Super-Resolution Based on Residual Dense Network for Agricultural Image

Guojiang Liu, Jiangshu Wei*, Yu Zhu, Yi Wei

College of Information Engineering, Sichuan Agricultural University, Ya’an, Sichuan, 625014, China

*Corresponding author’s e-mail: weijiangshu66@163.com

Abstract. With the rapid development of convolutional neural network in image processing, its image processing performance is becoming more and more outstanding. With the development of agricultural informatization, our demand for effectively accessing to agricultural information is also increasing. For example, greenhouse image monitoring equipment for crops is only in a low level. In this paper, we propose a method to apply image super resolution to the processing of greenhouse monitoring images, and use residual dense network to enhance the learning of global and local features in agricultural images. Therefore, more detailed information can be obtained from the low resolution images. Experiments on the data sets are made from the greenhouse agriculture images, and experimental results show that our method can effectively enhance the quality of agricultural monitoring images. It can better meet the needs of agricultural monitoring.

1. Introduction

The research of image processing technology was developed as a new subject in the early 1960s. The application of image processing in agriculture starts late, but the prospect is very broad. In the field of agriculture, the development of intelligent agriculture and fine agriculture is an inevitable trend with the development of science and technology. As we move toward intelligent development, we need to obtain agricultural information on a large scale, in real time and efficiently and then give it to the computer to process accordingly. In addition, the development of Internet Protocol Version 6 (Ipv6) is increasingly able to meet the real-time and efficient transmission of information. Therefore, traditional data acquisition methods relying on manual collection and wired measurement can hardly meet the requirements of precision agriculture in terms of real-time, accuracy and convenience. Therefore, the application of image processing technology in agriculture can not only monitor the growth of crops in real time. It can also discriminate pests and diseases of crops and promote the adoption of corresponding treatment measures. In addition, the fruits of crops can be tested and classified, so as to ensure the harvest and high yield of agriculture.

The primary purpose of single image super-resolution (SR) is to restore a satisfactory high resolution image (HR) from a low resolution (LR) image. Because the high resolution image in the field of digital image application can provide people with more detailed information than the low resolution image. Super-resolution reconstruction and related techniques play an important role in image information acquisition. Now SR has been applied to many computer vision tasks, such as security and surveillance imaging [2], medical imaging [3] and image generation [4]. A large number of image SR algorithms have been proposed, which can be roughly divided into three categories.
according to their theoretical basis. It includes interpolation class method [5], method based on model reconstruction [6] and method based on machine learning [7-15].

Nowadays, most fine agricultural places such as agricultural greenhouses use single-chip microcomputer as the main controller of the control system, and various sensors are used to obtain various parameters in the greenhouses [18]. However, the inherent defects of the sensor, defocusing, atmospheric conditions and various types of noise seriously affect the quality of digital images. Image resolution is an important criterion to judge image quality. It is usually determined by image acquisition equipment, but in many practical applications, expensive high-precision sensors and other hardware equipment may be an important limiting factor. In order to solve the problem of low quality of agricultural monitoring image, we chose a residual dense network (RDN) that can make full use of all the layered features and residual difference blocks in the original LR image (figure 1). Each layer has direct access to the original LR input, leading to implicit deep monitoring [19]. The results show that the RDN network can generate SR image of the original LR image well, so it can be used to support the plant growth monitoring, identification of afterwards, and control of pests and diseases. In summary, our main contributions are three-fold:

- The image super resolution is applied to agricultural image processing to obtain better quality agricultural monitoring image.
- The residual dense network, which can make full use of LR image information, is selected to realize the super resolution of agricultural images, so that more details can be recovered.
- Through the experiment and test of agricultural greenhouse image data set, it is proved that it can improve the quality of agricultural monitoring image.

2. Related work

Some developed countries have formed their own set of standards for smart greenhouses, and smart greenhouses have basically achieved full coverage. Now, China is vigorously developing modern intelligent agriculture. Yang has been able to collect environmental data and video in the greenhouse and upload them to the database through WIFI, so as to realize data detection and storage [20]. Pan has developed an android app, which can conduct real-time data and detection with mobile phones [18]. At the same time, under the vigorous development of the country, a series of major projects, a series of products in the agricultural Internet of things show, and achieved obvious results. But in general, the existing problem is that the greenhouse equipment is relatively simple, mostly artificial, in the image monitoring is quite weak, and the scope of monitoring is limited.

