Image Super-Resolution via Residual Blocks and Non-Negative Matrix Decomposition

Zhishuang Xue¹-a, Xiaofang Liu²-b, Hui Huang¹-c, Gen Zhang¹-d

¹School of Automation and Information Engineering, Sichuan University of Science and Engineering, Zigong, China
²School of Computer Science and Engineering, Sichuan University of Science and Engineering, Zigong, China

a465223378@qq.com, bxf1969@163.com, cschuang_hui@163.com, d465782728@qq.com

Abstract—In order to solve the problems that the existing super-resolution algorithms based on convolutional neural network do not fully utilize the feature information at all levels, which results in low reconstruction accuracy, and the deep network may result in the gradient disappearing, an image super-resolution reconstruction algorithm based on residual blocks and non-negative matrix decomposition is proposed. The proposed residual block can gradually learn the characteristics of the input image, the computational burden of the network is shared on each single residual block, thus the network complexity decreases. The input image and output image are fused by non-negative matrix decomposition to obtain more detail information of high-resolution image. Experiments on the benchmark dataset show that the proposed algorithm improves the PSNR, structural similarity and visual effect.

1. INTRODUCTION

The 5G era is an era of big data, pictures and videos [1] will become the main carriers for people to convey information. Low-resolution pictures and videos are used at the sending end, while high-resolution pictures and videos are recovered at the receiving end by super-resolution method, which will occupy less channel resources [2].

In recent years, the image super-resolution reconstruction (SR) algorithm based on deep learning [3-5] has become a hot topic for scholars. The main idea of these methods is to construct the mapping model from low-resolution image to high-resolution image by using the statistical prior knowledge obtained from large-scale datasets. The first application of convolutional neural network technology in the field of SR was Super-Resolution Convolutional Neural Network (SRCNN) [6] algorithm proposed by Dong et al. However, many details are still missing in the reconstructed image due to the small network layer and small network perception. Then, Dong et al. applied the deconvolution layer into SR and proposed Fast Super-Resolution Convolutional Neural Network (FSRCNN) [7], which realized the end-to-end reconstruction of the image. Kim et al. used skip-connection to deepen the number of convolutional layers, increased the receptive field and fitting ability of the network, and put forward Super-Resolution Using Very Deep Convolutional Networks (VDSR) [8]. Then, Lai et al. further used Laplacian super-resolution network (LAPSRN) [9] to integrate into the Laplacian pyramid, and utilized
the deconvolution layer to up-sample the residual images learned from each layer of the pyramid subnetwork gradually, and the reconstruction effect was good.

However, the above methods fail to make full use of the information extracted by convolution at all levels and is insensitive to tiny details, resulting in low reconstruction accuracy. Moreover, due to the large number of network layers and complex structure, network training is very difficult and difficult to converge quickly. Combined with the advantages of skip-connection that does not increase residual network parameters and non-negative matrix that can integrate more feature information, this paper proposes an image super-resolution reconstruction method based on residual blocks and non-negative matrix. The main contributions of this paper are summarized as follows:

1. The residual block with skip connection is introduced, which gradually inputs the learned image features into the network, improves the capability of feature reconstruction of low-resolution image, and highly restores the missing high-frequency semantic information of the image. At the same time, with the help of residual block, we can accelerate the convergence speed of the network.

2. Non-negative matrix decomposition is introduced into the improved network structure for feature fusion, fuse global and local features, and this method makes the reconstructed image contain more prior information of input images, which is conducive to the enhancement of image details.

![Residual Block 1](image1)

![Residual Block 2](image2)

![Residual Block 3](image3)

![Conv](image4)

![Deconv](image5)

![Bicubic Interpolation+Non-Negative Matrix Decomposition](image6)

**Figure 1. The overall residual network structure**

2. **THE PROPOSED IMAGE SUPER-RESOLUTION METHOD**

2.1 *The proposed residual block*

The model based on residual network is trained on Pytorch platform, which contains 8 convolutional layers and 1 fusion layer, the last layer is the fusion layer. The overall residual network structure is shown in Fig. 1. The first layer is the convolutional layer, has 64 convolution kernels with size of 5×5. The second layer to the seventh layer is composed of three residual blocks, as shown in Fig. 1, in which each layer has 64 convolution kernels with size of 3×3. The eighth layer is the deconvolution layer, which contains a 9×9 convolution kernel. The residual image obtained from eighth convolutional layer are fused with the input image after bilinear interpolation in the fusion layer. The low level features and high level are fused by using non-negative matrix decomposition theory. The fused image will contain more details, and the image quality becomes higher.

91-image and BSD500 are chosen as the training datasets, and BSD300 are using for verification. We test the proposed method on subsampled images. Training process is conducted on a computer with an Intel i7-4790 CPU 3.6GHz and GTX 970 GPU. We only use the luminance channel of the input image to train the network. The luminance channel (Y) of the output image is used for calculate PSNR/SSIM. For chrominance channels (Cb, Cr) of the input image, we only use bicubic interpolation to upsample them. Finally, we convert the YCbCr color space to RGB color space to obtain the high-resolution image. For scale factor ×2, ×3 and ×4, the images are cropped to the size of 14×14, 13×13 and 10×10, respectively. To speed up the convergence process, adaptive Moment Estimation (ADAM) is employed. The parameters are set as: $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\alpha = 0.001$ and $\varepsilon = 10^{-8}$. The minimum batch size is 128 and the initial learning rate is $10^{-4}$.

