2 | 89.63 | 90.52 | 90.56 | 90.89 | 90.52 | 90.53 | 91.41 |

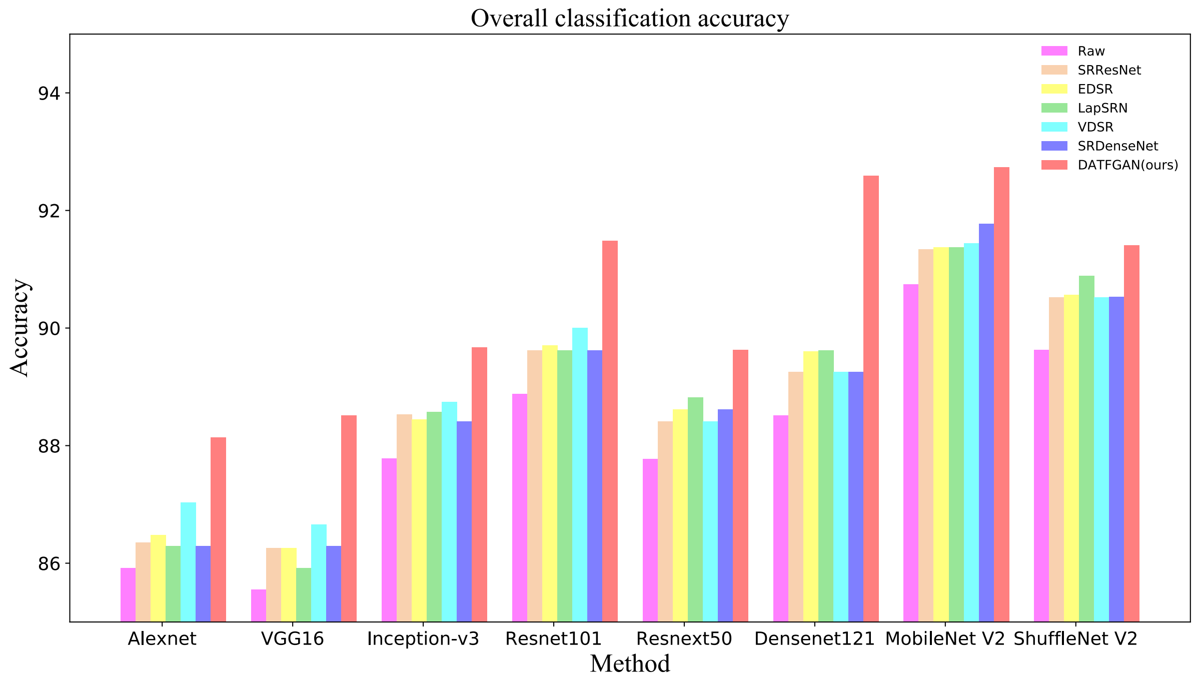



**FIGURE 10.** Overall classification accuracy.

DATFGAN is a GAN with dual-attention and topology-fusion mechanisms. To make use of both residual and dense connections, we implemented both types of connections in a single layer. Compared to residual networks, the generator of DATFGAN can preserve more information from previous states, which enables our network to maintain contiguous memory. Compared to dense networks, DATFGAN can reduce the channel growth rate by half, which significantly decreases the number of network parameters, making deeper structures trainable. We used two attention mechanisms called channel attention and texture attention. Channel attention can model the interdependencies of convolution channels, which can be learned autonomously to boost important channels and suppress useless channels. Texture attention can assign attention based on guidance from edges, meaning we can utilize edges as global spatial attention mechanisms for image reconstruction. Additionally, DATFGAN can effectively transform unclear images into clear and high-resolution images. Furthermore, the parameter sharing mechanism in our proposed network can significantly reduce the number of parameters. Although DATFGAN has many advantages over previous methods, there are still some limitations that must be overcome. Traditional deep-learning-based super-resolution methods typically use average pixel positions, making images overly smooth, but increasing PSNR. DATFGAN does not use averaging, resulting in better visual effects, but reducing PSNR compared to other super-resolution methods.

