Journal of Electronics and Information Science (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/jeis.2023.080107
ISSN 2371-9524 Vol. 8 Num. 1

Image super-resolution reconstruction based on residual compensation combined attention network

Xiyao Li[1,a,*]

[1]*College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou, 730050, China*
[a]*2330125984@qq.com*
[*]*Corresponding author*

Keywords: Super-resolution reconstruction; Convolutional neural networks; Deep separable convolution; Residual networks; Attention mechanisms

Abstract: For image reconstruction, the residual network ignores part of the residual information when extracting features. We propose an image super-resolution reconstruction based on residual compensation joint attention network (RCCN). Firstly, we construct a three-way residual network for compensating the feature information of the standard residual network; secondly, we design a joint attention module to complement the pixel-level image attention information by 3D attention while the channel attention learns the channel weight information; finally, our method has clearer results compared with other advanced methods, and the objective evaluation indexes are all greatly improved.

1. Introduction

Single image super-resolution is a classical image recovery problem[1] that aims to recover high resolution images from degraded low-resolution images. Current single-image super-resolution reconstruction techniques can be divided into three categories: interpolation-based methods[2], reconstruction-based methods[3] and learning-based methods[4]. In recent years, deep learning methods have developed rapidly and have shown great potential in the field of computer vision. Dong et al [5]first applied deep learning to the image super-resolution problem and proposed super-resolution reconstruction by convolutional neural networks, which achieved end-to-end mapping between LR and HR, but the SRCNN introduced additional computation using pre-up-sampling and the 3-layer convolution also resulted in limited extracted information. To address this problem, Dong et al [6]propose a super-resolution reconstruction based on fast convolutional neural networks, which use deconvolutional layers instead of dual cubic interpolation in the upsampling process and increase the depth of the network from three to eight layers. Later, many super-resolution reconstruction algorithms based on deep neural networks were proposed. Lim et al [7]proposed super-resolution reconstruction based on enhanced depth residual networks, and EDSR removed the BN layer while using the residual network, accelerating the convergence of the network.

Although all of the above deep learning-based SR methods have obtained good reconstruction results, there are still some problems. All of these methods ignore the fact that a lot of feature information is lost by single-path forward propagation, and although the use of residual networks can

alleviate this problem, there is also the problem of insufficient reconstruction details. Aiming at this problem, we propose an image super-resolution reconstruction with joint attention for residual compensation. The contributions of this paper are summarized as follows:

(1) To enhance the extraction of long-range features from residual networks, we introduce a parallel multi-path extraction module. This module can greatly enhance the detail extraction capability of the network and strengthen the generalization capability of the network model.

(2) A combined attention module has been designed by using a channel attention module and a pixel attention module. This module allows the recoding of feature information in the channel and pixel dimensions to enable adaptive selection of valid information by the network and suppression of interfering information.

(3) Experimental validation by using a standard test dataset. The experiments show that our method has good reconstruction performance and generalization capability.

2. Related work

2.1. Residual networks

In order to make deep neural network training easier, He et al. [8] proposed ResNet, which uses jumpers to connect adjacent feature layers to ensure that the feature information in the forward propagation process is not less, solves the problem of gradient disappearing in the network training process, and greatly deepens the deep learning network, the operation process can be expressed as:

$$x_{i+1} = F(x_i, \{W_i\}) + x_i \tag{1}$$

where x_i represents the input features divided into two paths, $F(\bullet)$ represents the residual mapping convolution operation and x_{i+1} represents the output features.

2.2. Attention mechanisms

The attention mechanism is an efficient feature selection mechanism. By generating the attention weighting function that focuses on the significant region of the feature and ignores the redundant feature, the accuracy of feature extraction can be improved by adding only a few parameters. Hu et al [9] proposed channel attention networks to improve the effective use of computational resources by adjusting the channel attention of network features so that the network focuses on useful features. Zhang et al [10] proposed that RCAN introduce channel attention and incorporate residual extraction so that the network guides the production of corresponding attention weights based on the image information of each channel. Zhao et al [11] extracted pixel attention (PA), produced three dimensional attention features to filter and introduce fewer additional parameters, and improved the reconstruction performance.

3. The method RCCN proposed in this paper

In this section, we introduce each module of the proposed method. Then, the whole framework of the proposed method is introduced.