Residual operation has a high processing efficiency in the process of image super-resolution. Fig. 2
shows the proposed residual block.

Figure 2. The proposed residual block

The residual block contains two convolutional layers and two activation layers. The residual block can be expressed as:

$$x_{i+1} = x_i + f(x_i, W_i)$$  \hspace{1cm} (1)$$

Where $x_i$ is the input of the i-th residual block, $W_i = \{w_{i,k} \mid 1 \leq k \leq 2\}$ is the set of weights and biases about the i-th residual block, and 2 is the number of convolutional layers in the residual block. $f$ is a residual function, $p$ is an activation function. Parametric Rectified Linear Unit (PReLU) is used here. In the model training stage, PReLU can bring smaller errors. PReLU can be expressed as:

$$p(x) = \max(x, 0) + \alpha \min(0, x)$$  \hspace{1cm} (2)$$

Where $x$ is the input of the activation function $p$, $\alpha$ is a constant that can be obtained from training.

For the convenience of description, we only consider $x$ is greater than 0 in Eq. (1). $x_{i+1}$ can be simplified as:

$$x_{i+1} = x_i + f(x_i, W_i)$$  \hspace{1cm} (3)$$

Multiple residual blocks are cascaded to obtain the following one:

$$x_K = x_1 + \sum_{m=1}^{K-1} f(x_m, W_m)$$  \hspace{1cm} (4)$$

It can be seen from Eq. (4) that between any block $i$ and block $K$, the model exists in the form of residual blocks, and Eq. (4) represents the sum of all previous residual functions.

Eq. (4) has good backward transfer ability, $\Phi$ is defined as a loss function, which can be obtained as follows:

$$\Phi_{x_i} = \Phi_{x_i}(1 + H_{x_i})$$  \hspace{1cm} (5)$$

$$H = \sum_{m=1}^{K-1} f(x_m, W_m)$$  \hspace{1cm} (6)$$

It can be seen from Eqs. (5) and (6) that $\Phi_{x_i}$ can be divided into $\Phi_{x_i}^'$ and $\Phi_{x_i} H_{x_i}^'$, but only $\Phi_{x_i}^'$ can transmit information through the convolutional layer directly. $\Phi_{x_i}^'$ transmits information directly to the next residual block. $H_{x_i}^'$ cannot be -1, so it cannot be eliminated. In other words, even if the weights are very small, the gradients of the convolutional layers will not disappear. Whether they are forward or backward, features can be transmitted from one residual block to any other residual block.

2.2 Non-negative Matrix Decomposition
The non-negative matrix decomposition theory is applied to the new residual network structure for the first time, and an image super-resolution reconstruction method based on global and local feature fusion
is proposed. The image fusion method makes the reconstructed image contains more low-frequency information, which is helpful to supplement the complete information of the image.

Based on the non-negative matrix image fusion theory, the image can be expressed as:

\[ V = WH + n \quad (7) \]

Where \( V \) is the observed image, \( W \) is the fused image, \( H \) is the weight, \( n \) is the noise. \( V \) is composed of high-resolution and low-resolution images, each of its columns contains an image, namely:

\[ V = [v_1, v_2] \quad (8) \]

\( V \) can be approximated by the product of \( W \) and \( H \). If \( v \) and \( h \) are the corresponding column vectors of \( V \) and \( H \), then:

\[ v = Wh \quad (9) \]

The non-negative matrix \( W \) can be seen as a set of bases which composes of high and low resolution image datasets \( V \), and \( H \) is the projection coefficient of \( V \) onto \( W \). When the dimension of the basis vector is 1, \( W \) can be represented as a fused image.

### TABLE 1. OBJECTIVE EVALUATION INDEXS

| Image            | Index | SRCNN | FSRCNN | VDSR  | LAPSRCNN | Ours   |
|------------------|-------|-------|--------|-------|----------|--------|
| Bird             | PSNR  | 36.71 | 37.04  | 37.52 | 37.54    | 37.76  |
|                  | SSIM  | 0.954 | 0.955  | 0.958 | 0.959    | 0.961  |
| Woman            | PSNR  | 36.73 | 37.06  | 37.53 | 37.52    | 37.78  |
|                  | SSIM  | 0.955 | 0.956  | 0.959 | 0.961    | 0.960  |
| Remote Sensing   | PSNR  | 34.56 | 34.89  | 35.01 | 35.25    | 35.44  |
|                  | SSIM  | 0.861 | 0.872  | 0.879 | 0.881    | 0.883  |

3. EXPERIMENTS

The experiment is implemented on an Inter Core I7 eight-core processor, 16GB memory, 1T hard disk and Windows operating system, and MATLAB R2017a is adopted as the simulation software.

In the test experiment, the scale factor is set as 2. The image quality evaluation index adopted subjective evaluation and objective evaluation. Figs. 3-5 show the experimental results of some images.

