



# Single Image Super-Resolution by Parallel CNN with Skip Connections and ResNet

Qimei Wang and Feng Qi<sup>(✉)</sup>

Business School, Shandong Normal University, Jinan, China  
474458464@qq.com, cliff@sdnu.edu.cn

**Abstract.** Image super-resolution has always been a research hotspot in the field of computer vision. Recently, image super-resolution algorithms using convolutional neural networks (CNN) have achieved good performance. But the existing methods based on CNN usually has to many (20–30) convolution layers, which has a large amount of calculations. In response to this problem, this paper proposes a lightweight network model based on parallel convolution, skip connections and ResNet. Parallel convolution means that different sizes of convolution kernels are set in the same convolutional layer to extract image features of different scales. In addition, in order to reduce the loss of image details, we combine the input and output of the previous layer as the input of the next layer, which is the skip connection. We also borrowed the idea of Residual Net. The network learns the residuals between high-resolution images and low-resolution images. Therefore, the algorithm proposed in this paper not only achieves the most advanced performance, but also achieves faster calculations.

**Keywords:** Image super-resolution · Parallel convolution · Skip connection · Residual net

## 1 Introduction

Due to the limitation of the equipment in real life, sometimes the image we obtain is low resolution and cannot meet our needs. In order to solve this problem, we can use Image Super-Resolution technology.

Image Super-Resolution refers to the recovery of high resolution images from a lowresolution image or image sequence. Image Super-Resolution is divided into Single Image Super-Resolution (SISR) and Multiple Image Super-Resolution (MISR). In this paper, we mainly focus on Single Image Super-Resolution (SISR). At present, SISR mainly includes three methods: interpolation-based method [1–4], reconstruction-based method [5–8], and learning-based [9–13] method. In recent years, with the wide application of artificial neural networks and the development of deep learning, the idea of deep learning has been introduced into the field of image super-resolution and many classical methods have been proposed. The current classic algorithm contains SRCNN [9], FSRCNN [10], VDSR [11], DRCN [12], SRDenseNet [13], RED [14].

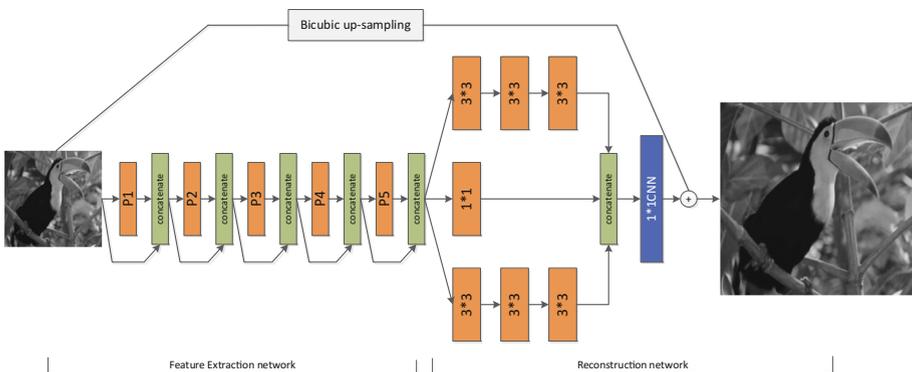
SRCNN is proposed by Dong [9] et al. SRCNN consists of three convolutional layer, which are image feature extraction layer, nonlinear mapping layer and reconstruction layer. Then, Dong [10] and others proposed the FSRCNN algorithm.

FSRCNN is an improved algorithm for SRCNN. It eliminates the steps of bicubic interpolation and adds a deconvolution layer to the final layer to amplify the image. Kim [11] and others proposed the VDSR algorithm, drawing on the idea of the ResNet to speed up the network convergence. In addition, Kim [12] and others also proposed the DRCN algorithm. The DRCN algorithm applies the existing recurrent neural network to the image super-resolution field for the first time. Tong [13] et al. proposed the SRDenseNet algorithm, and SRDenseNet links SISR and dense connections for the first time.

