Single Image Super-resolution Reconstruction Based on Enhanced Residual Network

Chunyu Liu, Wenhua Qian*, Dan Xu, Mengjie Jiang and Xiaojin Li
School of Information Science and Engineering, Yunnan University, Yunnan Kunming, 650500, China

*Corresponding author email: whqian@yun.edu.cn

Abstract. High-resolution images present richer detailed information and have stronger information expression capabilities. The increase of the network depth does not guarantee that the reconstructed image has a higher quality, and may cause problems such as overfitting. So this article proposes an enhanced residual network, which can fully extract input low-resolution image features and reconstruct high-resolution images. This paper introduces a deconvolution operation based on the residual module to expand the size of input features, and the connection between different modules promotes feature fusion, obtains more high-frequency details from the input low-resolution image. The objective experimental results show that the proposed method has improved the indicators PSNR and SSIM. In terms of visual effects, it can reconstruct clearer and more detailed images.

Keywords: Image super-resolution; Fusion; Residual learning; Deconvolution.

1. Introduction

Single image super-resolution reconstruction uses image processing and machine learning techniques to reconstruct the observed low-resolution images to reconstruct corresponding high-resolution images with more detailed information. At present, deep learning has been introduced to SISR, which has promoted the development of SISR. In 2014, Dong et al. [1] introduced deep learning to super-resolution reconstruction for the first time (SRCNN), compared with previous work, this algorithm has achieved excellent performance. Kim et al. [2] deepened the network depth to improve the effect of image reconstruction. Later, Dong et al. [4] improved SRCNN. At the end of the model, the deconvolution layer is used to enlarge the image size, which improves the calculation speed compared with SRCNN. Li et al. [3] used image feature information of different scales for image reconstruction. Dai et al. [5] improved the attention network, and the results proved that the SAN network has improved in terms of objective indicators and visual quality. Liu et al. [10] proposed a fusion residual module structure to solve the limitation that residual information is not fully utilized. Although super-resolution models based on CNNs can achieve good reconstruction results, there are still some limitations. Due to the limited receptive field, there is a lack of sufficient context information. In order to improve the performance of the network, the current models deepen the network, but it will also bring some new problems, for example, some key features will disappear during the propagation process, which hinders the expressive ability of the network.

This paper proposes an enhanced residual network. The main improvements are as follows: The model proposed adopts a feature fusion structure between levels, so that the network can make full use of the hierarchical features of the image and enhance the expressive ability of the network; The deconvolution operation can expand the size of the feature, so that the network can extract more useful
details, the simultaneous use of deconvolution layer and convolution layer can not only take into account the global upper and lower information, but also can effectively extract rich texture detail information; Since the end-to-end learnable upsampling layer is used, the computational complexity is reduced. At the same time, it avoids problems such as noise amplification and visible artifacts caused by methods of preprocessing operations (such as using bicubic interpolation).

2. Proposed Method

The network model proposed in this paper is shown in Figure 1. Its main structure is composed of a head module, a backbone module, and a reconstruction module.

![Figure 1. The architecture of our proposed model.](image)

2.1. Network Architecture

The head module mainly performs initial feature extraction. Assuming that represents the input low-resolution image, the shallow features of are extracted through the convolutional layer and activation function:

$$E_0 = \delta(H(F_{LR}))$$

Where $H$ and $\delta$ represent the convolution operation and activation function respectively, $E_0$ is the extracted feature map, and then $E_0$ continues to propagate to the backbone module to extract more detailed features. The backbone module is composed of $N$ enhanced residual modules:

$$E_n = \mathcal{R}_n(E_{n-1}) = \mathcal{R}_n(\mathcal{R}_{n-1}(\ldots(\mathcal{R}_0(E_0))))$$

