Super Resolution Reconstruction of Bidirectional Feature Flow Images Based on SRGAN

Kangliang Xiao[1], Shaozhang Xiao[1*], Yiyan Zhao[1]

[1] Huaiyin Institute of Technology, huai'an, 223001, China

* xiaoshaozhang@hyit.edu.cn

https://doi.org/10.70695/IAAI202504A15

Abstract

Over the years, in order to prevent the impact of leaf diseases on crop yields, deep learning technology has been introduced into the field of leaf disease recognition and has achieved good results. Deep learning techniques rely on high-quality images, and encountering low-quality images can cause a significant decline in model performance. To avoid performance degradation, low-quality images are often reconstructed using image super-resolution before being input into recognition models. However, existing super-resolution models often suffer from detail blurring and structural distortion due to their unidirectional feature transfer when dealing with complex textures and structures, which limits further performance improvement. To this end, this article innovatively introduces a bidirectional feature flow module, which achieves collaborative enhancement and more accurate reconstruction of multi-level image features through parallel processing and dynamic fusion of local details and global context. Compared to the original model, the PSNR and SSIM have improved by 0.47dB and 1.50% respectively.

Keywords Super-Resolution Reconstruction; Leaf Diseases; Bidirectional Features

1 Introduction

Every year, food losses caused by pests and diseases account for nearly one-third of the total food production, posing a serious threat to food security [1]. Corn, as one of the three major food crops in the world, has the advantages of high yield, strong adaptability, and wide application. However, diseases can cause a loss of more than 40% in corn yield, resulting in huge economic losses. At present, the identification of corn diseases is mainly based on manual identification, which has problems such as low efficiency and low success rate. With the development of technology, the combination of machine vision and image processing has been applied in the treatment of corn diseases, solving the problems of high manual recognition error rate, high demand for professional knowledge, and consumption of manpower and material resources. However, the manual feature extraction process is cumbersome, time-consuming, and the final recognition accuracy is limited, which severely limits its practical application. Since Hinton proposed deep learning, people have conducted detailed research on it and applied it to different fields. In the field of crop leaf disease control, it has been found that using deep learning can more accurately and quickly identify or classify diseases, and then carry out specialized treatment according to different methods, greatly reducing the losses caused by diseases, improving yields, and reducing the consumption of human and material resources [2]. Yang Bo et al. proposed an improved SPP-x YOLOv5 model for rice disease recognition to address the high computational complexity and slow speed of YOLOv5 [3]. The improved model achieved a detection speed of only 0.34 seconds per page; Agarwal et al. proposed a plant disease classification model using only 3 convolutional layers, 3 pooling layers, and 2 fully connected layers [4]. Compared with traditional VGG and MobileNet, the accuracy was improved by 14% and 29%, respectively; Wu Huarui used Bayesian optimization algorithm to reduce the difficulty of network training and introduced residual units to alleviate the problem of gradient explosion and disappearance [5-7]. Finally, the optimized model was successfully applied to tomato leaf disease recognition; Long Mansheng et al. used deep transfer learning techniques to transfer the general feature knowledge learned by the model on the ImageNet large-scale dataset to specific oil tea disease recognition tasks, effectively improving the performance and generalization ability of the model [8].

With the increasing use of classification or recognition models in the field of crop leaf diseases, it has been found that when encountering blurry and insufficiently detailed images, the accuracy of both models will significantly decrease and fluctuate. To address this issue, people attempt to perform super-resolution operations on low resolution images (LR) and convert them into high-resolution images (HR) before inputting them into classification or recognition models.

Image super-resolution reconstruction (ISRR) is a fundamental computer vision task. Unlike traditional methods that rely on manually selecting features, deep learning based super-resolution algorithms reconstruct images by fitting the mapping relationship from low resolution images to high-resolution images. Dong et al. first attempted to use convolutional networks in super-resolution algorithms, designing a super solution convolutional neural network (SRCNN) model that does not require manual design of complex image priors and restoration algorithms, but directly allows a deep learning model to learn the end-to-end mapping relationship between a large number of "low resolution high resolution" image pairs [9]. It mainly consists of three modules: fast image extraction and expression, nonlinear mapping, and reconstruction. Shi et al. proposed an Efficient Sub Pixel Convolutional Neural Network (ESPCN), which uses sub-pixel convolutional layers instead of convolutional layers to achieve image upsampling and improve image resolution [10].

