# Image Super-Resolution Based on Multi-scale Fusion Network

Leping Lin[1,2], Huiling Huang[2], and Ning Ouyang[1,2](✉)

[1] Key Laboratory of Cognitive Radio and Information Processing (Guilin University of Electronic Technology), Ministry of Education, Guilin 541004, Guangxi, China
lin_leping@163.com, ouyangning@guet.edu.cn

[2] School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, Guangxi, China
834597430@qq.com

**Abstract.** It is important and necessary to obtain high-frequency information and texture details in the image reconstruction applications, such as image super-resolution. Hence, it is proposed the multi-scale fusion network (MCFN) in this paper. In the network, three pathways are designed for different receptive fields and scales, which are expected to obtain more texture details. Meanwhile, the local and global residual learning strategies are employed to prevent overfitting and to improve reconstruction quality. Compared with the classic convolutional neural network-based algorithms, the proposed method achieves better numerical and visual effects.

**Keywords:** Multi-pathways · Residual learning · Multi-scale · Receptive field · Texture details

## 1 Introduction

Deep learning (DL)-based methods are well applied in super-resolution reconstruction [1, 2]. The advantage of the methods is the new nonlinear mapping learning idea. As the pioneer network model for SR, super-resolution convolutional neural network (SRCNN) [3] established the nonlinearlow resolution–high resolution (LR–HR) mapping via a fully convolutional network, and significantly outperformed the classical non-DL methods. However, the method loosed many textures, and converged slowly. Later, the accelerated version of SRCNN, namely FSRCNN [4], is proposed, where the de-convolution layer instead of the bicubic interpolation is made used to improve the reconstruction quality. Based on it, Shi proposed a more efficient sub-pixel convolution layer [5] to fulfill the upscaling operation, which is performed at the very end of the network.

Though the available models based on deep neural networks have achieved success both in reconstruction accuracy and computational performance, the high-frequency information and texture details of images are ignored. In this paper, it is proposed the image super-resolution method based on the multi-scale fusion network (MCFN), which enlarged the reception field [13] of the network by multiple pathways and

© Springer Nature Singapore Pte Ltd. 2020
Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 516,
https://doi.org/10.1007/978-981-13-6504-1_9

obtained more texture details. The network is composed of multiple pathways and trained by the residual strategies. The advantages of this network are twofolds:

- Three pathways that employed different sizes of convolution layers are designed to get different size of receptive fields, which could extract multi-scale features of the input image. This network design is expected to preserve more image details after the extracted features are fused for the reconstruction.
- The residual strategy is composed of two types: local and global residual learning. The former, which linked the input and the convergence end of the pathways, aimed at optimizing this network. The local residual learning, which linked the input and the end of the network, aimed at preventing overfitting. The use of residual learning could further improve the reconstruction quality.

The method will be validated based on the publicly available benchmark datasets, and compared with the available DL-based methods. By the experiments, it is shown that the proposed method achieves better results.

## 2 Multi-scale Fusion Network

The proposed multi-scale fusion network is composed of three parts: the block extraction and representation, the nonlinear mapping, and the reconstruction part. As shown in Fig. 1, the nonlinear mapping composed three pathways, which could obtain more high-frequency information and texture details than single pathway. The residual learning has two patterns: local residual learning and global residual learning, which could prevent overfitting training, and further improve the quality of reconstruction.

