IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 9, 2020, accepted April 27, 2020, date of publication April 30, 2020, date of current version May 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2991552

Agricultural Pest Super-Resolution and Identification With Attention Enhanced Residual and Dense Fusion Generative and 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: YouHua Zhang (zhangyh@ahau.edu.cn)

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, and in part by Anhui Provincial Natural Science Foundation.

ABSTRACT The growth of the most significant field crops such as rice, wheat, maize, and soybean are influenced because of various pests. And crop production is decreased due to various categories of insects. Deep learning technologies significantly increased the efficiency of identifying and controlling agricultural pests attack. However, agricultural pests images obtained are often obscure and unclear because of the sparse density of cameras deployed in the real farmland. This always makes pests difficult to recognize and monitor. Additionally, the existing classification and segmentation methods are not satisfying for the identification of low-resolution images because they are pre-trained on the clear and high-resolution datasets. Therefore, it is crucial to restore and upscale the obtained low-resolution pest images in order to improve classification accuracy and the recall rate of the instance segmentation. In this paper, we propose a generative adversarial network (GAN) with quadra-attention and residual and dense fusion mechanisms to transform low-resolution pest images. Compared with previous state-of-the-art PSNR-oriented super-resolution methods, our proposed method is more powerful in image reconstruction and achieves the state of the art performance. The experiment results show that after reconstructing with our proposed gan, the recall rate increased by 182.89% and classification accuracy also improved a lot. Besides, our proposed method could decrease the density of the camera layout in the agricultural Internet of Things (IOT) monitor systems and the cost of infrastructure, which is practical for real-world applications.

INDEX TERMS Agricultural pests, super-resolution, classification, object instance segmentation, deep learning, quadra-attention, residual and dense fusion.

I. INTRODUCTION

The production of crops is associate with many factors, for example, climate change, plant diseases, and insect pests. According to recent researches, about half of the crop yield in the world is lost to pest infestations and crop diseases [1]. Crop pests cause significant damage to crops and mainly affect the productivity of crop yield, whether in developing or developed countries. Hence, it is of great significance to identify insects in the crops at an early stage and select optimal treatments, which is an important prerequisite for reducing crop loss and pesticide use. There are too many types of insects and the number of individuals which belongs to the same species is enormous. However, traditional pest identification of insects is typically time-consuming and inefficient. Therefore, in order to improve the efficiency of agricultural production, a new effective recognition method should be proposed.

Nowadays, with the development of deep learning, many researchers apply this technology into different fields and many excellent approaches have been proposed. Because of the successful applications of deep learning in diverse areas, it also has been used in agriculture. Thenmozhi and Reddy [2] proposed a CNN model and used it to identify

The associate editor coordinating the review of this manuscript and approving it for publication was Chang-Hwan Son.

three pest datasets and the highest classification accuracy of 96.75%, 97.47%, and 95.97% was achieved in their proposed CNN model for three datasets respectively. A fine-tuned GoogLeNet model was proposed by Li *et al.* [3] to recognize their collected pest dataset and obtained an improvement of 6.22% compared to the state-of-the-art method. Tetila *et al.* [4] presented an analysis of the network weights for the automatic recognition of soybean leaf diseases applied to images taken straight from a small and cheap unmanned aerial vehicle (UAV). They evaluated four deep neural network models trained with different parameters for fine-tuning (FT) and transfer learning. They tested the data set created from real flight inspections in an end-to-end computer vision approach and results suggested that their method substantially improved the identification accuracy. A new method of recognizing apple leaf diseases through the region-of-interest-aware deep convolutional neural network was proposed by Yu and Son [5] Two subnetworks were first designed in their method. One is for the division of the input image into three areas: background, leaf area, and spot area indicating the leaf diseases. The experimental results proved that correct recognition accuracy can be increased using the predicted region-of-interest(ROI) feature map. It is also shown that the proposed method obtains better performance than the conventional state-of-the- art methods: transfer-learning-based methods, bilinear model, and multiscale-based deep feature extraction and pooling approach. Cheng *et al.* [6] used a fine-tuning method to classify and identify a 10-classes pest dataset with deep convolutional neural networks (DCNNs), which achieved satisfactory recognition results. Yue *et al.* [7] proposed a super-resolution method for agricultural pest image restoration and detection and also gained a high detection result. These previous works demonstrate the feasibility and effectiveness of applying deep learning in the field of pest identification. However, pest images obtained from real farms are typically unclear and very small in pixel because the arrangement of the cameras in the farmland is relatively sparse. Although the existing classification and segmentation methods are very mature, the classification and segmentation for low-resolution and small pixel scale pest images still could not reach a satisfactory result. Poor image quality significantly reduces the classification and segmentation results of pre-trained classifiers and segmentation systems, which are typically trained on clear high-resolution datasets.

In order to enhance the pest classification and segmentation systems, low-resolution pest images need to be upscaled to increase spatial resolution and reconstruct the high-frequency details of images. Therefore, in this paper, we propose a generative adversarial network (GAN) with quadra-attention, residual and dense fusion mechanisms to transform low-resolution pest images. The proposed network is named PSRGAN. And for the sake of evaluating our PSRGAN, we compare it to state-of-the-art super-resolution methods in terms of classification accuracy and instance segmentation recall rate when images are upscaled. We present experiments using eight classic classification networks with the pest images of Xie1 [8] dataset[1] and Xie2 [9] dataset.[2] Besides, we use the Mask RCNN [10] as the object instance segmentation model and choose Diostrombus, Chauliops, and Callitettix as our research objects.

Experimental results testify that classification accuracy and segmentation recall rate could be improved if images are transformed using super-resolution methods. And in comparison with the state-of-the-art super-resolution methods considered in this research, PSRGAN provides superior performance in improving pest image classification accuracy and segmentation recall rate. Our main contributions can be summarized as follows.

1) We propose a novel image super-resolution method for upscaling agricultural pest images.
2) To the best of our knowledge, our proposed method is the first to introduce GANs into agricultural pest image processing.
3) According to benchmark tests, PSRGAN outperforms state-of-the-art super-resolution methods in terms of visual quality, improving classification accuracy and segmentation recall rate.

II. RELATED WORKS

Previously, plenty of automatic pest recognition systems based on machine learning (ML) have been proposed. Wen *et al.* [11] proposed an effective feature-based insect automatical classification for orchard insects using six machine learning algorithms. And five common pest species in orchards were used to verify the classification method. The maximum classification accuracy of 89.5% was observed. Faithpraise *et al.* [12] demonstrated the combination of the k-means clustering algorithm and the correspondence filter to achieve pest detection and recognition. The detection used the relative filter to identify different types of pests, which is time-consuming and ineffective when the dataset is huge. An automatic identification system designed by Wang *et al.* [13] to identify insect specimen images at the order level. The system adopted artificial neural networks (ANNs) and support vector machine (SVM) as the pattern recognition methods for the identification tests and the system performed with good stability and accuracy reached 93%. Xie *et al.* [8] developed an insect recognition system using advanced multiple-task sparse representation and multiple-kernel learning (MKL) techniques, which could combine multiple features of insect species to enhance the recognition performance. And their experimental results on 24 common pest species they collected outperformed other state-of-the-art methods of the generic insect categorization. Traditional machine learning algorithms [14]–[18] were proved to perform well when the number of pest species was small, but they cannot match the accuracy provided by deep learning methods when multiple features need to be extracted manually.

[1] http://www2.ahu.edu.cn/pchen/web/insectRecognition.htm.
[2] http://www2.ahu.edu.cn/pchen/web/DLFautoinsects.htm.

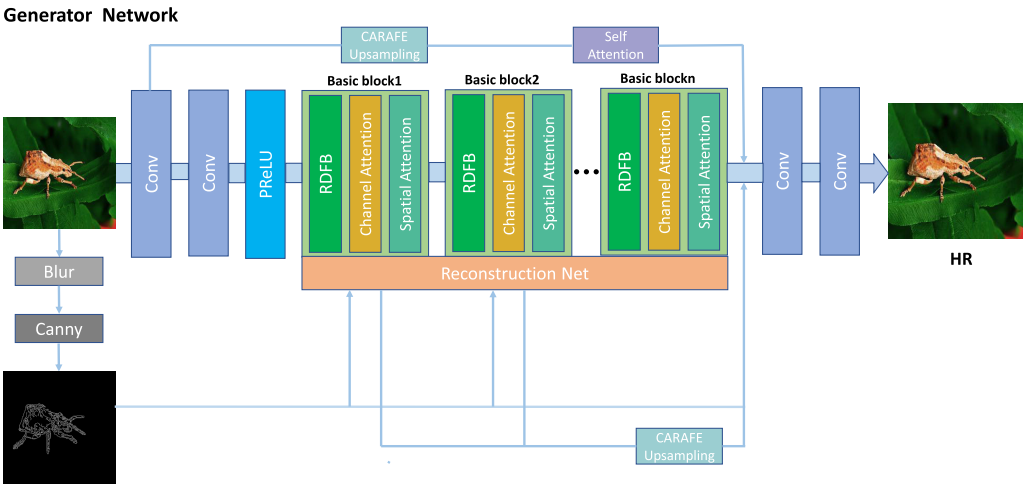

FIGURE 1. Generator network of PSRGAN.