On the super resolution of single image, machine learning processing technology has made breakthrough achievements and remarkable effects. On the one hand, it can not only learn the information carried by the image itself, but also learn the information change when the image scale changes, which can make full use of the image information. Deep Learning is a branch of machine learning methods. Recently, Deep learning based methods have also made great achievements in computer vision in terms of neural networks [21-24]. Dong [10] et al. proposed SRCNN, which established an end-to-end mapping between LR image and HR image for the first time. This baseline can be improved by increasing the depth of the network and the weight of the Shared network. VDSR [13] and IRCNN [25] stack more convolution layers by using residual learning to increase the depth of the network. DRCN [16] proposes to use recursive learning of deep network for parameter sharing. Tai et al. introduced recursive module in DRRN [26] and memory module in deep network in MemNet [27]. All the above methods need to first interpolate the original LR image to the required size, and then input it into the network. Such preprocessing not only increases the computational complexity [28], but also makes LR image more blurred and overly smooth, and loses details. Because the interpolated LR image is used to extract features, it is also impossible to establish end-to-end mapping between the original LR image and HR image.

Based on the above reasons, the monitoring of agricultural images cannot reach a high level, so the image quality is limited. But plant growth monitoring, plant diseases, plant identification, maturity and classification of the fruits. We all need a clear detailed representation of the leaves and fruits in crop
monitoring images. In addition, with the development of image SR, Dong[28], Shi[29] et al. added the feature input of the original LR image and the introduction of residual learning [30]. The network depth and performance of SR were further improved. Therefore, we chose residual dense network (RDN) to extract the features of the original LR image. The network has enough depth to extract the hierarchical features of the image and can also adaptively fuse the hierarchical features and global features. Therefore, we can make full use of the information in the agricultural image to make the image SR have a better effect, so as to meet our agricultural use needs. The results will be presented in section 4.

3. The proposed network structure

3.1. Network Structure
As shown in figure 1, this RDN is mainly composed of four parts: shallow feature extraction network (SFENet), residual dense block (RDBs), dense feature fusion (DFF), and UPNet. We take ILR and ISR as the input and output of RDN. To be specific, we first used two Conv layers to extract the shallow feature \( B^{-1} \) and \( B^0 \) we entered in the image, the result \( B^0 \) is then used as input to the residual block RDBs. In this case, we can represent \( B^{-1} \) and \( B^0 \) as:

\[
B^{-1} = F_{SFE\cdot1}(I_{LR}) \quad (1)
\]
\[
B^{0} = F_{SFE\cdot2}(B^{-1}) \quad (2)
\]

Where \( F_{SFE\cdot i}(\cdot) \) represents the \( i \)-th convolution operation in the shallow feature extraction network. Next, we assume that there are \( M \) residual dense blocks, so the input and output of the \( m \)-th RDB can be expressed as:

\[
B_{m} = F_{RDB\cdot m}(B_{m-1}) \quad (3)
\]

Where \( F_{RDB\cdot m}(\cdot) \) represents the convolution operation of the residual-dense block. And \( m \) is the \( m \)-th RDB operation. \( F_{RDB\cdot m}(\cdot) \) is the operation of a compound function, like the convolution operation and rectified linear units (ReLU) [31]. Then we extract the hierarchical features of RDBs and conduct a dense feature fusion (DFF), which includes global feature fusion (GFF) and global residual learning (GRL). DFF makes full use of the characteristics of all the previous layers and makes a corresponding fusion, which can be expressed as:

\[
B_{DF} = B_{GF} + B_{-1} = F_{DFF}(B_{-1},B_{0},B_{1},\cdots,B_{M}) \quad (4)
\]

Where \( B_{DF} \) is the feature image output after dense feature fusion (DFF) obtained by composite function \( F_{DFF}(\cdot) \). After extracting local and global features in LR space, we stacked an UPNet in SR space. For reference [17], we use ESPCNN [29] in UPNet, followed by a convolution layer. In this way, the output of the entire network can be expressed as:

\[
I_{SR} = F_{UP}(B_{DF}) \quad (5)
\]

Where \( F_{UP}(\cdot) \) represents the compound operation carried out in the up-sampling process.
3.2. Residual Dense Block
The dense residuals block contains the connection layer, local feature fusion (LFF) and local residuals learning (LRL), and can carry out continuous signal memory (CM). CM means N convolution operations in each RDB block, the characteristic graphs used for the n-th convolution operation are all from the series of the previous RDB block and n-1 convolution operation. Therefore, the features of each layer can be fully utilized. After N times of convolution operation in the current RDB block, another local feature fusion will be conducted in order to adaptively fuse the information of the previous RDB block and all current convolution layers. Meanwhile, in order to reduce the number of feature graphs after fusion, according to the article [27], we use a convolution layer of 1 × 1 to control the output information adaptively. Since there are multiple convolution layers in each RDB block, so we end up with a local residual learning, That's the output $B_{m-1}$ of the last RDB block. This will further improve the expression and performance of the network.

3.3. Dense Feature Fusion
After a series of RDBs operations, we extracted local dense features, and then we should carry out dense feature fusion (DFF) to make good use of global layered features. DFF includes global feature fusion (GFF) and global residual learning (GRL).