Super resolution technology can effectively restore image details, providing higher quality data input for subsequent classification or recognition tasks. Therefore, reconstructing low resolution corn disease images into high-resolution images through super-resolution models and inputting them into recognition models is a technical path with clear value. However, existing methods still commonly suffer from key issues such as blurred edges of leaf lesions, overall smoothness of the image, and structural distortion when processing leaf disease images, which seriously restricts the further improvement of recognition model performance.

To address the aforementioned issues, this paper proposes a super-resolution reconstruction method based on Bidirectional Feature Flow Generative Adversarial Network (BFFSRGAN). This module constructs a bidirectional information pathway between local and global regions, and utilizes gating mechanisms to achieve adaptive feature filtering and fusion, thereby synergistically enhancing the model's ability to reconstruct details and maintain structure. Ultimately, it generates super-resolution results with clearer texture, more realistic vision, and reasonable structure.

2 Related Work

2.1 SRGAN Model

Ledig et al. proposed using Generative Adversarial Networks (GANs) for super-resolution SRGAN(super-resolution generative adversarial network) [11-12]. The key contribution of SRGAN is to shift the focus of super-resolution from pursuing high Peak Signal to Noise Ratio (PSNR) to pursuing high-resolution images that are visually more realistic and natural to the human eye. Before SRGAN was proposed, most super-resolution models used mean square error (MSE) as the loss function. Although this approach can achieve very high PSNR values, MSE also has a serious drawback: it tends to learn the average of all high-resolution images, resulting in images that are too smooth and lack high-frequency details. SRGAN has changed the approach of using MSE as the loss function, proposing that we should not only pursue precise correspondence between pixels, but also pursue the core idea of making the generated image indistinguishable from real high-definition images in the eyes of human observers. It consists of a generator (G) and a discriminator (D). The two are a process of mutual competition and learning, until reaching Nash equilibrium.

The generator task is to receive feedback from the discriminator for learning and generate fake images to attempt to deceive the discriminator again; The discriminator task is to determine whether the image is the original image or a fake image generated by the generator.

The network core of the generator consists of multiple stacked residual blocks, which enable direct cross layer transmission of information. The core idea of this design is to learn the complex mapping relationship from low resolution images to high-resolution images through stacked residual blocks. The goal is no longer simply to restore pixel accuracy, but to generate a high-resolution image with realism and rich details. After the image input generator, it first goes through a shallow feature extraction block, and then enters multiple concatenated residual blocks. The residual blocks are connected by skip connections to add the input and convolution output and send them to the next residual block. This structure can greatly solve the problem of vanishing gradients in deep networks, allowing the network to be trained very deeply, thus possessing strong feature expression ability and learning complex mapping relationships. SRGAN adopts an upsampling module that includes sub-pixel convolutional layers. It expands the feature map in the channel dimension through convolution, and then achieves high-resolution spatial reconstruction through a periodic shuffling operation. Compared with traditional deconvolution or interpolation operations, this sub-pixel convolution can more efficiently integrate information and generate clearer details. Figure 1 shows the structure of the generator.