In recent years, to improve crop management and health, many deep learning methods have been applied to identify pests and achieved state-of-the-art results. Shijie *et al.* [19] proposed a detection algorithm on leaf images and constructed the convolution neural network model to detect tomato pests and diseases based on VGG16 [20] and transfer learning. The detection model achieved an average classification accuracy of 89%. In [21], an 8-layer CNN network was developed for the visual localization and classification of agricultural pest insects by computing a saliency map and applying deep convolutional neural network (DCNN) learning. And achieved a mean Accuracy Precision (mAP) of 0.951 for the classification of 12 important paddy insect species. Dawei *et al.* [22] put forward a diagnostic system based on transfer learning for pest detection and recognition, which achieved an accuracy of 93.84% and the performance of their proposed method is comparable to human experts and the traditional neural network. They also used their model to recognize two types of weeds and achieved an accuracy of 98.92%. In [23], the deep convolutional neural network (DCNN) was used in their study to classify crop pests image. On the ground of their experiments, in which LeNet-5 and AlexNet were used to classify pests image. Furthermore, 82 common pest types have been classified, with the accuracy reaching 91%, which proves that their proposed method is not only feasible but preeminent.

According to these previous works, deep learning is a useful tool in the field of pest identification. Nevertheless, the pest images captured from farmland are typically low-resolution and very small in pixels, which have a bad impact on the improvement of pest image identification accuracy. In order to enhance pest segmentation and recognition systems, low-resolution pest images need to be upscaled. Thereby, a generative adversarial network (GAN) with quadra-attention and residual and dense fusion mechanisms is proposed to upscale low-resolution pest images.

III. PROPOSED METHOD

A. NETWORK ARCHITECTURE

1) OVERALL ARCHITECTURE

Our proposed PSRGAN consists of a generator and discriminator. The generator network of our PSRGAN is shown in Figure 1, low-resolution images were put into the generator network and divided into two branches. One was put into the CARAFE [24] upscale module after the first convolution layer in the generator network and then, this branch went through the self-attention module. The other was fed into the reconstruction net to predict the details after went through the second convolution layer and PReLU activation layer. And the reconstruction net used the global residual learning and combined the upscaled images and edges with the predicted details before a convolution layer to generate the high-resolution images.

To discriminate real high-resolution(HR) images from generated super-resolution(SR) images, we design a discriminator network which is illustrated in Figure 2. We use LeakyReLU activation and avoid max-pooling throughout the network. The discriminator network is trained to solve the maximization problem. It is made up of seven convolutional layers with an increasing number of filter kernels, increasing by a factor of 2 from 64 to 512 kernels. Stridden convolutions are used to reduce the image resolution each time the number of features are doubled. The resulting 512 feature maps are followed by a final LeakyReLu activation function and two linear layers to obtain the probability for sample classification.

2) RESIDUAL AND DENSE FUSION

With the deepening of the network, the phenomenon that the accuracy of the training dataset decreases and the error of rate rises is the problem of degeneration. It stands to reason that a deeper model should not have a higher error rate than its shallower model. This is not due to overfitting, but because when the model is complex, SGD optimization becomes more

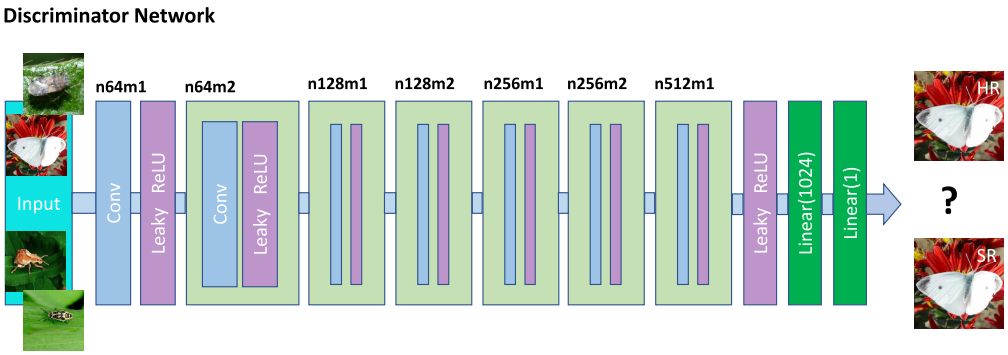

FIGURE 2. Discriminator network of PSRGAN.

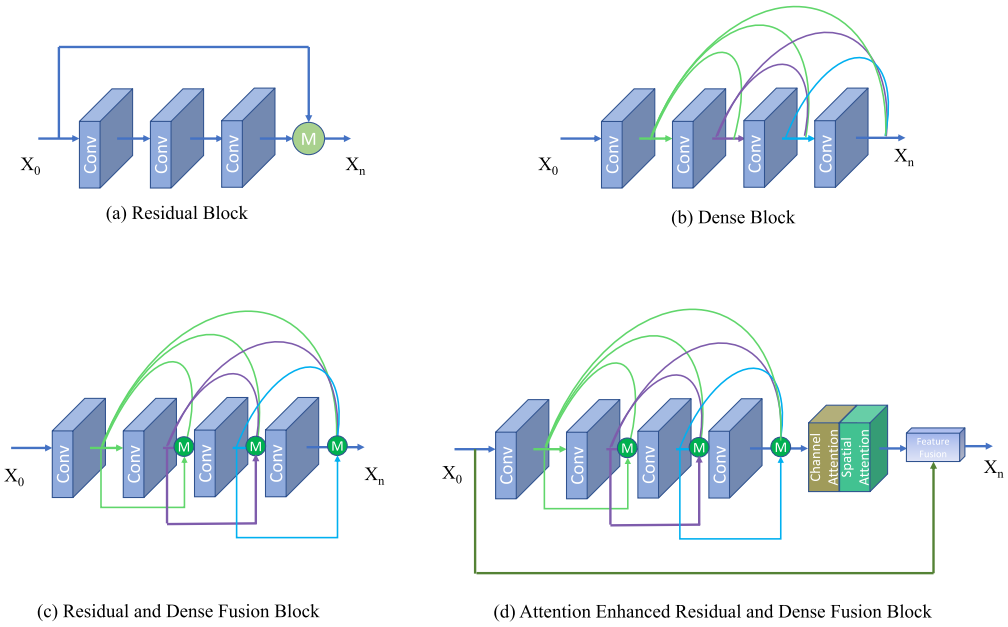

FIGURE 3. Comparison of different network structures. M denotes the mix operation.

difficult, resulting in a model that is not good. ResNet [25] was proposed to solve the problem of degradation in deep learning. The residual block in ResNet [25] which is shown in Figure 3 (a) was implemented by the residual connection. The input and output of the block were added by element-wise by the residual connection. This simple addition does not add extra parameters and calculations, but it could greatly increase the training speed of the model, and improve the training effect. When the number of layers of the model is increased, the residual block can solve the degradation problem well.

Compared to ResNet [25], DenseNet [26] proposed a more aggressive dense connection mechanism that connects all layers, and each layer will use all the layers in front of it as its additional input. In DenseNet [26], each layer will concat with all the previous layers in the channel dimension and serve as the input for the next layer. Due to the dense connection method, DenseNet [26] improves the gradient backpropagation, making the network easier to train. The dense block in DenseNet [26] which is shown in Figure 3 (b) was implemented by the dense connection. Since each layer can directly reach the final error signal, implicit deep supervision is realized. In addition, DenseNet [26] realizes feature reuse and uses a small growth rate. The unique feature map of each layer is relatively small, which makes the parameters smaller and the calculation more efficient.

To make good use of the residual and dense connections, we combine them in a block named residual and dense fusion block, which is shown in Figure 3 (c). Compared with residual networks, the generator of our PSRGAN could preserve more information from the previous layers. Compared with dense networks, our proposed generator could decrease the channel growth rate by half which significantly decreased the network parameters and made a deeper structure trainable. And the fusion operation could be calculated as Formula 1.