3.1. Residual compensation combined attention network

In order to increase the ability of the residual network to extract feature information, this paper proposes an image super-resolution reconstruction algorithm with residual compensation combined with attention network. The overall framework of the network is shown in Figure 1. Our proposed

RCCN consists of three modules, containing a shallow feature extraction module, a nonlinear mapping module for residual compensation combined with attention and a reconstruction module.

The ILR and ISR represent the input and output of the network, respectively, and the initial features of the LR image are first extracted using standard convolution to extract shallow feature information, as described below:

$$F_0 = H_0(I_{LR}) \tag{2}$$

where $H_0(\cdot)$ is the convolution operation for the initial extracted features and F_0 is the initial extracted features.

Secondly, F_0 the feature information is extracted step by step through a composition of n RCCN end-to-end connections. In this module, features are extracted using a combination of residual compensation and combined attention, with residual compensation capturing some of the features lost by the residual network and improving high frequency reconstruction performance. The extraction process is shown in the following equation:

$$F_{LD} = C_B\left\{F_{CR_1}, F_{CR_2}, F_{CR_3}, \cdots F_{CR_n}\right\} \tag{3}$$

where $F_{CR_i}(i = 1, 2, 3, \cdots n)$ represents the i th CRUB module for feature extraction. $C_B\{\cdot\}$ indicates the operation of the n modules above using end connection fusion. This results in a feature image F_{LD} for deep feature extraction.

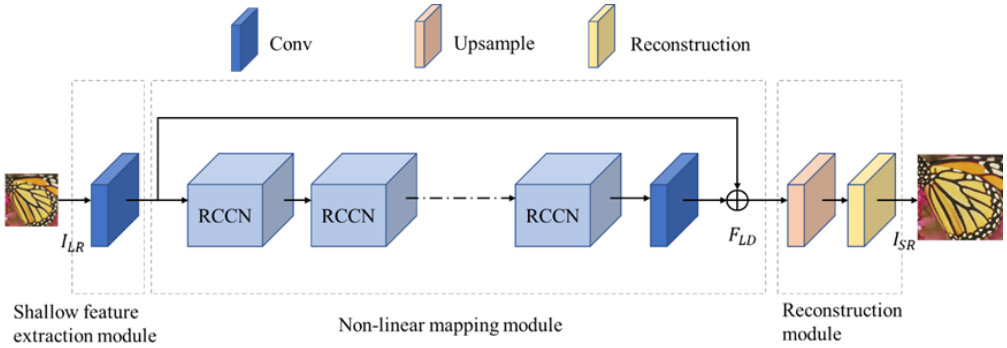

Figure 1: RCCN network structure diagram

Finally, the resulting deep image features are used as input for upsampling, by which the feature images are scaled to the desired magnification, i.e:

$$F_{UP} = H_{UP}(F_{LD}) \tag{4}$$

where H_{UP} represents the upsampling operation and F_{UP} represents the feature image at the desired magnification, after which F_{UP} is passed to the reconstruction module to obtain the corresponding SR image:

$$I_{SR} = H_{RE}(F_{UP}) \tag{5}$$

where $H_{RE}(\cdot)$ is the reconstructed mapping function and I_{SR} is the reconstructed high resolution image.

3.2. Residual compensation module

As shown in Figure 2. Our residual compensation network, using group convolution and depth-separable convolution as the basic units, introduces a residual structure in the forward propagation of

features to prevent the effects of gradient explosion and gradient disappearance. We then introduce channel shuffle, which reduces the effect of group convolution and depth-separable convolution on the network.

For the features of the input network, residual compensation is first performed in the low channel dimension in order to reduce the computational effort during the compensation process. Split the dimension into q_1-dimensional and q_2-dimensional features by using the channel dimension split function for features with input in dimension q, where dimension q_1 is equal to dimension q_2 and the sum of dimension q_1 and dimension q_2 is q. The purpose of this is to promote the fusion of information after the feature mapping dimension and increase the richness of the extracted information. The specific operation details are as follows:

$$F_i = S_{q_1}(F_P) = S_{q_2}(F_P) \tag{6}$$

Where $S_{q_1}(\cdot)$ and $S_{q_2}(\cdot)$ denote the channel splitting operation. Then, for the input features split into two, one way is used for long path information extraction, and the other way is retained after one convolution. The long path information extraction process starts with a group convolution with a convolution kernel of 1 and an activation function for feature mapping to extract shallow information. After performing a two-layer group convolution operation, and incorporating a depth-separable convolution, deeper detailed information can be extracted. When the image feature information F_i is passed into a layer of group convolution, the feature information F_{KL} is obtained.