In this paper, we propose a new Image Super-Resolution model called CPCSCR in which ideas of skip connection [13], parallel convolution [15] and Residual Net [16] were employed. This network mainly includes four advantages. First, this algorithm directly processes the original image, eliminating the pre-processing steps and retaining the details of the image. Second, parallel convolution is used to extract image features of different scales. Third, skip connection enables the network to more fully learn the image features. Fourth, the network learns the residuals between high-resolution images and low-resolution images, achieves more efficient calculations. Therefore, this model is lighter, less computational, and better performing than the models we mentioned above.

## 2 Method

As shown in Fig. 1, CPCSCR is a fully convolutional neural network with feature extraction network and reconstruction network. We use the original image as input, extract features of different scales of the image through the parallel convolution module, and then import the extracted features into the reconstruction layer to reconstruct the image details. In addition, our model learns the residuals between low-resolution images and high-resolution images.



**Fig. 1.** CPCSCR network overall architecture

## 2.1 Feature Extraction Network

As shown as Fig. 2, feature extraction layers consists of five parallel convolutional modules, each modules contains  $1 \times 1$  [17] and  $3 \times 3$  kernels, bias and Parametric ReLU. Five modules have the same structure. The input and output of each module are spliced and used as input for the next module.

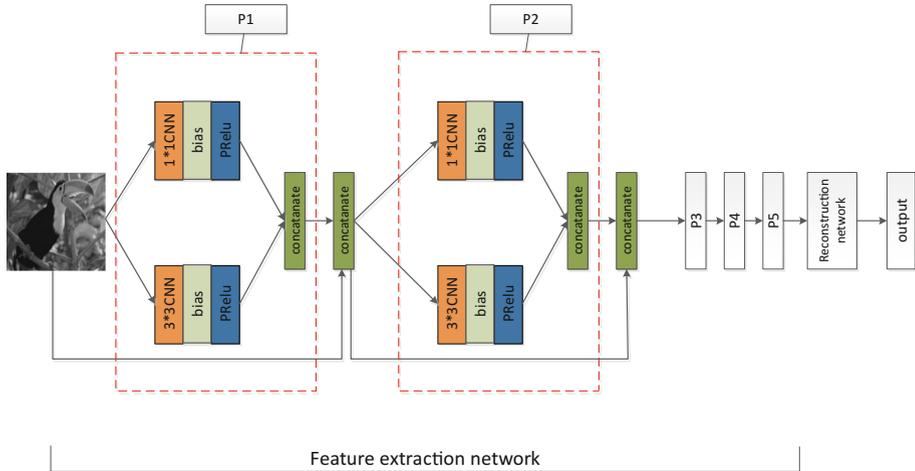


Fig. 2. Construction of parallel convolution modules

In the case of feature extraction, the parallel convolutional proposed by Google [15] is typically used. The parallel convolutional means process an image with multiple different convolution kernels simultaneously and splice different feature maps together. In CPCSCR, we use  $1 \times 1$  and  $3 \times 3$  convolutional kernels. The role of the  $1 \times 1$  convolution kernel is to control the number of feature maps so that it can be easily concatenated with the feature map generated by the  $3 \times 3$  convolution kernel. Because the original image is the input of the network, the size of the original image is smaller than the image after preprocessing, so a large-sized convolution kernel is not required [10]. A  $3 \times 3$  convolution kernel is sufficient to cover the entire image information.

In addition, we have found that the network has a good performance when the network contains 5 modules. Since then, when the number of modules has increased again, the performance of the network has not been greatly improved.

## 2.2 Reconstruction Network

As shown as Fig. 1, the reconstruction network consists of three parallel branches, the first branch is three serial  $3 \times 3$  convolution kernels, the second branch is a  $1 \times 1$  [16] convolution kernel, and the third branch is the same as the first branch. As the name suggests, the reconstruction layer up-sampling the image to complement the image details.

In previous models, deconvolution (also known as transposed convolution) was mostly used in the reconstruction layer to up-sampled the image. The process of transposed convolutional layer is similar to the usual convolutional layer and the reconstruction ability is limited. This means that the deeper the deconvolution layer, the better the reconstruction performance, but it also means that the computational burden is increased. So we propose a parallelized CNN structure, which usually consists of  $1*1$  [16] and  $3*3$  convolutional kernels.