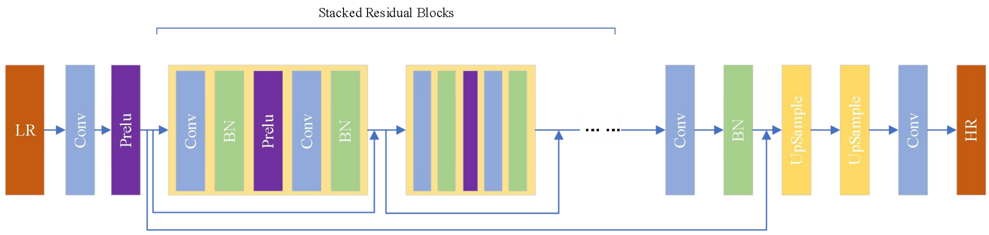

Fig. 1. SRGAN generator structure

The discriminator of SRGAN is a typical discriminative network, whose core task is a binary classification problem: to determine whether the input image is a real high-resolution image or a high-resolution image forged by the generator. The architecture of the discriminator is a very classic image classification network, but the target is more specific. It receives images as input and then goes through a series of convolutional layers with Leaky ReLU activation functions. These convolutional layers share a common feature: they all use convolutions with a stride of 2 to gradually downsample feature maps, increasing the number of image channels while decreasing the height and width, gradually extracting abstract features from pixel level information, from texture, edges to overall layout. After each convolutional layer, use the Leaky ReLU activation function to help the model alleviate the gradient vanishing problem. Finally, a fully connected layer is used in conjunction with the Sigmoid function to aggregate all information into a single scalar value between 0 and 1, representing the likelihood that the input image is a real image. The task of the discriminator is to determine whether an input image is a high-resolution image or a fake image generated by the generator. (SR) The following is the structural model of the discriminator.

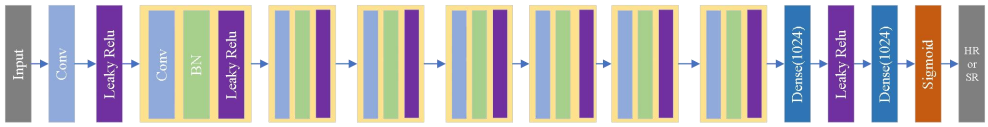

Fig. 2. SRGAN discriminator structure

2.2 BFFSRGAN Model

The original intention of adding BN layer in SRGAN is to reduce internal variable offset, accelerate the training speed of deeper networks, further improve network stability, and have a certain regularization effect. It is precisely because the BN layer has such an effect that it has achieved great success in tasks such as image classification. However, in super-resolution models, the BN layer can bring many problems: the BN layer relies on the normalization of the current batch, which means that the output value depends not only on the input image, but also on other collected and selected images in the same batch. This can lead to unnatural, checkerboard like artifacts or speckle like noise in super-resolution reconstruction, which is called "artifacts". The phenomenon of artifacts can seriously damage the visual quality of images.

In order to improve the performance of SRGAN models in super-resolution reconstruction and address the limitations of traditional residual block feature extraction, we propose a Bidirectional Feature Flow (BFF) module to replace the residual blocks in the original model. There are two main improvements to this improvement: firstly, to avoid artifacts in the generated images, we removed the BN layer; Secondly, we introduced a bidirectional feature flow mechanism, which achieves collaborative optimization from details to structure by parallel processing of local and global information, thereby enhancing the model's multi-scale perception ability of image content.

The core idea of the bidirectional feature flow mechanism module is to simultaneously capture high-resolution local details and low resolution global context, and dynamically fuse them through an adaptive gating mechanism. The local feature flow processes the input image through a series of convolutions, focusing on image texture and edge information to maintain high spatial resolution; The global feature flow downsamples the feature map to a fixed size through adaptive average pooling and then processes it to capture the overall structure and semantic content of the image.

In the feature fusion stage, an interactive gating network is introduced, which receives the concatenated results of local and global features and generates a weight map. This gating mechanism allows the model to adaptively balance local details and global structure: after multiplying local features with gating weights, the detail regions are further strengthened; Multiplying global features by inverse weights can further enhance the guiding role of structural consistency. Finally, merge the convolutions and perform residual linking with the input image to ensure the stability of gradient flow and promote feature reuse. BFF is shown in Figure 3, and the overall structure of BFFSRGAN is shown in Figure 4.