Where $\mathcal{R}_n$ represents the $n$-th enhanced residual module function, $E_{n-1}$ is the input of the $n$-th module, and $E_n$ is the corresponding output. Then, $1 \times 1$ convolution integrates the extracted features, and the useful hierarchical information contained in the previous module is transmitted to the end of the network to generate more representative features:

$$F_{up} = \delta(H(E_0, E_1, E_2, \ldots, E_n))$$

Where $\delta$ represents the activation function, $H$ is the convolution operation, $E_i (i \neq 0)$ represents the output of the $i$-th enhanced residual module, $E_0$ is the output of the first convolutional layer, $[E_0, E_1, E_2, \ldots, E_n]$ represents the connection operation, and $F_{up}$ is the output of the corresponding hierarchical feature fusion structure. Finally, super-resolution reconstruction is carried out through the reconstruction part, and the extracted deep feature map is enlarged using the sub-pixel cleaning layer, and then the up-sampled feature map is reconstructed into the final SR image through a convolutional layer:

$$I_{SR} = P(F_{up}) = G(I_{LR})$$

Where $P$ is the reconstruction function, $G$ is the entire SR network function, and $I_{SR}$ is the
reconstructed super-resolution image.

![Reconstructed Super-Resolution Image](image)

**Figure 2.** Left: The structure of residual module. Right: The structure of enhanced residual module.

### 2.2. Enhanced Residual Module

Figure 2 (left) describes the basic residual modules used in image super-resolution. The current work is almost choosing to connect the residual blocks together to increase the depth or breadth of the network, so as to obtain more contextual information in the image, but at the same time it will also bring about a significant increase in storage and computing costs. In this paper, the residual module is enhanced to solve the above problems, in order to achieve a more powerful feature representation. Figure 2 (right) shows the specific structure of the enhanced residual module. The residual module consists of two sets of deconvolution layers and convolution layers. The deconvolution layer can expand the feature size and therefore make the extracted features more representative. The deconvolution and convolution process in the enhanced residual module can be expressed by the following formula:

\[ E_n = \delta(H^1(\delta(D^1(E_{n-1})))) \]

\[ E_n = H^2(\delta(D^2(E_n))) \]

Where \( E_n \) represents a deconvolution and convolution operation, \( H^i \) represents the i-th convolution operation, \( D^i \) represents the i-th deconvolution operation, and \( E_n \) is the output of an enhanced residual module.

### 3. Experiments

#### 3.1. Settings

The experiment in this paper is based on the DIV2K data set, using 800 images with clear details to train the proposed model, and expanding the data set through rotation, cropping, etc. The Set5, Set14, BSD100, and Urban100 standard image super-resolution reconstruction evaluation data sets are selected as the experimental test data. In order to obtain a pair of LR and HR image training sets, the corresponding LR image is obtained by bicubic downsampling for each initial HR image block. This article selects the L2 loss function for optimization, the formula of the L2 loss function is:

\[ \ell_{SR}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \| H_{\theta}(I_{LR}^i) - I_{HR}^i \|_2^2 \]

#### 3.2. Comparison and Analysis of Experimental Results

In order to prove the effectiveness of the proposed method, this article compares the proposed model with other SR models. The comparison methods include traditional algorithms (Bicubic, A+[6]), and algorithms based on deep learning (SRCNN[1], FSRCNN [4], VDSR[2], DRCN[8], LapSRN[7], DRRN[9]).

Table 1 compares the quantitative index values obtained by different models trained on different data.
test sets. It can be seen from the table that the proposed model based on the enhanced residual network can improve the expression performance of the network and meet the human requirements for image quality.

Figure 3 shows the visual comparison between the proposed model and other 4x super-resolution works. The test image is selected from the Set14 and Set5 data sets. The model proposed in this paper can reconstruct higher-quality images, it is clear that the reconstructed image can restore sharper and clearer lines and edges, showing more texture details. From the comparison of objective indicators and subjective effects, we can see that the method based on the enhanced residual network that we propose can effectively extract image features and restore more high-frequency details, allowing people to observe some details more clearly.