$$F_{KS} = G_S^1(F_i) \tag{7}$$

$$F_{KL} = G_L^1(F_i) \tag{8}$$

where F_{KS} is the short path retention feature and F_{KL} is the feature information extracted from the first layer of long path information extraction,

$$F_{KE} = DW_C(G_L^2(F_{KL})) \tag{9}$$

where $G_L^2(\cdot)$ denotes the second layer group convolution with activation function operation, and $DW_C(\cdot)$ denotes the depth-separable convolution operation. And F_{KE} denotes the resulting long path information.

$$F_D = G_E(C(A_{CON}(F_{KS}, F_{KE}))) \tag{10}$$

where A_{CON} is the recovery function for the channel dimension, C represents the Channel Shuffle, and $G_E(\cdot)$ represents the 1×1 dimension-holding convolution.

Although low latitude residual compensation maintains the convolution operation at low latitudes, the reduced dimensionality results in incomplete feature information being extracted by the network. To obtain fuller feature information, we introduce a two-way constant dimensional residual compensation branch, shown in the right-hand two-way diagram. In the first branch of the constant dimensional residual compensation, we use the same convolution settings as in the long path information extraction part of the low latitude residual compensation, with the difference that the constant dimensional residual compensation does not compress the dimensions, i.e. no dimensional splitting of the input features is performed. In the second branch of the constant dimensional residual compensation, we use only one group convolution operation and one depth-separable convolution operation. To increase the exchange of information, both paths are followed by a channel-mixing operation, and then the feature information from the different branches is stacked together through

the Add layer to fuse all the feature information.

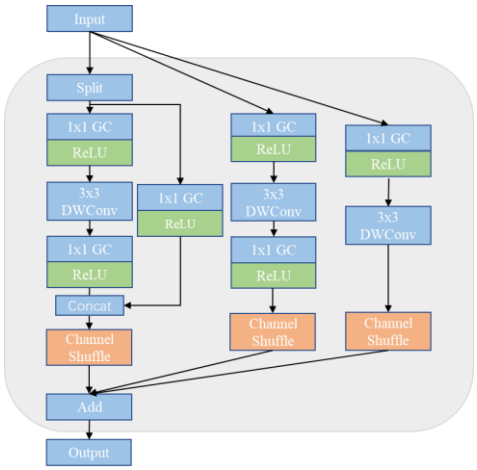

Figure 2: Structure of the residual compensation module

3.3. Combined Attention Module

In traditional convolutional neural networks, the features obtained from all convolutional layers are aggregated directly to get the output, which cannot effectively use the useful information in feature extraction and focus on information helpful for image detail recovery, thus limiting the learning efficiency of convolutional networks for feature information. For this purpose the combined attention module is designed in this paper and the structure is shown in Figure 3.

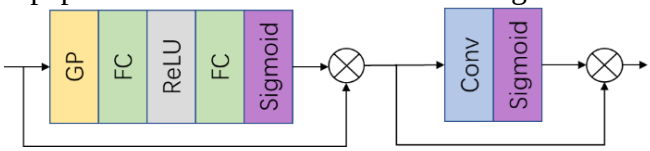

Figure 3: Structure of the combined attention module

In the figure, the input features go through the global average pooling, full connectivity, activation function, full connectivity and Sigmoid function in turn, and this is done to obtain information on the weights of the input features. Since a single attention can only get the weight share of one dimension of image features, for this reason, we introduce pixel-level attention and propose a combined attention module for channel-level and pixel-level feature attention on the input image. As shown in the second module in Figure 3, pixel-by-pixel multiplication gets pixel attention feature information.

4. Experimental results and analysis

4.1. Experimental setup

To evaluate the effectiveness and accuracy of the proposed algorithm, we experimented with LR images and HR images, using DIV2K [12] as the training dataset, and enhanced the data with 90 °, 180 °, 270 ° rotations and random horizontal flips on the training dataset. Four commonly used standard datasets Set 5 [13], Set 14 [14], B100 [15] and Urban100 [16] were used as test datasets. All experiments in this paper are based on the RGB triple channel, and tests are performed by converting the image colour space from RGB to YCbCr, in its Y channel.