3.3.1. Global feature fusion. This is mainly used to fuse and extract all the hierarchical features generated by residual dense blocks (RDB) to get $B_{GF}$. Firstly, the feature graph generated by $B_1, \ldots, B_M$ is connected in series, and then, like the residual dense block, a convolution layer of 1 × 1 is introduced to adaptively fuse the feature information of different layers. Then, a convolution layer of 3 × 3 is introduced in order to better extract features from global residual learning. This can be referred to [32].

3.3.2. Global residual learning. This is used to obtain the feature graph before entering the up-sampling network, such as formula (4). The previously mentioned residual dense blocks(RDBs) has made full use of the feature information of each layer in the block, and then we carried out global feature fusion to adaptively fuse the RDB feature information of all different layers, and finally we got $B_{DF}$.

3.4. Implementation Details
In this RDN, the kernel size of all our convolution layers is 3 × 3. Except in feature fusion, the specific purposes kernel size is 1 × 1 convolution layer. Where the convolution layers of kernel size is 3 × 3, we pad zeros to each side of the input to keep size fixed. The shallow feature extraction layers, local and global feature fusion layers all have $G_0=64$ filters. For all RDB blocks, all convolution layers have $G_0$ filters, followed by ReLU [31]. According to the article [17], we used ESPCNN [29] to promote the low resolution feature to the high resolution of UPNet. The final convolution layer is the output of 3 channels to ensure our SR color image.

4. Experimental

4.1. Datasets and Metrics
In this section, the main purpose is to verify the effect of this network structure on agricultural image restoration. We specially selected the agricultural greenhouse, now this kind of more representative fine agricultural site with monitoring function. Among the monitoring images collected by us, there are monitoring images of the growth of group vegetables in the greenhouse, as well as a series of monitoring images of the growth of individual plants and fruits. It is closer to the actual needs of agricultural monitoring occasions. Because the existing fine agriculture cannot reach the level of general ultra-clear monitoring, the data sets we made are all images of hundreds of thousands of phase
elements. There are 500 training images and 100 test images. The final result of SR is evaluated by the Y channel (brightness) PSNR [33] value obtained from YcbCr spatial transformation.

4.2. Experimental Settings

In order to compare and reflect the processing results of agricultural images by this network, we used the bicubic option in Matlab function as a degradation model to simulate LR images. This method is called BI. This model can simulate agricultural LR images with a scaling factor of \( \times 2, \times 3, \times 4 \). Then according to Settings [17], in each batch of training, we randomly extract 64 LR RGB patches with the size of 16 \( \times 16 \) as inputs. Then we added patches with random horizontal and vertical moves and 90° rotations. Repeated back propagation of 1000 times constitutes an epoch. We used the Torch7 framework and updated it with Adam optimizer [34]. The learning rate for all layers is initialized to \( 10^{-4} \), and cut it in half every 200 periods. All experiments are carried out using the pytorch framework on NVIDIA GTX2080 GPUs.

4.3. Image testing

We carried out experiments and tests with the above methods. There are basic network parameters: the number of RDB blocks (referred to as M), the number of convolutional layers in each RDB block (referred to as N) and the growth rate (referred to as G). The relationship between them is that larger M and N will make the network perform better. This is because the increase of M and N makes the network structure deeper. In addition, the existence of local feature fusion (LFF) enables us to have a higher growth rate (G), and the increase of G also helps improve the performance of the network. In this experiment, our \( M=16, N=8, G_0=64 \). Table 1 shows the quantitative comparison of SR results of degradation models with different scaling factors. Figure 2 shows the SR results of picture "505", which represents the monitoring of specific plants or fruits in agricultural greenhouses. As for the pictures which are processed by bicubic \( \times 4 \), the edge contours and the surface patterns are blurred. However, after our SR treatment, local surface details can still be restored.

Table 1. SR results of different scaling factors in BI degradation model.

| Model               | Scale | PSNR  |
|---------------------|-------|-------|
| RDN(M=16,N=8,G0=64) | 2     | 32.52 |
| RDN(M=16,N=8,G0=64) | 3     | 28.09 |
| RDN(M=16,N=8,G0=64) | 4     | 26.26 |

5. Conclusions

In this paper, we propose a method which uses a deep residual-dense network to realize the super-resolution of the existing agricultural monitoring images. The network can fully learn the global and local features of images and train them stably. It can not only recover the complex information of agricultural image, but also has a good effect on the detail processing.
Acknowledgement
This work was supported by the CERNET Innovation Project (No.NGII20170605) and Key Laboratory of Agricultural Information Engineering of Sichuan Agricultural University.