$\odot$ means the concatenate operation, F_{i-1}^{1} and F_{i-1}^{2} denote the sliced parts of the features from the previous layer.

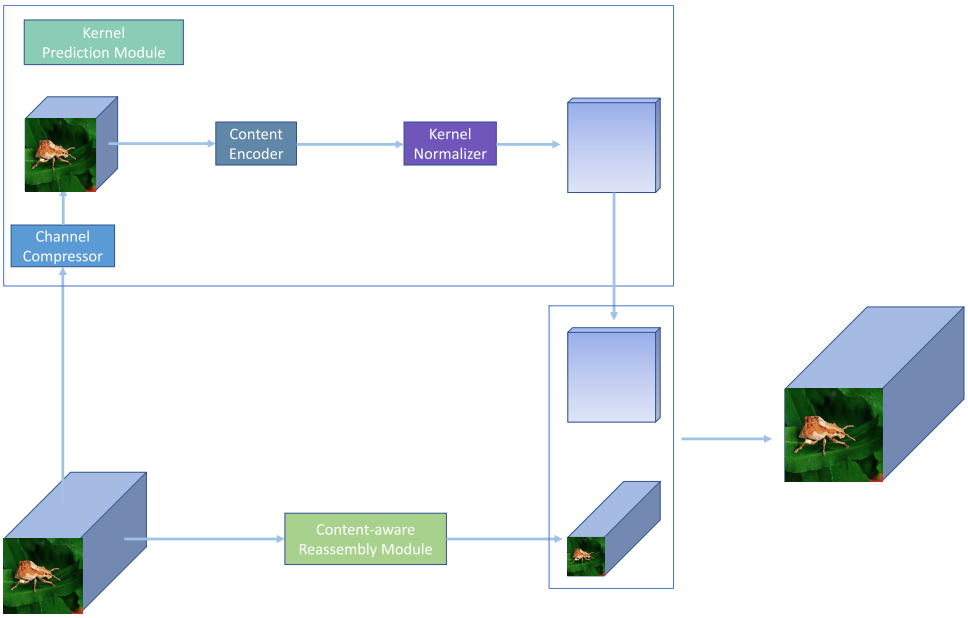

FIGURE 4. Overall framework of CARAFE.

F_i^1 and F_i^2 mean the sliced parts of the features from the current layer. The addition between F_i^1 and F_{i-1}^2 makes the residual fusion and the concatenation between $F_i^1 + F_{i-1}^2, F_i^2$ and F_{i-1}^1 make the dense fusion. These operations could make our network become a partial residual network [25] and dense network [26]. At the end of each block, we use a convolution layer to reshape the feature map to the original size, which could be calculated as Formula 2, where W_t means the weight of a 1×1 convolution for block-feature fusion, F_{j-1} denotes the feature of the proceeding residual and dense fusion block and F_j is the output feature of the current residual and dense fusion block. What's more, with the help of the residual and dense fusion, our network could synchronously do both residual and dense connections at the same time, which decreases half of the network growth parameters and improves network performance.

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

$$F_j = W_t(F_{j-1}) + b, \tag{2}$$

3) CARAFE UPSAMPLING

The upsampling operation can be expressed as a dot product of the upsampling kernel at each location and the pixels in the corresponding neighborhood in the input feature map, which is called feature recombination. The nearest neighbor or bilinear upsampling determines the upsampling kernel only by the spatial position of the pixels and does not use the semantic information of the feature map. It can be regarded as a uniform upsampling, and the perception domain is usually small. The upsampling kernel of the deconvolution operator is not calculated from the distance between pixels but learned through the network. The same upsampling kernel is applied to each position of the feature map, and the information of the feature map content cannot be captured. A large number of parameters and calculations are introduced, especially when the size of the upsampling kernel is large. Dynamic filter predicts a different set of upsampling kernels for each position of the feature map, but the amount of parameters and calculations is more explosive, and it is recognized that it is difficult to learn.

Compared with the previous upsampling operator, the CARAFE [24] operator which is shown in Figure 4 can have a larger receptive field during recombination and will guide the recombination process according to the input characteristics. At the same time, the entire operator is relatively lightweight. Specifically, the CARAFE [24] operator first uses the input feature map to predict the upsampling kernel, and the upsampling kernel at each position is different, and then performs feature reassembly based on the predicted upsampling kernel. In different tasks, CARAFE [24] has achieved significant improvements, while bringing only a small amount of additional parameters and calculations.

$$W_l = \psi(N(\chi_l, k_{encoder})) \tag{3}$$

$$\chi_l' = \phi(N(\chi_l, k_{up}), W_l) \tag{4}$$

CARAFE [24] consists of two steps, the first step is to predict a reassembly kernel for each target location according to its content, and the second step is to reassemble the features with predicted kernels. The first step could be shown in Formula 3, the kernel prediction module ψ predicts a location-wise kernel W_l for each location l which is based on the neighbor of χ_l. The second step is shown in Formula 4, ϕ denotes the content-aware reassembly module, which reassembles the neighbor of χ_l and with the kernel W_l. We use the CARAFE [24] upsampling operator in

the generator of our PSRGAN without introducing too many parameters and calculations and get good results in segmentation tasks.

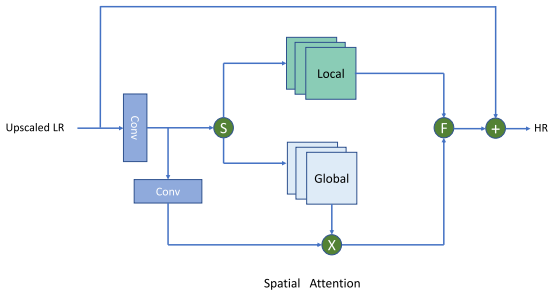

FIGURE 5. Overall framework of spatial attention. S means split operation, ⊗ denotes add operation, F means fusion operation and ⊕ denotes element-wise addition.

B. QUADRA ATTENTION

1) SPATIAL ATTENTION

We use a spatial attention module [27] in the generator network of our proposed PSRGAN, which could help enhance the details in the reconstructed process and improve the effect of residual and dense fusion. As is shown in Figure 5, the spatial attention module uses the global residual learning and the features of the global residual learning are increased to twice the original. Half of the channels after getting through the convolutional layer are weighted by the global information and the other half retains the original information. Then, both of them make global and local fusion, which improves the quality of the high-resolution image generated by the reconstruction net. The following formulas denotes the process of the spatial attention module.

$$T_1, T_2 = S(Q(x)) \tag{5}$$

$$I_{HR} = \sum_{n=b}^{N} \omega^b \frac{1}{2}(T_1 + T_2 P(Q(x))) + U(I_{LR}) \tag{6}$$

where b denotes the basic blocks, ω^b means the weight for the images and S denotes a slice operation. T_1 and T_2 mean the features after the slice operation, P denotes the channel squeeze convolution, Q means the channel multiplier convolution. U means the subpixel shuffler [28] for upscaling which increases the width and height of I_{LR} to the desired size.

2) CHANNEL ATTENTION

The channel attention module [27] we used in the generator could be learned by itself to enhance the important channels and suppress the useless channels, which could help decrease the parameters of our proposed PSRGAN and make the network easier to converge. As is shown in Figure 6, the feature maps will be squeezed into a global pooling layer after passing through the first convolution layer. After that, two 1×1 convolution layers generate a bottleneck. In the end, a Sigmoid layer is used to normalize the information

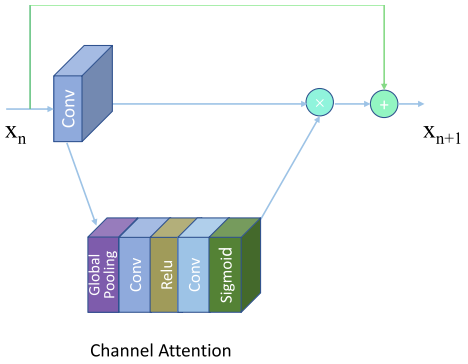

FIGURE 6. Overall framework of channel attention. ⊗ denotes dot product operation and ⊕ denotes element-wise addition.

and generates an output, which is used to reweight the original output to generate a self-learned channel-wise attention. The whole process could be calculated as the following formulas:

$$S(F) = \frac{1}{HW} \sum_{m}^{H} \sum_{n}^{W} F(m, n), \tag{7}$$

where $S(.)$ is the squeeze operation, which squeezes the features in every channel to a global mean, H and W mean the height and width of the input.

$$C(F) = \Phi(W_b \Gamma(W_a S(F))) * F, \tag{8}$$

where $C(.)$ means the channel attention operation, Γ denotes the ReLU function, W_b and W_a mean two 1×1 convolution. W_a first decreases the channels to 1/16 of the original feature map, then W_b expands the feature map to the original shape which generates a bottleneck, Φ denotes the sigmoid function which normalizes the weights for each channel. We use these weights to improve the useful channels and suppress the useless channels.