4.2. Experimental results and analysis

In order to objectively evaluate the algorithm proposed in this paper, We selected Bicubic, SRCNN[5], FSRCNN[6], VDSR[17], DRCN[18], CARN[19], MSRN[20] and other algorithms for comparison. The results are shown in Table 1. At 2x, 3x and 4x magnification, the PSNR values and SSIM values of the above algorithms are compared on the test data sets Set5, Set14, BSD100 and Urban100, respectively.

Table 1: Quantitative evaluation of 9 SR methods tested on four benchmark sets at different magnifications

Method	Scale	Set5	Set14	BSD100	Urban100
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic	2	33.66/0.9299	30.24/0.8688	29.56/0.8431	26.88/0.8403
SRCNN		36.66/0.9542	32.42/0.9063	31.36/0.8879	29.50/0.8946
FSRCNN		37.00/0.9559	32.75/0.9098	31.51/0.8939	29.87/0.9065
VDSR		37.53/0.9587	33.03/0.9124	31.90/0.8960	30.77/0.9140
DRCN		37.63/0.9588	33.08/0.9118	31.08/0.8942	30.41/0.9133
MSRN		38.08/0.9605	33.74/0.9170	32.23/0.9013	32.22/0.9326
CARN		37.76/0.9590	33.52/0.9166	32.09/0.8978	31.92/0.9256
RCCN		38.13/0.9610	33.67/0.9185	32.22/0.9001	32.41/0.9302
Bicubic	3	30.39/0.8682	27.55/0.7742	27.21/0.7385	24.46/0.7349
SRCNN		32.75/0.9090	29.28/0.8209	28.41/0.7863	26.24/0.7989
FSRCNN		33.02/0.9135	29.49/0.8271	28.50/0.7937	26.41/0.8161
VDSR		33.66/0.9213	29.78/0.8314	28.82/0.7976	27.14/0.8279
DRCN		33.82/0.9226	29.79/0.8311	28.82/0.7963	27.07/0.8276
MSRN		34.38/0.9262	30.34/0.8395	29.08/0.8041	28.08/0.8554
CARN		34.29/0.9255	30.29/0.8407	29.06/0.8034	28.06/0.8493
RCCN		34.48/0.9278	30.37/0.8417	29.12/0.8054	28.34/0.8549
Bicubic	4	28.42/0.8104	26.00/0.7027	25.96/0.6675	23.14/0.6577
SRCNN		30.48/0.8628	27.49/0.7503	26.90/0.7101	24.52/0.7221
FSRCNN		30.66/0.8646	27.71/0.7562	26.98/0.7124	24.60/0.7221
VDSR		31.35/0.8838	28.02/0.7674	27.29/0.7251	25.18/0.7524
DRCN		31.35/0.8854	28.19/0.7670	27.23/0.7233	25.14/0.7510
MSRN		32.07/0.8903	28.60/0.7751	27.52/0.7273	26.04/0.7896
CARN		32.13/0.8937	28.60/0.7806	27.58/0.7349	26.07/0.7837
RCCN		32.24/0.8966	28.69/0.7833	27.63/0.7376	26.31/0.7912

As can be seen from Table 1, compared with the above comparison algorithms, the proposed algorithm has significantly improved PSNR and SSIM at 2x, 3x and 4x magnification, and is more obvious at 3x and 4X magnification. Compared with the CARN algorithm in Set5, Set14, BSD100 and Urban100 test sets, the average PSNR value of the proposed algorithm is improved to 0.19dB, 0.08dB, 0.06dB and 0.28dB, respectively, when compared with the second-best one. The enhancement of SSIM value is 0.0023, 0.0010, 0.0020 and 0.0056 respectively. It indicates that RCCN can show better performance in each data set.

As shown in Figure 4, in the upper outer column of Urban image024. The reconstruction effect of RCCN algorithm proposed in this paper is clearer, and compared with other algorithms, there is no artifact phenomenon, and it is closer to HR image. In Urban image076, the transverse decoration of the window light has some distortion and deformation of other algorithms, while the reconstruction

effect of RCCN algorithm has no deformation, which is a good restoration of the exterior light decoration, and the recovery effect of details is more accurate than other algorithms.