Table 1. PSNR and SSIM results of different models.

| Model  | Scale | Set5 PSNR/SSIM | Set14 PSNR/SSIM | BSD100 PSNR/SSIM | Urban100 PSNR/SSIM |
|--------|-------|----------------|-----------------|------------------|-------------------|
| Bicubic | ×4    | 28.42/0.810    | 26.10/0.704     | 25.96/0.669      | 23.15/0.659       |
| A+     | ×4    | 30.30/0.859    | 27.43/0.752     | 26.82/0.710      | 24.34/0.720       |
| SRCNN  | ×4    | 30.49/0.862    | 27.61/0.754     | 26.91/0.712      | 24.53/0.724       |
| FSRCNN | ×4    | 30.71/0.865    | 27.70/0.756     | 26.97/0.714      | 24.61/0.727       |
| VDSR   | ×4    | 31.35/0.882    | 28.03/0.770     | 27.29/0.726      | 25.18/0.753       |
| DRCN   | ×4    | 31.53/0.884    | 28.04/0.770     | 27.24/0.724      | 25.14/0.752       |
| LapSRN | ×4    | 31.54/0.885    | 28.19/0.772     | 27.32/0.728      | 25.21/0.756       |
| DRRN   | ×4    | 31.68/0.889    | 28.21/0.772     | 27.38/0.728      | 25.44/0.764       |
| Ours   | ×4    | 31.88/0.901    | 28.54/0.797     | 27.54/0.755      | 25.95/0.79        |
| Bicubic | ×3    | 30.39/0.868    | 27.55/0.774     | 27.21/0.739      | 24.46/0.736       |
| A+     | ×3    | 32.58/0.909    | 29.13/0.819     | 28.29/0.784      | 26.03/0.797       |
| SRCNN  | ×3    | 32.75/0.909    | 29.28/0.821     | 28.41/0.786      | 26.24/0.799       |
| FSRCNN | ×3    | 33.16/0.914    | 29.43/0.824     | 28.52/0.790      | 26.42/0.807       |
| VDSR   | ×3    | 33.66/0.921    | 29.77/0.831     | 28.82/0.798      | 27.14/0.828       |
| DRCN   | ×3    | 33.82/0.922    | 29.76/0.831     | 28.80/0.796      | 27.15/0.828       |
| LapSRN | ×3    | 33.78/0.921    | 29.87/0.833     | 28.81/0.797      | 27.06/0.827       |
| DRRN   | ×3    | 34.03/0.924    | 29.96/0.835     | 28.95/0.800      | 27.53/0.838       |
| Ours   | ×3    | 34.10/0.933    | 30.24/0.855     | 29.02/0.822      | 27.99/0.860       |

Figure 3. Results of different methods with a magnification factor of 4.
4. Conclusion
This paper proposes a model to enhance the residual network, the purpose is to improve the existing super-resolution reconstruction results in the blur of details and curve blur. By introducing the deconvolution layer to extract the features of the input LR image, the size of the input image feature is expanded to use more context information, and the feature fusion between different levels allows the features of the previous layer to be continuously used, which further ensures reconstruction a super-resolution image with high frequency details.

Acknowledgements
This research was funded by the grants (No. 61662087, 61761046) from the Research Natural Science Foundation of China, the Research Foundation of Yunnan Province (No. 2019FA044), Provincial Foundation for Leaders of Disciplines in Science and Technology (2019HB121), the Postdoctoral fund of the Ministry of education of China (No. 2017M621591), Jiangsu Planned Projects for Postdoctoral Research Funds in 2017.

References
[1] Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution[C]//European Conference on Computer Vision, Zurich, Switzerland, Sep 6-12, 2014. Springer-Verlag, 2014: 184-199.
[2] Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1646-1654.
[3] Li J, Fang F, Mei K, et al. Multi-scale residual network for image super-resolution[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 517-532.
[4] Dong C, Loy C C, Tang X. Accelerating the super-resolution convolutional neural network[C]//European conference on computer vision. Springer, Cham, 2016: 391-407.
[5] Dai T, Cai J, Zhang Y, et al. Second-order attention network for single image super-resolution[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2019: 11065-11074.
[6] Timofte R, De Smet V, Van Gool L. A+: Adjusted anchored neighborhood regression for fast super-resolution[C]//Asian conference on computer vision. Springer, Cham, 2014: 111-126.
[7] Lai W S, Huang J B, Ahuja N, et al. Deep laplacian pyramid networks for fast and accurate superresolution[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 624-632.
[8] Kim J, Kwon Lee J, Mu Lee K. Deeply-recursive convolutional network for image super-resolution[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1637-1645.
[9] Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 3147-3155.
[10] Liu J, Zhang W, Tang Y, et al. Residual Feature Aggregation Network for Image Super-Resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 2359-2368.