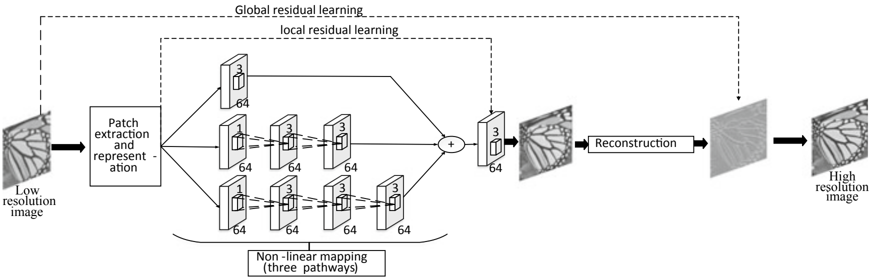


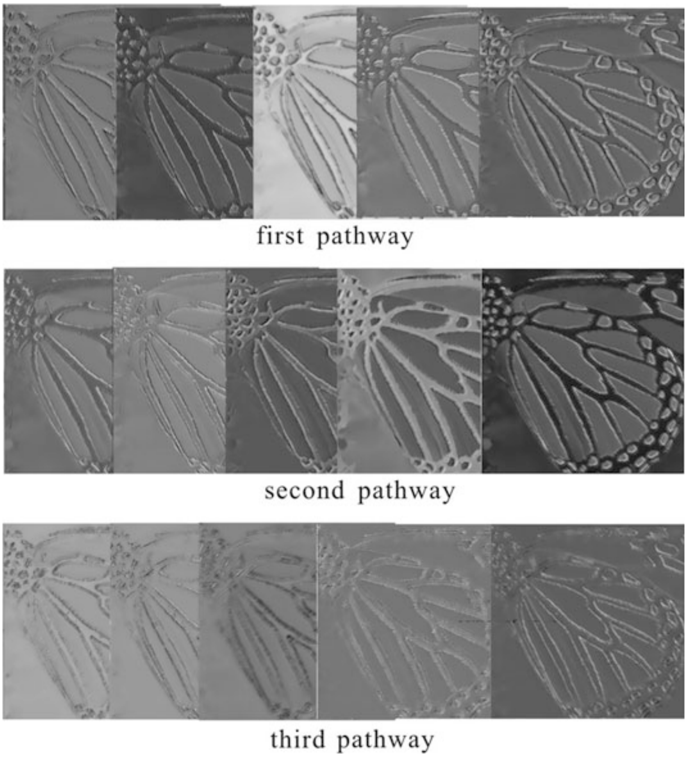
**Fig. 1.** Framework of the proposed MCFN method

### 2.1 Multiple Pathways Design

The three pathways in the nonlinear mapping part are designed to obtain feature details of diverse scales, which are followed by the fusion operation for reconstruction. The three pathways perform 1, 2, and 3 cascaded 3 × 3 convolution operations, respectively. Besides, a 1 × 1 filter is put on the beginning of the second and third pathway [6] to reduce the amount of parameters. The design is motivated by the GoogLeNet [7], where multiple paths are adopted to obtain various receptive fields. Besides, based on

GoogLeNet, Kim [8] proposed to remove the pooling operations, and to replace the small convolutional kernels by the big ones, which achieved good performance in image restoration. The paralleled pathways could easily be applied the parallel computations.

Figure 2 shows the features abstracted by the three pathways. It can be seen that different receptive fields and features of different scales are obtained from different paths. The first pathway learns a smooth feature, the second pathway learns texture details, and the third pathway learns contour information.



first pathway

second pathway

third pathway

**Fig. 2.** Different texture details abstracted by three pathways, respectively

The output formulas of these pathways as follows:

$$F_1 = \sigma(w_1 x) \tag{1}$$

$$F_2 = \sigma(w_4 \sigma(w_3 \sigma(w_2 x))) \tag{2}$$

$$F_3 = \sigma(w_8 \sigma(w_7 \sigma(w_6 \sigma(w_5 x)))) \tag{3}$$

where $w_1$, $w_2$, $w_3$, $w_4$, $w_5$, $w_6$, $w_7$, $w_8$ represent the filters of the convolution layers, and $\sigma$ denotes the rule operation. $F_1$, $F_2$, and $F_3$ are the output of the first, second, and third pathway, respectively.

## 2.2 Residual Learning

In order to optimize the training of the nonlinear mapping part, and to improve the convergence speed, a residual connection is designed between the input and output of the pathway. Since the LR image and the HR image compose the same structures, and share the same low-frequency information, it would be advantageous to model the connection [9–11]. Similarly, a residual connection is made between the input and the output of the network. This connection has proven to be an effective network structure.

The residual learning structures are shown in Fig. 1. The two parts adopt the structure of residual learning, which is achieved through the forward neural network and the residual connection. The connection between the convolutional layers is a sequential connection using batch normalization (BN) and rectified linear units (Relu). This is a forward-activated structure proposed by He [12], which is easy to train and easier to fit. At the same time, the residual connection is equivalent to simply performing the equivalent mapping, without generating additional parameters or increasing computational complexity to prevent overfitting of the network.