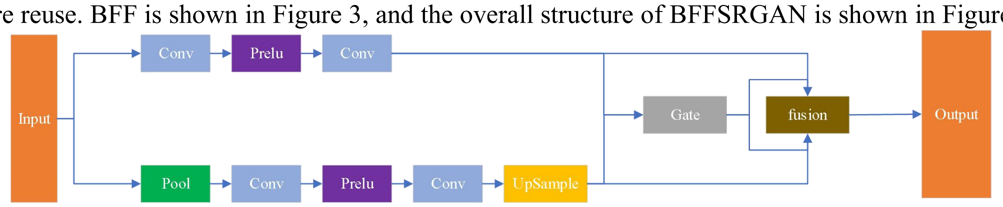

Fig. 3. BFF model

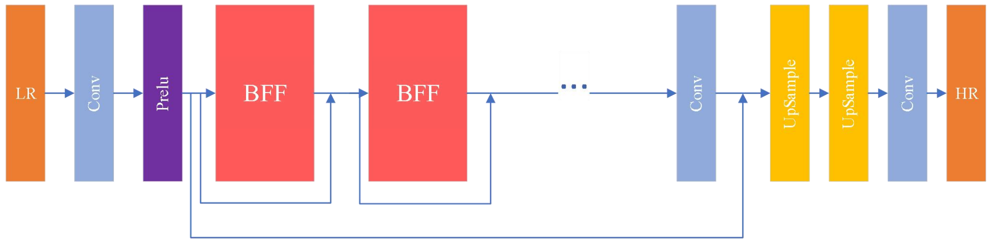

Fig. 4. BFFSRGAN model

3 Experiments

3.1 Experimental Environment

The hardware environment for this experiment is CPU: Intel(R) Core(TM) i9-14900KF@ 3.20 GHz，GPU: NVIDIA GeForce RTX 5070 Ti, 16G Visual memory. The software environment is Python 3.10.16 programming language; Windows 11 system; Pytorch 2.8.0 deep learning framework; Cuda 12.8 accelerates learning.

In response to the research question, the publicly available leaf disease dataset Crop Disease was used for corn disease images, and inverted images were used for data augmentation, with a total of 4556 images. All images in this dataset have a resolution of 1024 * 1024 or higher, and low resolution images are created by double triple downsampling the original image with a magnification factor of 4.

3.2 Evaluation Metric

This article uses PSNR and Structural Similarity Index Measure (SSIM).

PSNR is an image quality evaluation metric based on pixel level error, which quantifies reconstruction quality by calculating the mean square error between the original image and the processed image. It is currently one of the most common and widely used evaluations.

The proposal of SSIM is to overcome the shortcomings of PSNR and attempt to simulate the human eye's perception of the overall structural information of an image. Evaluating image quality comprehensively from three dimensions: brightness, contrast, and structure is more in line with the perceptual characteristics of the human visual system.

3.3 Quantitative Analysis

To comprehensively evaluate the performance of the proposed BFFSRGAN model in the super-resolution reconstruction task of maize leaf disease images, this paper selected the current mainstream super-resolution algorithms SRGAN, CBSRGAN, and RCPGAN as comparison benchmarks.

As shown in Table 1, in terms of PSNR index, BFFSRGAN reached 23.636 dB, significantly better than SRGAN's 23.166 dB, CBSRGAN's 22.943 dB, and RCPGAN's 23.524 dB. This indicates that BFFSRGAN has stronger pixel level detail recovery ability and can effectively suppress irrelevant noise. Compared with the original SRGAN, it has improved by 0.47dB, indicating that the introduced bidirectional feature flow mechanism effectively enhances the model's ability to learn local texture information and avoids the problem of detail blurring caused by traditional one-way paths.

In terms of SSIM indicators, BFFSRGAN also showed the best performance, reaching 0.680, higher than SRGAN's 0.665, CBSRGAN's 0.650, and RCPGAN's 0.674. This means that BFFSRGAN maintains higher consistency at the visual structure level, verifying the effectiveness of the bidirectional feature fusion strategy in preserving the semantic structure of images.