What's more, as is shown in Figure 3(d), we use the spatial attention module and the channel attention module in the residual and dense fusion block to build the basic block of the generator of our PSRGAN.

3) TEXTURE ATTENTION

The high-frequency details of an image usually located on the edges, thus it is important to give and attention with the guidance of edges and we use texture attention in the reconstruction network. Figure 7 shows the texture attention module we used in the generator network, which could be calculated as the Formula 9-10.

$$F_j^1, F_j^2 = Divide(E(F_{j-1})), \tag{9}$$

$$F_{j+1} = Up(Canny(F_0)) * F_j^1 + F_j^2, \tag{10}$$

As is shown in Formula 9, E denotes expanding the original number of channels and the initial input features are increased to twice the original and the expanded features are divided into two parts. As is shown in Formula 10, Up means upsampling operation and Canny denotes an operator of extracting

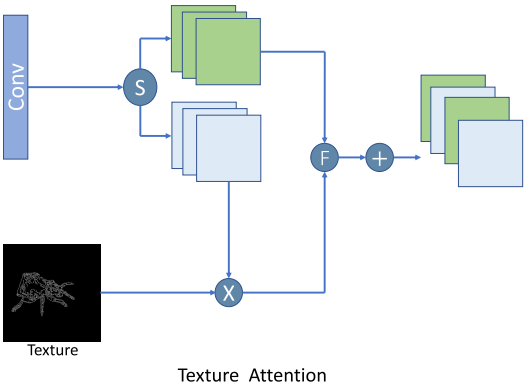

FIGURE 7. Overall framework of texture attention. S means split operation, ⊗ denotes add operation, F means fusion operation and ⊕ denotes element-wise addition.

edge features, F_0 means the initial input features. The result of edge features after upsampling operation times half of the initial input features and uses the large scale of the pixel maps to guide the small maps and then, adds with the other half of the initial input features to sum and average as the fusion.

4) SELF-ATTENTION

The traditional super-resolution model is easy to learn texture features, but it is not easy to learn specific structural and geometric features. In a convolutional neural network, the size of each convolution kernel is very limited, so each convolution operation can only cover a small area around the pixel. It is not easy to capture the features which are far away, because the multi-layer convolution and pooling operations will make the height and width of the feature map smaller and smaller. Self-attention [29] obtains the global geometric features of the image in one step by directly calculating the relationship between any two pixels in the image, so the self-attention mechanism could learn the dependencies among the global features well.

Figure 8 shows the self-attention module for the generator of our proposed PSRGAN. The feature maps are increased to twice the original before passing through the three 1×1 convolutional layers. Then, half of the feature maps are fed into the ConvQ layer and the other half are put into the ConvK layer. And the extra layer of convolution maps added by the initial feature maps are put into the ConvL layer, which helps the self-attention module learn more parameters. After feature maps in the ConvQ layer are transposed, they make the matrix multiplication and softmax operation with feature maps in the ConvK layer to get the attention maps. Finally, feature maps in the ConvL layer make the matrix multiplication with the attention maps and get through the 1×1 convolutional layer to obtain the self-attention maps. In fact, self-attention could be seen as a feature map multiplied by its own transposition, so that the pixels at any two positions have a direct relationship, which can help learn the dependency relationship between any two pixels to get global features. And the following formulas could explain the intact process.

$$f(x) = W_q x, g(x) = W_k x, \tag{11}$$

$$\alpha_{j,i} = \frac{exp(m_{ij})}{\sum_{i=1}^{N} exp(m_{ij})}, m_{ij} = f(x_i)^T g(x_j), \tag{12}$$

$$h(x_i) = W_h x_i, u(x_i) = W_u x_i, \tag{13}$$

$$O = u(\sum_{i=1}^{N} \alpha_{j,i} h(x_i)), \tag{14}$$

The feature maps from the previous layer X are transformed into two features f, g to calculate the attention. And $\alpha_{j,i}$ indicates the extent to which the model attends to the i^{th} location when synthesizing the j^{th} region. N denotes the number of feature locations of features from the previous layer and the output of the attention layer is O. In the above formulas, W_q, W_k, W_h, W_u are the learned weight matrices, which are implemented as 1×1 convolution layers.

C. ADVERSARIAL TRAINING

Adversarial training is an important way to enhance the robustness of deep neural networks. In the process of adversarial training, the samples will be mixed with some small perturbations, and then the neural network will adapt to this change, thus being robust to the adversarial samples. In the process of training our proposed PSRGAN, we use the adversarial training for generating more visually pleasing images instead of straightforward MSE loss between the input images and the output. The following formula shows the adversarial loss.

$$L_{GAN} = \mathbb{E}[D(G(I_{LR}))] - \mathbb{E}[D(I_{HR})], \tag{15}$$

where G(.) means the generator of our proposed gan and D(.) denotes discriminator. And I_{HR} denotes real-world high-resolution images and I_{LR} means the generated pseudo high-resolution images. And the total loss of our network could be calculated in Formula 16.

$$L = \psi L_{GAN} + L_{content} \tag{16}$$

where L is the total loss and L_{GAN} denotes the adversarial loss while the $L_{content}$ means the total perceptual loss for the content. And we set ψ to 0.01 in this work

IV. EXPERIMENTS

A. EXPERIMENT SETUP

The experiment was performed on a server with a 8 cores CPU which was accelerated by an NVIDIA RTX2080Ti GPU. NVIDIA RTX2080Ti has 4,352 CUDA cores and 11 GB memory and the core frequency is up to 1545 MHz. Pytorch was used as the framework to build the network and additional configuration parameters are listed in Table 1.

B. DATASETS

In this experiment, we first used DIV2K [30] dataset to train our proposed super-resolution model. We used bicubic

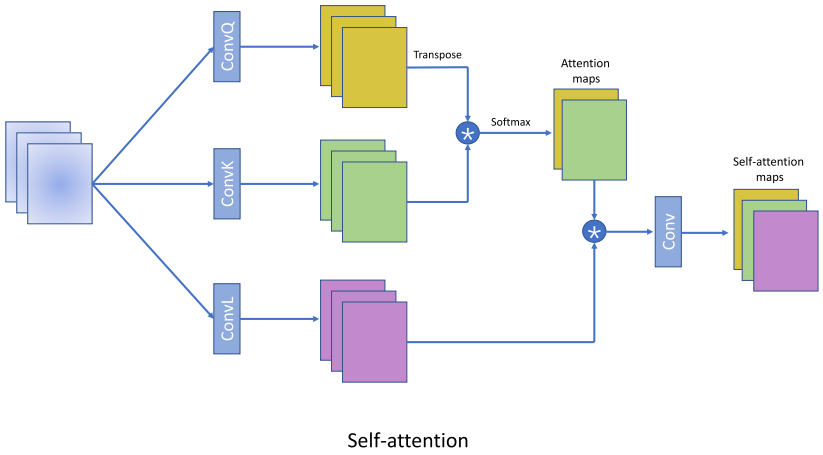

FIGURE 8. Overall framework of self-attention. The ∗ denotes matrix multiplication.

TABLE 1. Experiment setup.

Hardware	Software
CPU: 8 Cores	Windows10
RAM: 32 GB DDR4	CUDA10.0 + CUDNN7.0
GPU: NVIDIA RTX2080Ti (11GB GDDR6)	Pytorch 1.2.0 + Python 3.6

interpolation to downsample the images of the DIV2K [30] dataset and also added additive gaussian noise to the low-resolution images to create clear and unclear image pairs. For pest classification, we used the 24 insect classes of the Xie1 [8] dataset and 37 insect classes of the Xie2 [9] dataset. 1200 pest images of Xie1 dataset were used to train the classification networks and 240 pest images to test. For the Xie2 dataset, we used 3618 pest images to train the classification networks and 667 pest images to test. We randomly rotated and flipped the images and used batch normalization for data augmentation and all images have been pre-processed for better training and testing results. Besides, we also made appropriate adjustments to the Xie1 dataset. Figure 9 and Figure 10 show the examples of the Xie1 dataset and the Xie2 dataset. Categories and pictures' number of the Xie1 dataset and the Xie2 dataset could be found in Table 2 and Table 3.