In the nonlinear mapping stage, it is achieved through three pathways. In this part of the input and output residual connection, this connection structure is different from Resnet, as shown in the following formula:

$$y_l = h(x_l) + F(x_l, w_l) \tag{4}$$

$$x_{l+1} = f(y_l) \tag{5}$$

where $x_l$ and $x_{l+1}$ are the input and output of the first layer, respectively. $F$ is a residual function. $h(x_l) = x_l$ is an identity mapping and $f$ represents the Relu. The activation function of this connection structure $x_{l+1} = f(y_l)$ has an impact on two aspects, as shown in the following equation:

$$y_{l+1} = f(y_l) + F(f(y_l), w_{l+1}) \tag{6}$$

This is an asymmetrical approach that allows the activation function $\hat{f}$ to affect $l$ arbitrarily. Only the $F$ path can be affected, given by:

$$y_{l+1} = y_l + F(\hat{f}(y_l), w_{l+1}) \tag{7}$$

$F(\hat{f}(y_l), w_{l+1})$ is the sum of the nonlinear mappings of the three pathways, as shown as following:

$$F(\hat{f}(y_l), w_{l+1}) = f\{F_1 + F_2 + F_3) \tag{8}$$

The first row on the right side of the equation represents the output of the first pathway. The second and third rows of the equation represent the second pathway and the third pathway, respectively.

To optimize and estimate the network model, the mean squared error (MSE) is used as a loss function to estimate the network parameters, given by:

$$L(\theta) = \frac{1}{2N}\sum_{l=1}^{N}\left\|\tilde{x}^{(l)} - D(x^{(l)})\right\|^2 \tag{9}$$

$\{x^{(l)}, \tilde{x}^{(l)}\}_{i=1}^{N}$ is the training set, in which $N$ is the number of training fast, and $\tilde{x}^{(l)}$, $x^{(l)}$ are the ground truth value corresponding to the LR block. $\theta$ is a set of parameters to evaluate.

## 3 Experiments and Analysis

In order to make a comparison with the existing methods, a widely used training set of 91 images was used, and a magnification of 3 was used to train the data. A random gradient tailored training mode was used. The algorithms and comparison algorithms are on the same experimental platform (Intel CPU 3.30 GHZ and 8G memory), and MATLAB R2014a and Caffe are applied. The four common datasets are evaluated separately: Set5, Set14, BSD100, and Urban100. Quantitative assessment is used by

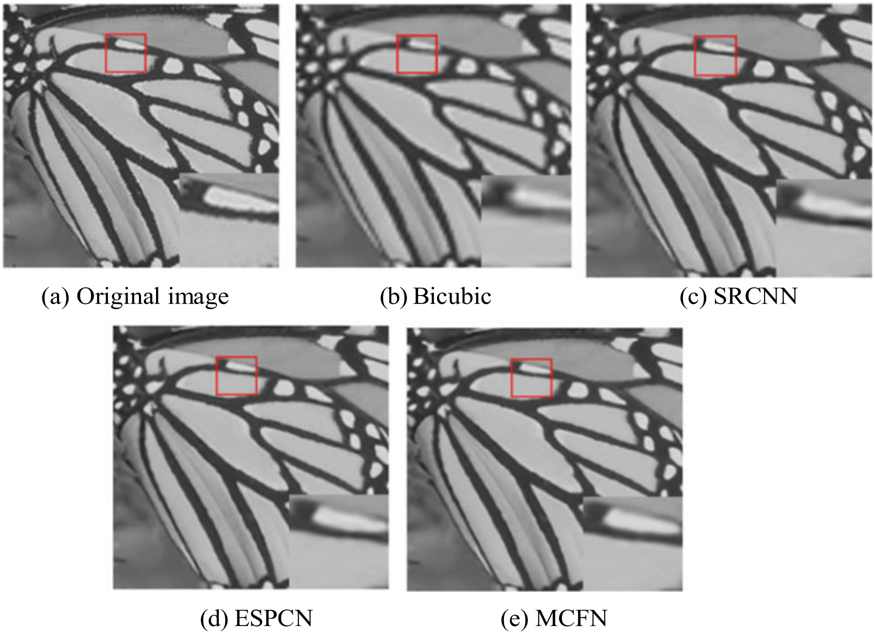

(a) Original image
(b) Bicubic
(c) SRCNN
(d) ESPCN
(e) MCFN

**Fig. 3.** Comparison of reconstruction results of original HR and each pathway of butterfly

PSNR and SSIM, because humans are more sensitive to changes in brightness than colors. Following most existing methods, super-resolution reconstruction of luminance channels is performed only in the YCbCr color space in order to display patterns, and the other two channels are simply performed by bicubic interpolation.