Fig. 3 shows the results of Bird image. (a) is the input low-resolution image, (b) is the result of SRCNN, (c) is the result of FSRCNN, (d) is the result of VDSR, (e) is the result of LAPSRN, (f) is the result of the proposed method. It can be seen from Fig. 3 that the result of SRCNN is worsen than other methods, especially in edges and textures, and the whole picture is not clear after processing. Compared with SRCNN, texture reconstruction of our method is closer to the original image, and obvious edge diffusion phenomenon appears in SRCNN around the eyes. FSRCNN, VDSR and LAPSRN have little difference after processing, all the texture around the eyes are distorted. Therefore, compared with the traditional algorithms, the reconstruction effect of the proposed method is obviously improved.
Fig. 3 gives the results of Bird image. Compared with VDSR, the overall effect of the proposed method is better, there are obvious distortion in VDSR, facial color and contour compared with the original image is not true. At the same time, look at the texture area around the cuff, there is an obvious distortion in VDSR. See the eye areas, the results of other reconstruction methods all have blurred eyes, but the eyes processed by the proposed method are clearer and more similar to the real situation.

Fig. 4 gives the results of Woman image. Compared with VDSR, the overall effect of the proposed method is better, there are obvious distortion in VDSR, facial color and contour compared with the original image is not true. At the same time, look at the texture area around the cuff, there is an obvious distortion in VDSR. See the eye areas, the results of other reconstruction methods all have blurred eyes, but the eyes processed by the proposed method are clearer and more similar to the real situation.

Fig. 5 shows the results of remote sensing images in mountainous areas, and the images on the right are local enlarged images of the cloud areas. Compared with other methods, it can be seen that after the super-resolution reconstruction of the proposed method, the high-frequency texture details are more abundant and the ground object information is easier to obtain. Further data mining and the extraction of near-infrared, red and green bands information can quickly determine the location of mountains and cloud areas. The researchers can better judge the cloud cover, predict the trend of atmospheric motion.
In order to further verify the reconstruction effect, objective evaluation indexes are adopted to measure the quality of the reconstructed images. The objective evaluation is determined by PSNR value and SSIM value. The higher the value is, the better the reconstruction effect will be. Tab. 1 shows the values of PSNR and SSIM for different methods.

It can be seen from Tab. 1 that the values of the proposed method is higher than that of other comparison methods, no matter from PSNR or SSIM. Therefore, the objective evaluation index can also indicate the superiority of the super-resolution reconstruction performance of the proposed method.

4. CONCLUSIONS
An image super-resolution reconstruction method based on residual blocks and non-negative matrix decomposition is proposed. The proposed residual block combines different network layers by means of skip-connection and has the ability to learn high frequency information. The non-negative matrix decomposition is used to fuse the high frequency information and the low frequency information to make the image information more complete. The future work will continue to optimize the network structure and improve the ability to reconstruct the detail features of the model.
ACKNOWLEDGMENT
This research was supported in part by Sichuan Science and Technology Program under Grant 2017GZ0303, in part by Fund Project of Sichuan Provincial Academician (Expert) Workstation under Grant 2016YSGZZ01, and in part by Special Fund for Training High Level Innovative Talents of Sichuan University of Science and Engineering under Grant B12402005.

REFERENCES
[1] O. K. Wade and J. B. Woehrstein, et al, "124-Color super-resolution imaging by engineering DNA-PAINT blinking kinetics," Nano letters, vol. 19, no. 4, pp. 2641-2646, Mar. 2019.
[2] A. Lucas and S. Lopez-Tapia, et al, "Generative adversarial networks and perceptual losses for video super-resolution," IEEE Transactions on Image Processing, vol. 28, no. 7, pp. 3312-3327, Jan. 2019.
[3] C. Wang and W. Gao, et al, "Single Image Super-Resolution Based on Cascaded Recursive Residual Convolutional Neural Network," 2019 IEEE 11th International Conference on Advanced Infocomm Technology (ICAIT). IEEE, 2019, pp. 113-117.
[4] K. H. Viet and J. Ren, et al, "Deep learning based single image super-resolution: a survey," International Journal of Automation & Computing, vol. 46, no. 4, pp. 413-426, Jul. 2019.
[5] Y. Wang and L. Wang, et al, "End-to-end image super-resolution via deep and shallow convolutional networks," IEEE Access, no. 7, pp. 31959-31970, Mar. 2019.
[6] C. Dong and C. C. Loy, et al, "Learning a deep convolutional network for image super-resolution," European conference on computer vision. Springer, 2014, pp. 184-199.
[7] C. Dong and C. C. Loy, et al, "Accelerating the super-resolution convolutional neural network," European conference on computer vision. Springer, 2016, pp. 391-407.
[8] J. Kim and L. J. Kwon, et al, "Accurate image super-resolution using very deep convolutional networks," IEEE conference on computer vision and pattern recognition. IEEE Computer Society, 2016, pp. 1646-1654.
[9] W. S. Lai and J. B. Huang, et al, "Deep laplacian pyramid networks for fast and accurate super-resolution," Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, 2017, pp. 5835-5843.