As stated in the model, because there are a lot of connection operations in the feature extraction layer, the input data dimension of the reconstruction layer is very large. So we use  $1*1$  [16] CNNs to reduce the input dimension before generating the HR pixels.

The last CNN, represented by the dark blue color in Fig. 1, compensating for the dimensional reduction caused by the parallel convolution structure.

## 3 Experiment

### 3.1 Experiment Setup

Experimental training datasets are Yang 91 [18] and BSDS200 [19]. Then we expand the training-sets by performing data augmentation operations on each image. The specific operations are that each image is vertically flipped, flipped horizontally, flipped horizontally and vertically. The total number of training images is 1,164 and the total size is 259 MB. In order to compare with existing image super-resolution algorithms, this paper converts color (RGB) images to YCbCr images and only processes Y-channels (Y-channel represents brightness). Each training image is divided into 32 steps, 32 steps, 16 steps, and 64 patches are used as the minimum batch. We use BSDS100 [19], SET 5 [20] and SET 14 [21] as test datasets.

The experiment uses an internationally accepted evaluation criterion to measure the network performance: Peak Signal-to-Noise Ratio (often abbreviated as PSNR). The unit of PSNR is dB. The larger the value is, the smaller the picture distortion is, and the better the network performance is.

We Initialize each convolution kernel using the method of He et al. All biases and PReLU in the network are initialized to zero. The dropout rate of  $p = 0.8$  during training. The Mean Squared Error function is used as a loss function to calculate the difference between the network output value and the true value. In addition, we used Adam [22] with an initial learning rate = 0.002 to optimize the algorithm to minimize the loss. If the loss value does not decrease after 5 training steps, the learning rate drops by 2 times. Training ends when the learning rate is less than 0.00002. An example of the results is shown in Fig. 3.

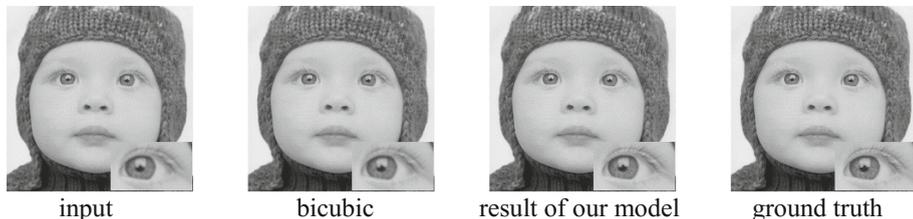


Fig. 3. An example of our results of img\_001 in Set5

### 3.2 Comparisons with State-of-the-Art Methods

**Comparisons of PSNR:** We use PSNR to objectively evaluate the processing result of the algorithm. Table 1 show the results of objective tests at scale = x2, scale = x3 and scale = x4 respectively. Except that the objective index of RED30 algorithm for BSD100 data set is slightly higher than our algorithm, the other test results show that the PSNR value obtained by our algorithm is higher than other algorithms, which fully demonstrates that the algorithm has better processing effect. Although the performance of our algorithm on the BSD100 dataset is slightly lower than RED30, it greatly reduces the computational complexity.

Table 1. Comparisons of PSNR with other SR Method. (scale = x2)

Dataset	SRCNN	DRCN	VDSR	RED30	CPCSC(ours)
Set5(x2)	36.66	37.63	37.53	37.66	37.75
(x3)	32.75	33.82	33.66	33.82	34.04
(x4)	30.49	31.53	31.35	31.51	31.70
Set14(x2)	32.45	33.04	33.03	32.94	33.18
(x3)	29.28	29.76	29.77	29.61	29.90
(x4)	27.49	28.02	28.01	27.86	28.24
BSD100(x2)	31.36	31.85	31.90	31.99	32.00
(x3)	28.41	28.80	28.82	28.93	28.86
(x4)	26.90	27.23	27.29	27.40	27.35

**Comparison of computational complexity:** Since each implementation is performed on a different hardware device or platform, it is unfair to compare test execution times. Here we calculate the computational complexity of each method. The approximate computational complexity of each method is shown in Table 2. Therefore, we can see that our DCSCN has the most advanced reconstruction performance, and the computational complexity is much smaller than VDSR [11], DRCN [12] and RED30 [14].