As shown in Fig. 3, compared with classic networks, the reconstruction image of MCFN is more clear, which have more detailed textures which improves the reconstruction effect.

The comparison of the average numerical results is shown in Table 1. Through the comparative experimental analysis, it can be seen that MSFN has higher PSMR and SSIM, intuitively indicating that the information extracted by the MSFN is conducive to reconstruction.

**Table 1.** PSNR (dB) comparison of test image reconstruction results

| Dataset | Enlargement factor | Bicubic | SRCNN | ESPCN | MSFN |
|---|---|---|---|---|---|
| Set5 | 3 | 30.39 | 32.39 | 32.55 | 33.19 |
| Set14 | 3 | 27.54 | 29.00 | 29.08 | 29.48 |
| Urban100 | 3 | 27.05 | 27.94 | 28.12 | 28.40 |
| BSD100 | 3 | 27.57 | 28.45 | 28.65 | 28.96 |

## 4 Conclusions

This paper proposes a super-resolution reconstruction method based on multi-scale fusion network, which is composed multiple pathways and adopted residual learning strategies. The multiple pathways could capture the texture information of different scales, which increases the diversity of features, and improves the accuracy of reconstruction. Using the residual learning strategies can not only enhance the ability of the network model to fit features, but also optimize network model. The residual links at the input and convergence ends of the three pathways, as well as at the input and output ends of the network. By the experiment results, the method is shown effective.

**Acknowledgments.** This work is partially supported by the following foundations: the National Natural Science Foundation of China (61661017); the China Postdoctoral Science Fund Project (2016M602923XB); the Natural Science Foundation of Guangxi province (2017GXNSFBA198212, 2016GXNSFAA38014); the Key Laboratory Fund of Cognitive Radio and Information Processing (CRKL160104, CRKL150103, 2011KF11); Innovation Project of GUET Graduate Education (2016YJCXB02); the Scientific and Technological Innovation Ability and Condition Construction Plans of Guangxi (159802521); the Scientific and Technological Bureau of Guilin (20150103-6).

## References

1. Jiao LC, Liu F, et al. The neural network of seventy years: review and prospect. Chin J Comput. 2016;39(8):1697–716.
2. Jiao LC, Zhao J, et al. Research progress of sparse cognitive learning calculation and recognition. Chin J Comput. 2016;39(4):835–51.
3. Dong C, Loy CC, He K, et al. Image super-resolution using deep convolutional networks. IEEE Trans Pattern Anal Mach Intell. 2016;38(2):295–307.
4. Dong C, Loy CC, Tang X. Accelerating the super-resolution convolutional neural network. In: European conference on computer vision. Springer International Publishing; 2016. p. 391–407.
5. Shi W, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 1874–83.
6. Lin M, Chen Q, Yan S. Network in network [EB/OL]. (2013-1-31)[2018-3-25] (2013) https://arxiv.org/abs/1312.4400.
7. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 1–9.
8. Kim Y, Hwang I, Cho NIA. New convolutional network-in-network structure and its applications in skin detection, semantic segmentation, and artifact reduction [EB/OL]. (2017-1-22) [2018-3-25] https://arxiv.org/abs/1701.06190.
9. Huang J, Yu ZL, Cai Z, et al. Extreme learning machine with multi-scale local receptive fields for texture classification. Multidimension Syst Sig Process. 2017;28(3):995–1011.
10. Kim J, Kwon Lee J, Mu Lee K. Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 1637–45.
11. Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 1646–54.
12. He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks. In: European conference on computer vision. Springer International Publishing; 2016. p. 630–45.
13. Luo W, Li Y, Urtasun R, et al. Understanding the effective receptive field in deep convolutional neural networks. In: Advances in neural information processing systems; 2016. p. 4898–906.