For pest instance segmentation, we chose Diostrombus, Chauliops, and Callitettix as our research objects, which are all very common in the actual farmland environment and do great harm to agricultural production. They look similar to the living environment, which makes them hard to be found from the background. We built a dataset with 58 images for each class of pest as the instance segmentation samples. All pest images were bicubic downsampling to reduce the size to one-quarter of their original size to simulate low pixel resolution and blurred insect morphology.

C. TRAINING DETAILS

In the preprocessing, we used bicubic interpolation to downsample the images and we added additive gaussian noise to the low-resolution image to create clear and unclear image pairs. we augmented the training data and reorganized the training set by making mirrors and rotate with four directions. All the input images were converted from RGB channels to YCbCr color space, and only the Y channel input is retained. Each time 32 images were selected to build a mini batch, the initial learning rate during training is 0.0001. The momentum was set to 0.9 and weight decay was set to 0.0001. Then we adopted the decay method of the learning rate, which reduced the learning rate by 10 times every 60 epochs and eventually trained 300 epochs using the RMSProp optimizer. What's more, we also pretrained our discriminative model using the trained VGG19 model to supply an initialization when training our PSRGAN to avoid undesired local optima.

D. COMPARE WITH STATE OF THE ARTS

1) VISUAL COMPARISON

Figure 11 and Figure 12 respectively show the visual comparison between our proposed method and other state-of-the-art PSNR-oriented methods including SRdenseNet [31], DSRNLP [7], VDSR [32], SESR [33], LapSRN [34] and PSNR and SSIM are also provided for reference. It could be observed from Figure 11 and Figure 12 that our proposed PSRGAN outperforms the previous state-of-the-art PSNR-oriented methods in details and sharpness. Our proposed PSRGAN has significant advantages in reconstructing the overall profile and body details of insects. However, the previous state-of-the-art PSNR-oriented super-resolution methods tend to generate blurry results and introduce unpleasant artifacts. Besides, the generated textures of previous super-resolution methods are unnatural and contain unpleasant noise. Therefore, PSRGAN can reconstruct the detailed body of pests better than the previous state-of-the-art PSNR-oriented super-resolution methods and improve classification accuracy and object instance segmentation recall rate more.

2) IMAGE CLASSIFICATION

We used several classification networks in our experiment including AlexNet [35], VGG-16 [20], Inception-v3 [36], ResNet-101 [25], Resnext50 [37], DenseNet-121 [26],

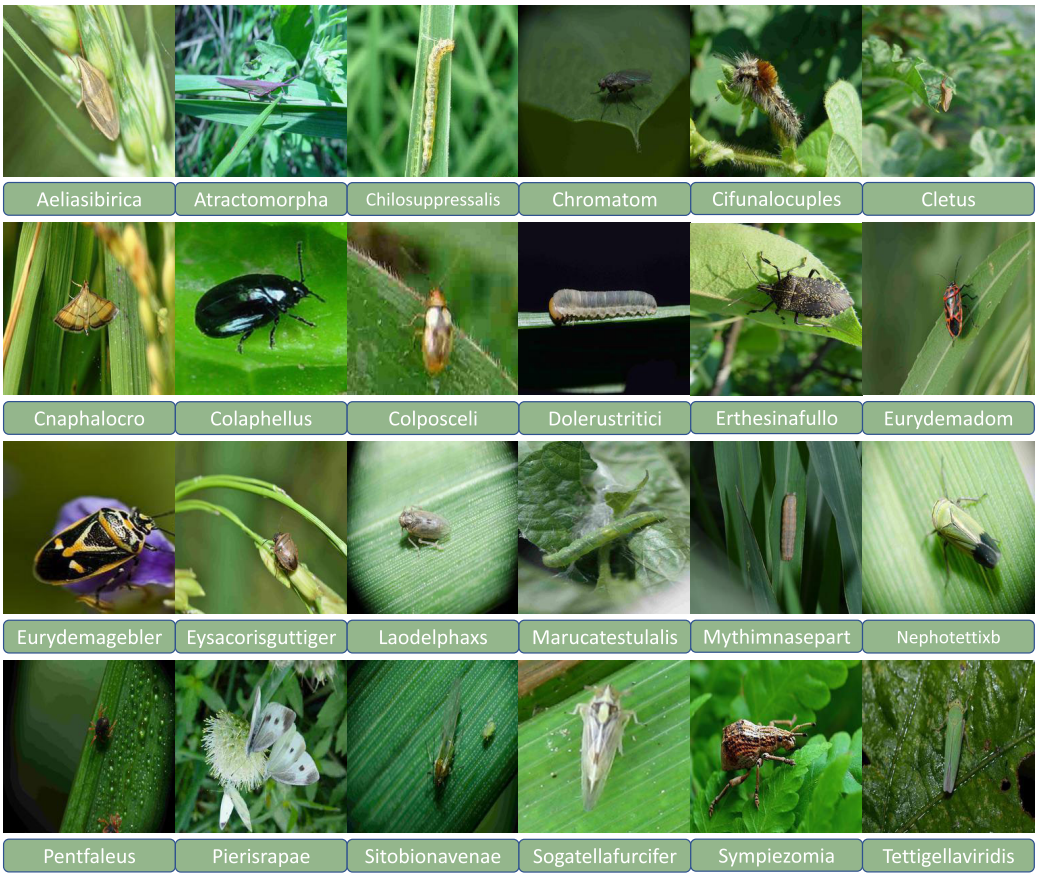

FIGURE 9. Sample images for 24-classes pest dataset.

TABLE 2. 24-classes pest dataset.

Name	Amount	Name	Amount
Aeliasibirica	60	Atractomorphasinensis	60
Chilosuppressalis	60	Chromatomyiahorticola	60
Cifunalocuples	60	Cletus punctiger	60
Cnaphalocrocismedinalis	60	Colaphellusbowvingi	60
Colposcelissignata	60	Dolerustritici	60
Erthesinafullo	60	Eurydemadominulus	60
Eurydemagebleri	60	Eysacorisguttiger	60
Laodelphaxstriatellua	60	Marucatestulalis	60
Mythimnaseparta	60	Nephotettixbipunctatus	60
Pentfaleus major	60	Pierisrapae	60
Sitobionavenae	60	Sogatellafurcifera	60
Sympiezomiasvelatus	60	Tettigellaviridis	60
Total	1440		

MobileNet V2 [38], ShuffleNet V2 [39]. Fine-tuning was used when training the classification networks and we retained most of the weights of the original networks and only trained the soft-max layer. We used 1200 pest images of 24-classes pest dataset to train the classification networks and 240 pest images were used to test. And for the 37-classes pest dataset, we chose 3618 pest images to train the classification networks and 667 pest images to test. What's more, cross-entropy was used as the loss function and we selected Adam as the optimizer. More training details of classification networks could be found in Table 4 and Table 5.

Table 6 and Table 7 show the classification accuracy of raw images and images restored by SRdenseNet [31], DSRNLP [7], VDSR [32], SESR [33], LapSRN [34], RCAN [40], SAN [41] and our proposed PSRGAN respectively. And we can derive from Table 6 and Table 7 that if pest images were super-resolved, it could help enhance the classification accuracy. Moreover, our proposed PSRGAN outperforms than the state-of-the-art PSNR-oriented

FIGURE 10. Sample images for 37-classes pest dataset.

super-resolution methods and improves classification accuracy more.

3) OBJECT INSTANCE SEGMENTATION RECALL RATE

We used the Mask RCNN [10] as the object instance segmentation model and train a segmentation system for the Diostrombus based on the resnet50 model. The training utilized the resnet50 model pretrained on the COCO dataset. Fine-tuning was carried out on the basis of this and we used Adam optimizer with a fixed learning rate of 0.001 to make the network converge. And after 20,000 iterations, the model basically converged. Figure 13, Figure 14 and Figure 15 respectively show the visualized pest instance segmentation results of Diostrombus, Chauliops, and Callitettix.

TABLE 3. 37-classes pest dataset.

Name	Amount	Name	Amount
Aulacophora indica	78	Bemisia tabaci	147
Callitettix versicolor	156	Ceroplastes ceriferus	100
Ceutorhynchus asper	146	Chauliops fallax Scott	68
Chilo supperssalis	93	Chromatomyia horticola	114
Cicadella viridis	138	Cletus punctiger	170
Corythucha ciliata	90	Corythucha marmorata	98
Dicladispa armigera	150	Diostrombus politus	238
Dolerus tritici Chu	88	Dolycoris baccarum	87
Dryocosmus	50	Empoasca flavescens	133
Eurydema dominulus	150	Graphosoma rubrolineata	116
Iscadia inexacta	79	Leptocorisa acuta	133
Luperomorpha suturalis	101	Lycorma delicatula	92
Maruca testulalis Gryer	73	Nezara viridula	175
Phyllotreta striolata	187	Pieris rapae	71
Plutella xylostella	68	Porthesia taiwana Shiraki	141
Riptortus pedestris	110	Scotinophara lurida	117
Sesamia inferens	126	Spilosoma obliqua	66
Spodoptera litura	130	Stollia ventralis	72
Strongyllodes variegatus	135	total	4285

TABLE 4. Training details of classification networks using 24-classes pest dataset.