In summary, BFFSRGAN has achieved leading advantages in both PSNR and SSIM key indicators, fully demonstrating the effectiveness of the proposed bidirectional feature flow module in synergistically enhancing local details and global structure. This module is capable of capturing fine textures while maintaining overall structural integrity, thus achieving higher quality visual output in the super-resolution reconstruction task of maize leaf disease images.

Table 1. Four indicators reconstructed by each algorithm on the dataset

Metric	SRGAN	CBSRGAN	RCPGAN	BBFSRGAN
PSNR	23.166	22.943	23.524	23.636
SSIM	0.665	0.650	0.674	0.680

3.4 Ablation Experiment

To comprehensively evaluate the effectiveness of each component in the proposed Bidirectional Feature Flow module, a series of ablation experiments were designed in this paper. The standard residual blocks in the original SRGAN architecture have been gradually replaced with different variants, including retaining only local flows, retaining only global flows, and removing interactive gating mechanism configurations. By conducting fair comparisons under the same training settings and dataset, we compared and analyzed the PSNR and SSIM metrics, and systematically analyzed the specific contributions of designs such as local global feature collaboration and adaptive gating fusion to the performance of super-resolution reconstruction.

As shown in Table 2, when using local flow alone, the PSNR is 23.188 and SSIM is 0.672, which is an improvement compared to the original model. The reason is that the BN layer has been removed to reduce artifacts, and local details have been better restored; When using global flow alone, the performance of the model decreases. This is because global flow is mainly responsible for compensating local flow for global information in the image. The pooling layer is used to compress features, which leads to the loss of high-frequency details, manifested as edge blurring and texture distortion. Without the supplement of local flow, the model cannot effectively reconstruct local texture details; The module that retains both local and global flows, as the fusion of their information complements each other, preserves both local details and global information, resulting in significant improvements compared to retaining only local or global flows; The complete BFFSRGAN model adds an interactive gating mechanism on the global and local flow modules. The interactive gating mechanism dynamically suppresses the redundancy of the global flow and activates the details of the local flow, achieving common gain and achieving the highest model performance.

Table 2. Results of ablation experiment

	SRGAN	w/o Global	w/o Local	w/o Gate	BBFSRGAN
PSNR	23.166	23.188	22.475	23.410	23.636
SSIM	0.665	0.672	0.638	0.674	0.680

3.5 Qualitative Analysis

To visually evaluate the improved visual performance of the BFFSRGAN model, typical samples were selected for visual comparison. As shown in Figure 5, in the first set of images, after SRGAN reconstruction, the overall structure is basically preserved, but the leaf vein details are relatively blurry and there are artifacts at the edges; CBSRGAN enhanced local texture, but the direction of leaf veins was not well restored; RCPGAN emphasizes structural consistency, which results in smooth images and consistent colors. BBFSRGAN not only reconstructed the branching structure of leaf veins, eliminated artifacts, but also restored the burrs on the leaves.

The second group demonstrated the recovery effect of lesions, which have irregular edges, blurry internal textures, and complex color transitions, posing a serious challenge to the detail recovery ability of super-resolution models; War. The presence of SRGAN results in blurry lesion contours and obvious artifacts. Although CBSRGAN enhances local contrast, there is an issue of excessive sharpening, resulting in unnatural highlights at patch boundaries and significant distortion of internal textures; Due to its emphasis on consistency, the patch distribution of RCPGAN tends to be reasonable, but it is still difficult to restore its internal fine cracks and grayscale gradient features. The BBFSRGAN model exhibits significant advantages in visual quality, as its reconstructed results not only preserve the overall shape and distribution pattern of the lesions, but also achieve more refined restoration of internal details, such as the recovery of some gray lesions.