One of the most important techniques in deep learning, including CNNs, is transfer learning, which is also as known as fine-tuning. In this study, we slightly modified state-of-the-art network architectures to avoid image size reduction and produce RGB images directly. In future studies, we will conduct network training using larger image datasets, such as ImageNet [37], and evaluate the resulting classification performance.

## VI. CONCLUSION

In this paper, we proposed a novel image restoration method for crop leaf disease images. To the best of our knowledge, our method is the first to introduce GANs into agricultural disease image processing. We took advantage of both residual and dense connections to reduce the number of network parameters significantly and make deeper structures trainable. Furthermore, a dual-attention mechanism provided a significant performance boost. Channel attention can boost important channels and suppress useless channels. Texture attention can assign attention based on the texture features and utilize textures as global spatial attention mechanisms during image reconstruction. According to our experimental results, DATFGAN outperforms state-of-the-art methods in terms of both visual quality and classification performance. Based on topology fusion and effective attention mechanisms, DATFGAN can not only enhance classification accuracy, but also reduce the number of network parameters, making it very practical for real-world applications.

## REFERENCES


[1] X. Cheng, Y. Zhang, Y. Chen, Y. Wu, and Y. Yue, "Pest identification via deep residual learning in complex background," *Comput. Electron. Agricult.*, vol. 141, pp. 351–356, Sep. 2017.

[2] Y. Yue, X. Cheng, D. Zhang, Y. Wu, Y. Zhao, Y. Chen, G. Fan, and Y. Zhang, "Deep recursive super resolution network with Laplacian pyramid for better agricultural pest surveillance and detection," *Comput. Electron. Agricult.*, vol. 150, pp. 26–32, Jul. 2018.

[3] Y. Kawasaki, H. Uga, S. Kagiwada, and H. Iyatomi, "Basic study of automated diagnosis of viral plant diseases using convolutional neural networks," in *Proc. Int. Symp. Vis. Comput.* Berlin, Germany: Springer, 2015, pp. 638–645.

[4] J. Sun, W. Tan, H. Mao, X. Wu, Y. Chen, and L. Wang, "Recognition of multiple plant leaf diseases based on improved convolutional neural network," *Trans. Chin. Soc. Agricult. Eng.*, vol. 33, no. 19, pp. 209–215, 2017.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[6] S. Savary, A. Ficke, J.-N. Aubertot, and C. Hollier, "Crop losses due to diseases and their implications for global food production losses and food security," *Food Secur.*, vol. 4, no. 4, pp. 519–537, Dec. 2012.

[7] E.-C. Oerke, "Crop losses to pests," *J. Agricult. Sci.*, vol. 144, no. 1, pp. 31–43, Feb. 2006.

[8] Z. Chuanlei, Z. Shanwen, Y. Jucheng, S. Yancui, and C. Jia, "Apple leaf disease identification using genetic algorithm and correlation based feature selection method," *Int. J. Agricult. Biol. Eng.*, vol. 10, no. 2, pp. 74–83, 2017.

[9] H. Waghmare, R. Kokare, and Y. Dandawate, "Detection and classification of diseases of grape plant using opposite colour local binary pattern feature and machine learning for automated decision support system," in *Proc. 3rd Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Feb. 2016, pp. 513–518.

[10] D. Al Bashish, M. Braik, and S. Bani-Ahmad, "Detection and classification of leaf diseases using $K$-means-based segmentation and," *Inf. Technol. J.*, vol. 10, no. 2, pp. 267–275, 2011.

[11] S. Arivazhagan, R. N. Shebiah, S. Ananthi, and S. V. Varthini, "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features," *Agricult. Eng. Int., CIGR J.*, vol. 15, no. 1, pp. 211–217, 2013.