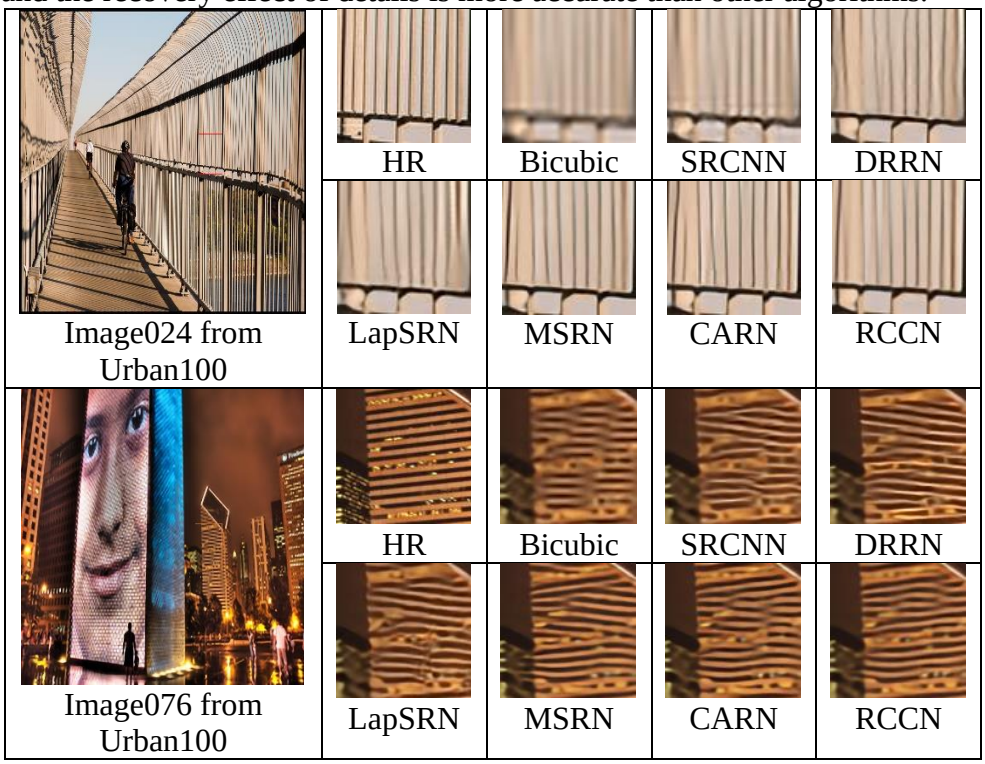

Figure 4: Visual quality of the Urban100 test set 4x magnification of RCCN compared to advanced methods

5. Conclusion

In order to solve the problem of insufficient extraction of residual information and insufficient use of feature information in residual network, a new super-resolution reconstruction algorithm based on residual compensation combined attention network is proposed in this paper. We designed a three-channel residual feature extraction module to extract more feature information. The main channel adopted the low-latitude residual extraction method to extract information while reducing the number of parameters. The two auxiliary channels adopted different convolution connections to extract features of different levels. At the same time, in order to make the features extracted from the residual compensation pass effectively and improve the extraction efficiency, the combined attention method is used to further select the extracted residual information, increasing the efficiency of the network, allowing better passage of low frequency information and enriching the edge information of the network.

Acknowledgements

The author would like to thank the referees and an editor for providing useful comments.

References

[1] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," IEEE transactions on image processing, vol. 19, no. 9, pp. 2345-2356, 2010.
[2] F. Zhou, W. Yang, and Q. Liao, "Interpolation-based image super-resolution using multisurface fitting," IEEE Transactions on Image Processing, vol. 21, no. 7, pp. 3312-3318, 2012.

[3] R. Tsai, "Multiple frame image restoration and registration," Advances in Computer Vision and Image Processing, vol. 1, pp. 1715-1989, 1989.
[4] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," International journal of computer vision, vol. 40, pp. 25-47, 2000.
[5] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 2, pp. 295-307, 2015.
[6] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14, 2016, pp. 391-407: Springer.
[7] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017, pp. 136-144.
[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.
[9] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132-7141.
[10] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 286-301.
[11] H. Zhao, X. Kong, J. He, Y. Qiao, and C. Dong, "Efficient image super-resolution using pixel attention," in Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16, 2020, pp. 56-72: Springer.
[12] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017, pp. 126-135.
[13] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," 2012.
[14] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7, 2012, pp. 711-730: Springer.
[15] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, 2001, vol. 2, pp. 416-423: IEEE.
[16] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 5197-5206.
[17] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1646-1654.
[18] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1637-1645.
[19] N. Ahn, B. Kang, and K.-A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 252-268.
[20] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 517-532.