Method	Learning rate	Batch size	Epochs
Alexnet	0.0001	25	50
VGG16	0.0001	25	50
Inception-v3	0.0001	25	65
Resnet101	0.0001	25	60
Resnext50	0.0005	25	60
Densenet121	0.0005	25	50
MobileNet V2	0.0001	25	60
ShuffleNet V2	0.0001	25	70

TABLE 5. Training details of classification networks using 37-classes pest dataset.

Method	Learning rate	Batch size	Epochs
Alexnet	0.0001	50	50
VGG16	0.0001	60	50
Inception-v3	0.0001	60	50
Resnet101	0.0001	60	60
Resnext50	0.0001	60	80
Densenet121	0.0005	60	60
MobileNet V2	0.0001	60	60
ShuffleNet V2	0.0001	60	60

These figures indicate that low-resolution pest images and images upscaled by bicubic were difficult to be segmented in the system while the image restored by super-resolution methods was successfully segmented. Besides, the recognition confidence and the accuracy of the bounding box of the target pest of our proposed method is higher than other previous state-of-the-art PSNR-oriented super-resolution methods.

The experiment used 58 low-resolution images to test each class of pest. Table 8 shows the recall rate results for untreated pest images and after restored by Bicubic interpolation, SRdenseNet [31], DSRNLP [7], VDSR [32], SESR [33], LapSRN [34], RCAN [40], SAN [41] and our proposed PSRGAN respectively. Furthermore, we also calculated the improvement of recall rate between our proposed method and the baseline on the three test datasets, which could be found in Table 8.

The experiment results show that after reconstructing with our proposed PSRGAN, the recall rate increased by 182.89% compared to the original low-resolution and unclear pest images. And in comparison to other super-resolution methods, the recall rate also increased a lot. The experiment manifests that super-resolution enhanced small pixel scale object instance segmentation result was significantly improved by our proposed method.

4) SPEED

In this part, we researched on the running time of models. We reproduced SRdenseNet [31], DSRNLP [7], VDSR [32], SESR [33], LapSRN [34] with PyTorch on an NVIDIA RTX2080Ti GPU. We selected 20 pest images and 16 pest images from the Xie1 dataset and Xie2 dataset respectively. As shown in Figure 16 and Figure 17, PSRGAN could run at a high speed and achieves the best classification accuracy in MobileNet V2 [38] among the state-of-the-art methods.

V. DISCUSSION

PSRGAN is a GAN with residual and dense fusion and quadra-attention mechanisms. To take good advantage of both residual and dense fusion mechanism, we carry out both residual and dense connections in a single layer. In comparison with the residual network, the generator of PSRGAN could retain more information from previous layers, which enables our network to reserve contiguous memory. Compared to the dense network, PSRGAN can decrease the channel growth by half, which crucially reduces the number of network parameters, making the deeper structure trainable. We also use four attention mechanisms, namely spatial attention, channel attention, texture attention, and Self-attention. Spatial attention could help enhance the details in the reconstructed process and improve the effect of residual and dense fusion. And channel attention helps to learn autonomously

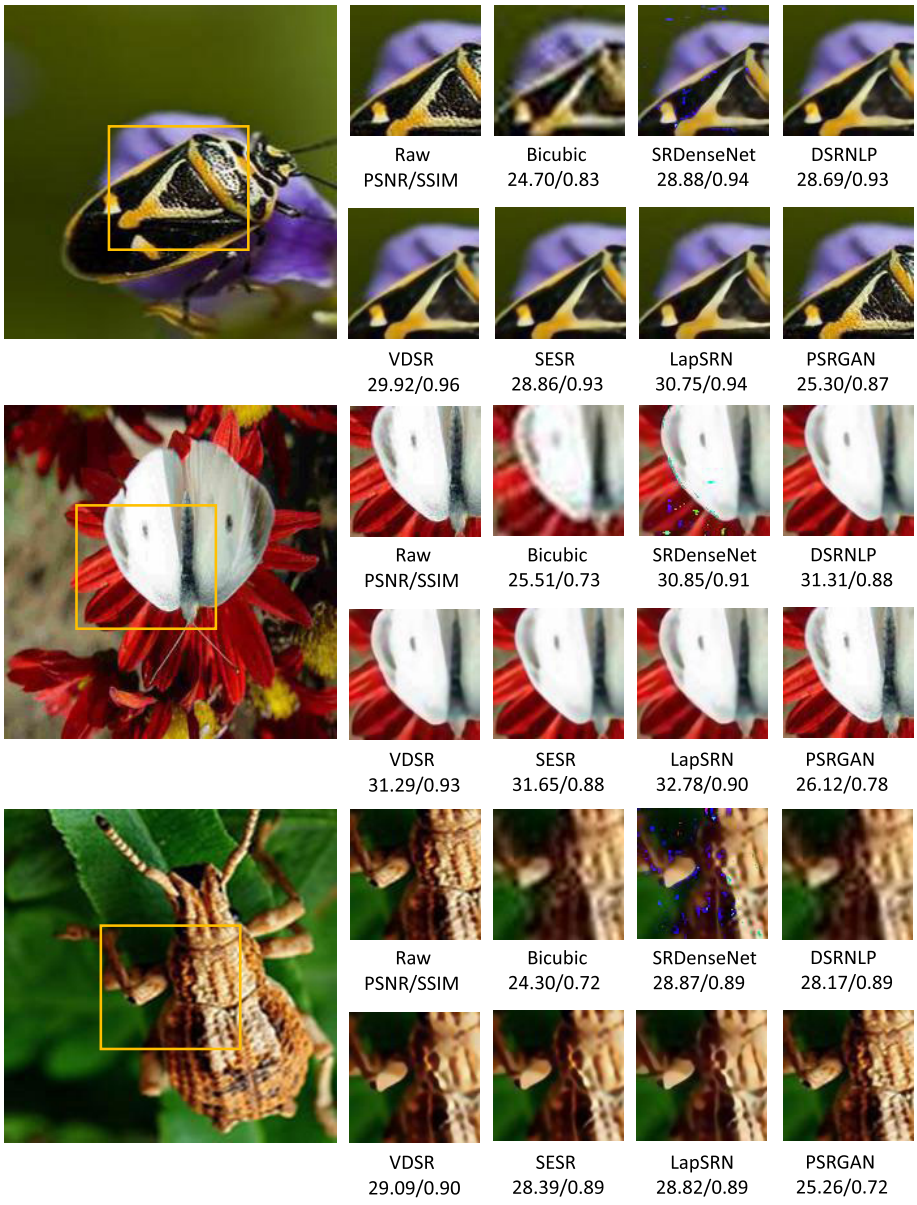

FIGURE 11. Reconstruction quality comparison of 24-classes pest dataset.

TABLE 6. Classification accuracy of 24-classes pest dataset.

Method	Accuracy(%)								
	Raw	DSRNLP	SESR	LapSRN	VDSR	SRDenseNet	RCAN	SAN	PSRGAN(Ours)
Alexnet	85.83	86.25	86.69	86.66	86.50	86.45	86.96	87.41	89.43
VGG16	85.41	86.32	86.65	86.75	86.48	86.50	86.81	87.23	89.11
Inception-v3	89.58	90.56	90.88	90.90	90.85	90.75	90.88	91.53	93.64
Resnet101	90.83	91.36	91.75	91.69	91.51	91.43	91.91	92.63	94.15
Resnext50	88.75	89.61	90.45	90.83	89.85	90.41	90.78	91.88	93.31
Densenet121	89.16	90.89	91.52	91.45	90.95	91.33	91.86	92.56	93.73
MobileNet V2	91.25	91.89	92.15	92.42	91.91	92.08	92.73	93.21	94.15
ShuffleNet V2	89.58	90.53	90.93	90.68	90.66	90.58	91.12	91.68	93.32

to boost significant channels and suppress useless channels. Texture attention assigns attention based on guidance from edges, which help utilize edges as a global spatial attention mechanism for image reconstruction. Self-attention can obtain the global geometric features of images in one step by straightly calculating the relationship between any