References
[1] Li, S. k., Zhang, X. (1998) Studies on the technique of multi-media image processing for plant shape information of crop. Acta Agronomica Sinica, 24, pp: 265-271.
[2] Zou, W. W., P. C. Yuen. (2012) Very low resolution face recognition problem. TIP.
[3] Shi, W., Caballero, J., Ledig, C., Zhuang, X., Bai, W., Bhatia, K., de Marvao, A. M. S. M., Dawes, T., ORegan, D., Rueckert, D. (2013) Cardiac image super-resolution with global correspondence using multi-atlas patchmatch. In MICCAI, pp: 9-16.
[4] Karras, T., Aila, T., Laine, S., Lehtinen. J. (2017) Progressive growing of gans for improved quality, stability, and variation. submitted to ICLR 2018.
[5] Zhang, L., Wu, X. (2006) An edge-guided image interpolation algorithm via directional filtering and datafusion. IEEE Transactions on Image Processing, pp: 2226-2238.
[6] Zhang, K., Gao, X., Tao, D., Li, X. (2012) Single image superresolution with non-local means and steering kernel regression. IEEE Transactions on Image Processing, pp: 4544-4556.
[7] Timofte, R., De, V., Gool, L. V. (2013) Anchored neighborhood regression for fast example-based super-resolution. In ICCV.
[8] Timofte, R., Smet, V. De., Van Gool, L. (2014) A+: Adjusted anchored neighborhood regression for fast super-resolution. In ACCV, pp: 111-126.
[9] Peleg, T., Elad, M. (2014) A statistical prediction model based on sparse representations for single image super-resolution. IEEE Transactions on Image Processing, pp: 2569-2582.
[10] Dong, C., Loy, C. C., He, K., Tang, X. (2014) Learning a deep convolutional network for image super-resolution. European Conference on Computer Vision, pp: 184-199.
[11] Schulter, S., Leistner, C., Bischof, H. (2015) Fast and accurate image upscaling with super-resolution forests. In CVPR.
[12] Huang, J.-B., Singh, A., Ahuja, N. (2015) Single image superresolution from transformed self-exemplars. In CVPR.
[13] Kim, J., Kwon Lee, J., Mu Lee, K. (2016) Accurate image superresolution using very deep convolutional networks. In CVPR.
[14] Tong, T., Li, G., Liu, X., Gao, Q. (2017) Image super-resolution using dense skip connections. In ICCV.
[15] Zhang, K., Zuo, W., Zhang, L. (2018) Learning a single convolutional super-resolution network for multiple degradations. In CVPR.
[16] Kim, J., Kwon Lee, J., Mu Lee, K. (2016) Deeply-recursive convolutional network for image super-resolution. In CVPR.
[17] Lim, B., Son, S., Kim, H., Nah, S., Lee, K. M. (2017) Enhanced deep residual networks for single image super-resolution. In CVPR.
[18] Pan, H. (2017) Intelligent greenhouse monitoring system based on android platform. Xi’an University of Science and Technology.
[19] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y. (2018) Residual Dense Network for Image Super-Resolution. In CVPR.
[20] Yang, Y. Q. (2015) Intelligent planting monitoring system based on Internet of things technology. Ocean University of China, Qing Dao.
[21] Li, K., Wu, Z., Peng, K.-C., Ernst, J., Fu, Y. (2018) Tell me where to look: Guided attention inference network. arXiv preprint arXiv:1802.10171.
[22] Zhang, H., Patel, V. M. (2018) Densely connected pyramid dehazing network. In CVPR.
[23] Zhang, H., Patel, V. M. (2018) Density-aware single image deraining using a multi-stream dense network. In CVPR.
[24] Zhang, H., Sindagi, V., Patel, V. M. (2017) Image de-raining using a conditional generative adversarial network. arXiv preprint arXiv:1701.05957.
[25] Zhang, K., Zuo, W., Gu, S., Zhang, L. (2017) Learning deep cnn denoiser prior for image restoration. In CVPR.
[26] Tai, Y., Yang, J., Liu, X. (2017) Image super-resolution via deep recursive residual network. In CVPR.
[27] Tai, Y., Yang, J., Liu, X., Xu, C. (2017) Memnet: A persistent memory network for image restoration. In ICCV.
[28] Dong, C., Loy, C. C., Tang, X. (2016) Accelerating the superresolution convolutional neural network. In ECCV, pp: 391-407.
[29] Shi, W., Caballero, J., Husz’ar, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., Wang, Z. (2016) Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In CVPR.
[30] He, K., Zhang, X., Ren, S., Sun, J. (2016) Deep residual learning for image recognition. In CVPR.
[31] Glorot, X., Bordes, A., Bengio, Y. (2011) Deep sparse rectifier neural networks. In AISTATS.
[32] Ledig, C., Theis, L., Husz’ar, 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. In CVPR.
[33] Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P. (2004) Image quality assessment: from error visibility to structural similarity. TIP.
[34] Kingma, D., Ba, J. (2014) Adam: A method for stochastic optimization. In ICLR.