[12] M. Kulin, T. Kazaz, I. Moerman, and E. De Poorter, "End-to-End learning from spectrum data: A deep learning approach for wireless signal identification in spectrum monitoring applications," *IEEE Access*, vol. 6, pp. 18484–18501, 2018.

[13] Y. Zhang, R. Gravina, H. Lu, M. Villari, and G. Fortino, "PEA: Parallel electrocardiogram-based authentication for smart healthcare systems," *J. Netw. Comput. Appl.*, vol. 117, pp. 10–16, Sep. 2018.

[14] X. Li, X. Cheng, W. Chen, G. Chen, and S. Liu, "Identification of forested landslides using LiDar data, object-based image analysis, and machine learning algorithms," *Remote Sens.*, vol. 7, no. 8, pp. 9705–9726, 2015.

[15] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, no. 3, pp. 657–682, Mar. 2014.

[16] J. Wäldchen and P. Mäder, "Machine learning for image based species identification," *Methods Ecology Evol.*, vol. 9, no. 11, pp. 2216–2225, Nov. 2018.

[17] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," *Comput. Intell. Neurosci.*, vol. 2016, Jun. 2016, Art. no. 3289801.

[18] J. Amara, B. Bouaziz, and A. Algergawy, "A deep learning-based approach for banana leaf diseases classification," in *Proc. BTW Workshops*, 2017, pp. 79–88.

[19] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[20] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers Plant Sci.*, vol. 7, pp. 1419, Sep. 2016.

[21] A. Ramcharan, K. Baranowski, P. McCloskey, B. Ahmed, J. Legg, and D. P. Hughes, "Deep learning for image-based cassava disease detection," *Frontiers Plant Sci.*, vol. 8, p. 1852, Oct. 2017.

[22] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Comput. Electron. Agricult.*, vol. 145, pp. 311–318, Feb. 2018.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[24] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[26] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[27] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 126–135.

[28] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.

[29] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 136–144.

[30] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4799–4807.

[31] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.

[32] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 624–632.

[33] X. Xia, C. Xu, and B. Nan, "Inception-v3 for flower classification," in *Proc. 2nd Int. Conf. Image, Vis. Comput. (ICIVC)*, Jun. 2017, pp. 783–787.

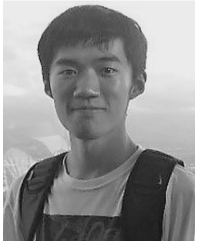
[34] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1492–1500.

[35] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[36] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet V2: Practical guidelines for efficient cnn architecture design," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 116–131.

[37] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

**QIANG DAI** was born in Anhui, China, in 1997. He is currently pursuing the B.S. degree with the School of Information and Computer, Anhui Agricultural University. His research interests include computer vision, deep learning, pattern recognition, and super resolution.



**XI CHENG** received the B.S. degree in computer science and technology from Anhui Agricultural University, Hefei, China. He is currently pursuing the Ph.D. degree in control science and engineering with the Nanjing University of Science and Technology, Nanjing, China. He is also a Research Assistant with the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education and the Jiangsu Key Laboratory of Image and Video Understanding for Social Security, Nanjing, China. His research interests include pattern recognition, object detection, image demoireing, and super resolution.

**YAN QIAO** received the Ph.D. degree in computer science from the Beijing University of Posts and Telecommunications, in 2012. She was a Postdoctoral Fellow with the School of Computer Engineering, Nanyang Technological University, Singapore. She is currently a Visiting Scholar with the Department of Computer Science, University of Victoria, Canada. She is also an Associate Professor with the School of Information and Computer, Anhui Agricultural University, China. Her current research interests focus on network monitoring and anomaly detection for the Internet of Things.

**YOUHUA ZHANG** received the B.S. and M.S. degrees from the Hefei University of Technology, Hefei, China, and the Ph.D. degree from the Institute of Intelligent Machine, Chinese Academics of Sciences, Hefei. From 2009 to 2019, he was a Professor with Anhui Agricultural University. His research interests include computer vision, agricultural informatics, and computer graphics.