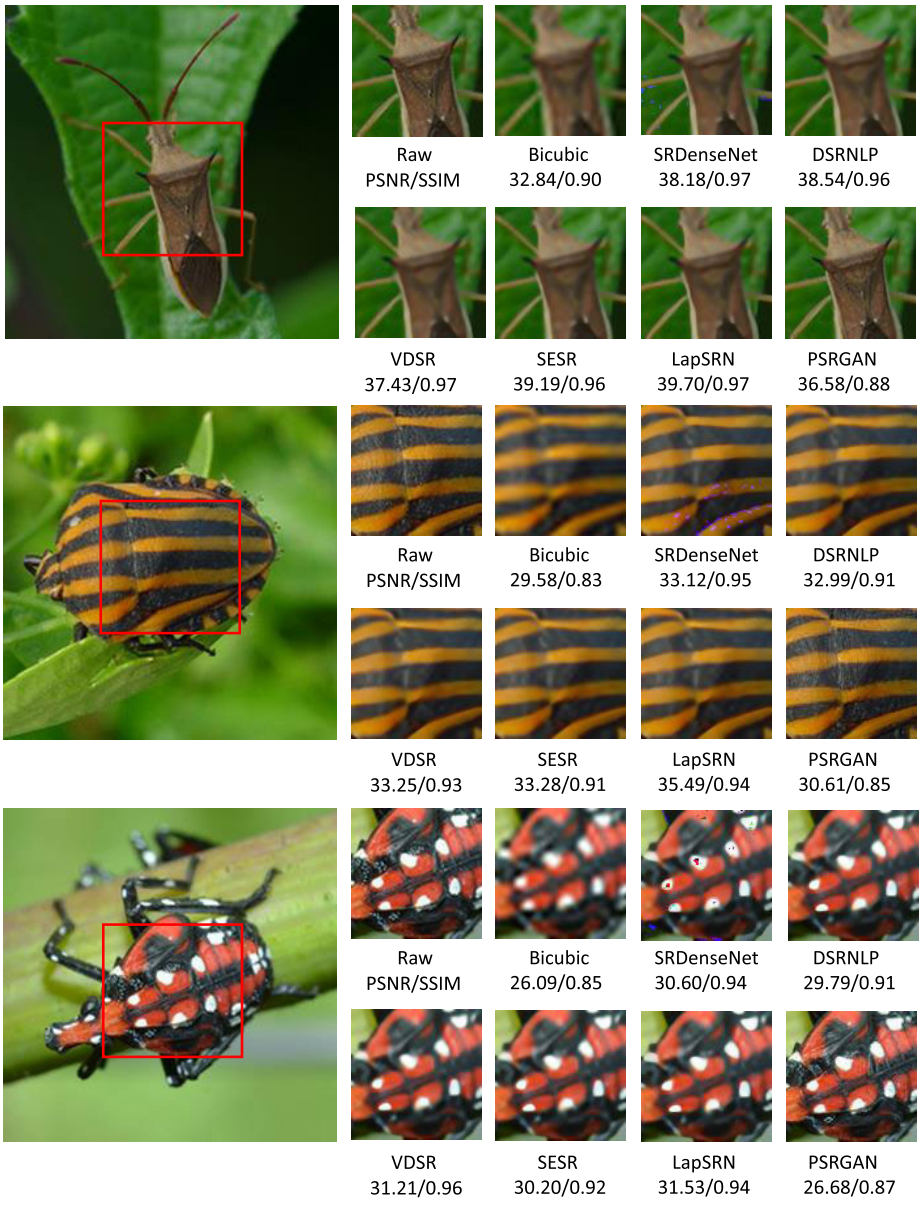

FIGURE 12. Reconstruction quality comparison of 37-classes pest dataset.

TABLE 7. Classification accuracy of 37-classes pest dataset.

Method	Accuracy(%)								
	Raw	DSRNLP	SESR	LapSRN	VDSR	SRDenseNet	RCAN	SAN	PSRGAN(Ours)
Alexnet	92.65	93.25	93.64	93.80	93.42	93.20	93.91	94.32	96.54
VGG16	94.60	95.46	95.52	95.99	95.56	95.15	96.18	96.84	97.96
Inception-v3	94.75	95.89	96.35	96.40	96.40	96.55	96.48	96.84	98.33
Resnet101	94.15	95.22	95.53	95.85	95.85	95.40	95.93	96.38	97.96
Resnext50	95.35	96.35	96.75	96.88	96.65	96.45	96.95	97.31	98.88
Densenet121	95.20	96.32	96.65	96.56	96.35	96.75	96.71	96.92	98.93
MobileNet V2	96.85	97.22	97.53	97.84	97.35	97.60	97.90	98.26	99.10
ShuffleNet V2	96.65	96.86	97.32	97.13	96.80	97.00	96.78	97.17	98.85

two pixels in the image, thereby the self-attention mechanism could study the dependencies between global features well. Additionally, CARAFE upsampling is used in the generator network of PSRGAN, which could have a larger receptive field during recombination and guide the recombination process according to the input characteristics. And CARAFE

TABLE 8. Performance comparison of recall rate.

Method	Callitettix	Chauliops	Diostrombus
Baseline	33.41%	25.48%	43.32%
Bicubic	45.62%	39.76%	53.62%
SRDenseNet	62.16%	68.86%	73.86%
VDSR	74.13%	70.32%	84.67%
DSRNLP	65.60%	60.24%	85.82%
SESR	60.31%	63.52%	83.68%
LapSRN	78.57%	73.68%	88.54%
RCAN	82.43%	76.82%	90.32%
SAN	85.36%	80.41%	92.56%
PSRGAN(ours)	88.68%	92.36%	95.64%
Improvement	165.42%	262.48%	120.77%
Average			182.89%

FIGURE 13. Sample images for instance segmentation results with Diostrombus.

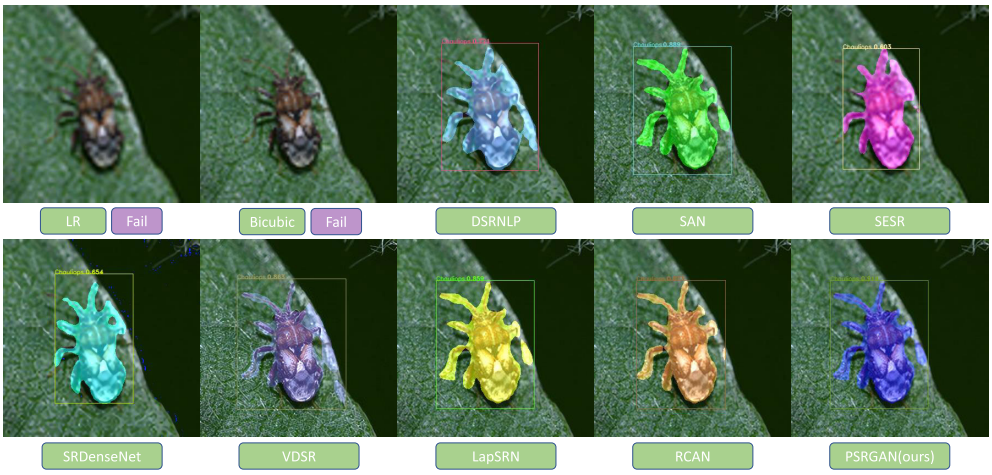

FIGURE 14. Sample images for instance segmentation results with Chauliops.

upsampling also assists our PSRGAN to gain good results in segmentation tasks.

According to experiment results, the classification accuracy and object instance segmentation recall rate for pest images transformed by super-resolution methods are more satisfying than low-resolution pest images obtained from real farms. This is because images upscaled by the super-resolution method could convey much more information, such as high-frequency details of images, in comparison with low-resolution images. Experiment results on two pest datasets clearly reveal this phenomenon that using the low-resolution images results in lower accuracy. These results

FIGURE 15. Sample images for instance segmentation results with Callitettix.

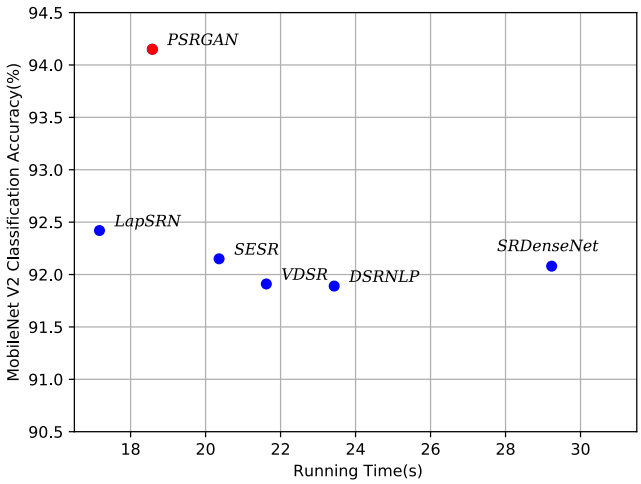

FIGURE 16. Running time comparison with other models on 24-classes pest dataset.