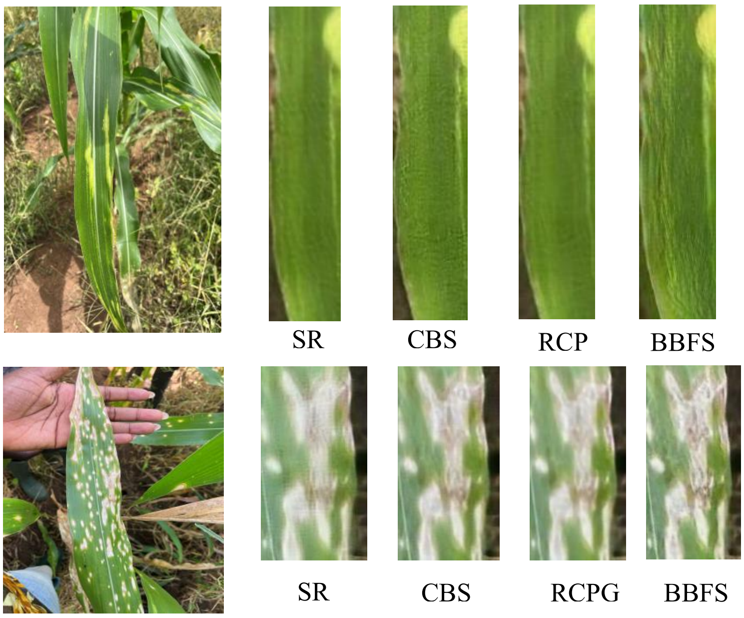

Fig. 5. Image comparison

3.6 Classification comparison experiment

To evaluate the effectiveness of the proposed super-resolution method, this paper selected a representative set of image samples related to corn diseases from the PlantVillage dataset for experimentation. Specifically, the original low-resolution image and the super-resolution image reconstructed by this method are used as inputs and fed into four classic deep convolutional neural network classification models - ResNet, AlexNet, GoogLeNet, and DenseNet - for performance testing. The evaluation metrics cover classification accuracy (Accu), precision (Prec), recall, and F1 score to comprehensively measure the gain effect of super-resolution reconstruction on downstream visual tasks. The relevant experimental results are summarized in Table 3.

Table 3. Classification model results

		Acc	Prec	Recall	F1
ResNet	LR	91.84	89.28	88.42	88.64
	SR	96.11	94.91	93.89	94.33
AlexNet	LR	95.85	94.70	94.36	94.33
	SR	98.19	97.84	97.67	97.53
GooLeNet	LR	87.31	84.26	83.45	83.78
	SR	93.65	92.04	91.19	91.24
DenseNet	LR	93.39	91.28	90.34	90.53
	SR	94.17	92.86	91.32	91.71

On the ResNet model, the classification accuracy of SR images increased from 91.84% to 96.11%, an increase of 4.27 percentage points; Meanwhile, the accuracy, recall, and F1 score were improved by

5.63%, 5.47%, and 5.69%, respectively, indicating that the model has achieved better discriminative ability in identifying subtle texture information such as plant diseases. Similarly, on AlexNet, SR input improved accuracy from 95.85% to 98.19%, and F1 score increased to 97.53%, indicating that this method effectively alleviates the problem of category confusion caused by image blur. For GoogLeNet, although it has strong multi-scale feature extraction capabilities, SR images still bring about an accuracy improvement of about 6.34%, and the recall rate has increased from 83.45% to 91.19%, reflecting the advantage of reconstructed images in preserving details of weak targets. DenseNet, as a model with a dense connection structure, has a high sensitivity to the local structure of the input image. Its accuracy under SR conditions has steadily increased from 93.39% to 94.17%, and F1 score has improved to 91.71%, further confirming the important impact of high-quality input on the robustness of the model.

4 Conclusion

This article proposes an image super-resolution reconstruction method based on bidirectional feature flow. The architecture processes local details and global structure in parallel, and utilizes a self use gating mechanism to achieve intelligent fusion of dynamic trade-offs and deep fusion between multi-scale features. It can effectively alleviate the common problems of detail blur and structural distortion in traditional methods. The next step will focus more on improving the perceptual characteristics of human vision by introducing a new loss function to make the generated images more in line with subjective quality evaluation; A brand new upsampling module is proposed for improvement, attempting to introduce learnable upsampling operators or content aware scaling mechanisms to enhance edge preservation and texture generation capabilities.