**Table 2.** Comparisons of approximate computation complexity with other SR Method. (scale = x2) For comparison, we chose f1, f2, f3, n1, n2 = (9,5,5,64,32) for SRCNN

	SRCNN (9,5,5)	DRCN	VDSR	RED30	CPCSC
CNN layers	3	20	20	30	13
CNN filters	32,64	256	64	64	32 to 96
Complexity[k]	229.5	78,083.2	2,668.5	4,152.8	80.9

## 4 Conclusion

This paper presents an image super-resolution method based on convolutional neural networks with parallel convolution, skip connection and ResNet. The algorithm uses parallel convolution to extract features of different scales of the image, and inputs the local and global features of each layer to the next layer by means of skip connections. In addition, the algorithm learns the residual between the low resolution image and the high resolution image. Another important point is, the model takes an image of the original size as an input, reducing image information loss. Using these methods, our model can achieve the most advanced performance with less computing resources. The experimental results also show that the model has better performance.

**Acknowledgements.** This work was supported by the Natural Science Foundation of China (No. 61502283). Natural Science Foundation of China (No. 61472231). Natural Science Foundation of China (No. 61640201).

## References

1. Tsai, R.Y.: Multiframe image restoration and registration. *Adv. Comput. Vis. Image Process.* **1**, 317–339 (1984)
2. Ur, H., Gross, D.: Improved resolution from subpixel shifted pictures. *CVGIP: Graph. Models Image Process.* **54**(2), 181–186 (1992)
3. Komatsu, T., et al.: Signal-processing based method for acquiring very high resolution images with multiple cameras and its theoretical analysis. *IEE Proc. I (Commun. Speech Vis.)* **140**(1), 19–25 (1993)
4. Komatsu, T., et al.: Very high resolution imaging scheme with multiple different-aperture cameras. *Sign. Proces. Image Commun.* **5**(5-6), 511–526 (1993)
5. Tappen, M.F., Russell, B.C., Freeman, W.T.: Exploiting the sparse derivative prior for super-resolution and image demosaicing. In: *IEEE Workshop on Statistical and Computational Theories of Vision* (2003)
6. Kim, K.I., Kwon, Y.: Single-image super-resolution using sparse regression and natural image prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 1127–1133 (2010)
7. Dai, S., et al.: Soft edge smoothness prior for alpha channel super resolution. In: *IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2007. IEEE* (2007)
8. Yang, X., et al.: An improved iterative back projection algorithm based on ringing artifacts suppression. *Neurocomputing* **162**, 171–179 (2015)
9. Dong, C., et al.: Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(2), 295–307 (2016)

10. Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 391–407. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46475-6\\_25](https://doi.org/10.1007/978-3-319-46475-6_25)
11. Kim, J., Lee, J.K., Lee, K.M.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)
12. Kim, J., Lee, J.K., Lee, K.M.: Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)
13. Tong, T., et al.: Image super-resolution using dense skip connections. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE (2017)
14. Mao, X.-J., Shen, C., Yang, Y.-B.: Image restoration using convolutional auto-encoders with symmetric skip connections. arXiv preprint [arXiv:1606.08921](https://arxiv.org/abs/1606.08921) (2016)
15. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015)
16. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)
17. Lin, M., Chen, Q., Yan, S.: Network in network. arXiv preprint [arXiv:1312.4400](https://arxiv.org/abs/1312.4400) (2013)
18. Yang, J., et al.: Image super-resolution via sparse representation. *IEEE Trans. Image Process.* **19**(11), 2861–2873 (2010)
19. Arbelaez, P., et al.: Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(5), 898–916 (2011)
20. Bevilacqua, M., et al.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding, p. 135-1 (2012)
21. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Boissonnat, J.-D., et al. (eds.) *Curves and Surfaces 2010*. LNCS, vol. 6920, pp. 711–730. Springer, Heidelberg (2012). [https://doi.org/10.1007/978-3-642-27413-8\\_47](https://doi.org/10.1007/978-3-642-27413-8_47)
22. Kinga, D., Ba Adam, J.: A method for stochastic optimization. In: *International Conference on Learning Representations (ICLR)*, vol. 5 (2015)