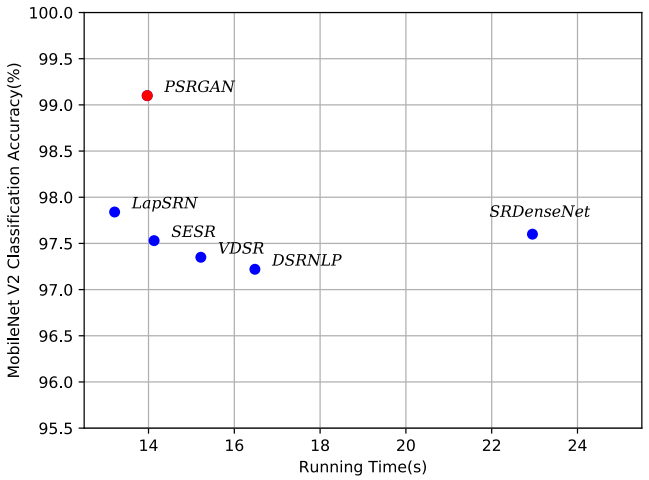

FIGURE 17. Running time comparison with other models on 37-classes pest dataset.

indicate that the super-resolution methods successfully reconstructed the high-frequency details of pest images and promoted the identification of pests. We also compared PSRGAN to the previous state-of-the-art PSNR-oriented super-resolution methods and found that our proposed method performs better in improving classification accuracy and object instance segmentation recall rate.

Although PSRGAN has many advantages over previous super-resolution methods, there are still some limitations that must be overcome. Previous state-of-the-art PSNR-oriented super-resolution methods typically utilize average pixel positions to make images overly smooth but increasing PSNR results. However, the previous state-of-the-art PSNR-oriented super-resolution methods tend to generate blurry results and introduce unpleasant artifacts. Besides, the generated textures of previous super-resolution methods are unnatural and contain unpleasant noise, which has a bad effect on improving classification accuracy and object instance segmentation recall rate. PSRGAN does not use averaging, which facilitates resulting in better visual effects but reducing PSNR results in comparison with other PSNR-oriented super-resolution methods. And PSRGAN has significant advantages in reconstructing the overall profile and body details of insects. Therefore, PSRGAN can reconstruct the detailed body of pests better than the previous state-of-the-art PSNRoriented super-resolution methods and improve classification accuracy and object instance segmentation recall rate more.

VI. CONCLUSION

In this work, we propose a novel image super-resolution method for agricultural pest images. To the best of our knowledge, our proposed method is the first to introduce GANs into agricultural pest image processing. We make use of both residual and dense connections to retain more information from previous layers to promote reserving contiguous memory, and reduce the number of network parameters significantly to make deeper structures trainable. Moreover, a quadra-attention mechanism provides a crucial performance increase. Spatial attention could boost enhancing the details in the reconstructed process. Channel attention

can increase important channels and suppress useless channels. Texture attention facilitates assigning attention based on the texture features and taking advantage of textures as a global spatial attention mechanism during image reconstruction. Self-attention could study the dependencies between global features well. Besides, on the basis of our experimental results, PSRGAN outperforms state-of-the-art super-resolution methods in terms of visual quality, classification, and instance segmentation performance. Based on effective attention mechanisms and residual and dense fusion, PSRGAN could not only help improve classification accuracy and object instance segmentation recall rate but also decrease network parameters. And our proposed method also makes it possible to deploy fewer cameras in the farmland and save costs, which is pretty practical for real-world applications.

REFERENCES

[1] S. Lim, S. Kim, S. Park, and D. Kim, "Development of application for forest insect classification using CNN," in *Proc. 15th Int. Conf. Control, Autom., Robot. Vis. (ICARCV)*, Nov. 2018, pp. 1128–1131.

[2] K. Thenmozhi and U. Srinivasulu Reddy, "Crop pest classification based on deep convolutional neural network and transfer learning," *Comput. Electron. Agricult.*, vol. 164, Sep. 2019, Art. no. 104906.

[3] Y. Li, H. Wang, L. M. Dang, A. Sadeghi-Niaraki, and H. Moon, "Crop pest recognition in natural scenes using convolutional neural networks," *Comput. Electron. Agricult.*, vol. 169, Feb. 2020, Art. no. 105174.

[4] E. C. Tetila, B. B. Machado, G. K. Menezes, A. Da Silva Oliveira, M. Alvarez, W. P. Amorim, N. A. De Souza Belete, G. G. Da Silva, and H. Pistori, "Automatic recognition of soybean leaf diseases using UAV images and deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 5, pp. 903–907, May 2020.

[5] H.-J. Yu and C.-H. Son, "Apple leaf disease identification through Region-of-Interest-Aware deep convolutional neural network," 2019, *arXiv:1903.10356*. [Online]. Available: http://arxiv.org/abs/1903.10356

[6] 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.

[7] 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.

[8] C. Xie, J. Zhang, R. Li, J. Li, P. Hong, J. Xia, and P. Chen, "Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning," *Comput. Electron. Agricult.*, vol. 119, pp. 123–132, Nov. 2015.

[9] C. Xie, R. Wang, J. Zhang, P. Chen, W. Dong, R. Li, T. Chen, and H. Chen, "Multi-level learning features for automatic classification of field crop pests," *Comput. Electron. Agricult.*, vol. 152, pp. 233–241, Sep. 2018.

[10] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2961–2969.

[11] C. Wen, D. E. Guyer, and W. Li, "Local feature-based identification and classification for orchard insects," *Biosyst. Eng.*, vol. 104, no. 3, pp. 299–307, Nov. 2009.

[12] F. Faithpraise, P. Birch, R. Young, J. Obu, B. Faithpraise, and C. Chatwin, "Automatic plant pest detection and recognition using k-means clustering algorithm and correspondence filters," *Int. J. Adv. Biotechnol. Res.*, vol. 4, no. 2, pp. 189–199, 2013.

[13] J. Wang, C. Lin, L. Ji, and A. Liang, "A new automatic identification system of insect images at the order level," *Knowl.-Based Syst.*, vol. 33, pp. 102–110, Sep. 2012.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

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

[19] J. Shijie, J. Peiyi, H. Siping, and S. Haibo, "Automatic detection of tomato diseases and pests based on leaf images," in *Proc. Chin. Autom. Congr. (CAC)*, Oct. 2017, pp. 2510–2537.

[20] 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

[21] Z. Liu, J. Gao, G. Yang, H. Zhang, and Y. He, "Localization and classification of paddy field pests using a saliency map and deep convolutional neural network," *Sci. Rep.*, vol. 6, no. 1, p. 20410, Apr. 2016.

[22] W. Dawei, D. Limiao, N. Jiangong, G. Jiyue, Z. Hongfei, and H. Zhongzhi, "Recognition pest by image-based transfer learning," *J. Sci. Food Agricult.*, vol. 99, no. 10, pp. 4524–4531, 2019.

[23] R. Wang, J. Zhang, W. Dong, J. Yu, C. Xie, R. Li, T. Chen, and H. Chen, "A crop pests image classification algorithm based on deep convolutional neural network," *TELKOMNIKA (Telecommun. Comput. Electron. Control)*, vol. 15, no. 3, p. 1239, 2017.

[24] J. Wang, K. Chen, R. Xu, Z. Liu, C. C. Loy, and D. Lin, "CARAFE: Content-aware ReAssembly of FEatures," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3007–3016.

[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. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[27] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T.-S. Chua, "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5659–5667.

[28] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.

[29] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," 2018, *arXiv:1805.08318*. [Online]. Available: http://arxiv.org/abs/1805.08318

[30] A. Ignatov and R. Timofte, "PIRM challenge on perceptual image enhancement on smartphones: Report," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, Jan. 2019, pp. 315–333.

[31] 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.

[32] 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.

[33] X. Cheng, X. Li, J. Yang, and Y. Tai, "SESR: Single image super resolution with recursive squeeze and excitation networks," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 147–152.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] 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.
[40] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.
[41] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, p. 11.

QIANG DAI was born in Anhui, China. 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 currently a Research Assistant with the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information, Ministry of Education, and the Jiangsu Key Laboratory of Image and Video Understanding for Social Security, Nanjing. His research interest includes 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 Post-doctoral 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 Agriculture University, China. She currently focuses on network monitoring, 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.