Acknowledgement

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Fan, X., Zhang, J., & Yan, X. (2021). Corn disease recognition under complicated background based on improved convolutional neural network. Transactions of the Chinese Society for Agricultural Machinery. https://doi.org/10.6041/j.issn.1000-1298.2021.03.023
2. Hinton, G. E., Osindero, S., & Teh, Y. (2006). A fast learning algorithm for deep belief nets. Neural Computation, 18(7), 1527-1554. https://doi.org/10.1162/neco.2006.18.7.1527
3. Yang, B., Huang, J., & Zhang, L. (2023). Identification and detection of rice leaf diseases by YOLOv5 neural network based on improved SPP-x. Journal of Chinese Agricultural Mechanization, 44(9), 190. https://doi.org/10.13733/j.jcam.issn.2095-5553.2023.09.027
4. Agarwal, M., Singh, A., Arjaria, S., Sinha, A., & Gupta, S. (2020). ToLeD: Tomato leaf disease detection using convolution neural network. Procedia Computer Science, 167, 293-301. https://doi.org/10.1016/j.procs.2020.03.225
5. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv. https://doi.org/10.48550/arXiv.1409.1556
6. Howard, A., Zhmoginov, A., Chen, L., Sandler, M., & Zhu, M. (2018). Inverted residuals and linear bottleneck s: Mobile networks for classification, detection and segmentation. arXiv. https://doi.org/10.48550/arXiv.1801.04381
7. Wu, H. (2019). Method of tomato leaf diseases recognition method based on deep residual network. Smart Agriculture, 1(4), 42. https://doi.org/10.12133/j.smartag.2019.1.4.201908-sa002
8. Long, M., Ouyang, C., Liu, H., & Fu, Q. (2018). Image recognition of Camellia oleifera diseases based on convolutional neural network & transfer learning. Transactions of the CSAE, 34(18), 194-201. https://doi.org/10.11975/j.issn.1002-6819.2018.18.024

9. Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2), 295-307. https://doi.org/10.1109/TPAMI.2015.2439281
10. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., & Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1874-1883). https://doi.org/10.1109/CVPR.2016.207
11. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2017). GAN (Generative Adversarial Nets). Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, 29(5), 177. https://doi.org/10.3156/jsoft.29.5_177_2
12. Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 105–114). https://doi.org/10.1109/cvpr.2017.19

Biographies

1. **Kangliang Xiao** Master, engaged in research on image processing;
2. **Shaozhang Xiao** Master, Senior experimenter, mainly engaged in research work in the fields of computer application technology, data mining, and artificial intelligence;
3. **Yiyan Zhao** Undergraduate student at Huaiyin Institute of Technology, with a main research focus on image vision.

基於SRGAN的雙向特徵流動圖像超分辨率重建

肖康亮[1]，肖紹章[1]，趙藝妍[1]

[1]淮陰工學院，淮安，中國，223001

摘要：歷年來爲了防止葉片病害對農作物產量造成的影響，深度學習技術開始引入到葉片病害識別領域並取得良好的效果。深度學習技術依賴於高質量圖像一旦遇到低質量圖像會造成模型性能劇烈下降。爲了防止模型性能下降，人們考慮將低質量圖像通過圖像超分辨率重建後再輸入識別模型中。然而，現有超分辨率模型在處理複雜紋理與結構時，其單向特徵傳遞方式往往導致細節模糊與結構失真，限制了性能的進一步提升。爲此本文創新性地引入雙向特徵流模塊，通過並行處理並動態融合局部細節與全局上下文，實現對圖像多層次特徵的協同增強與更精準重建。對比原有模型在PSNR、SSIM上分別提升0.47dB，1.50%。

關鍵詞：超分辨率重建；叶片病害；双向特征

1. 肖康亮，碩士，從事圖像處理部分研究；
2. 肖紹章，碩士，高級實驗師，主要從事計算機應用技術、數據挖掘和人工智能等領域的研究工作；
3. 趙藝妍，學士，主要研究方向爲圖像視覺方面。