Multiple Residual Learning Network for Single Image Super-Resolution

Renhe Liu, Sumei Li, Chunping Hou, Guoqing Lei
School of electrical and information engineering, Tianjin University, Tianjin, P.R. China
tjnklsm@163.com, hcp@tju.edu.cn

***Abstract*—Deep residual convolutional neural network (CNN) has recently achieved great success in image super-resolution (SR). Because residual learning accelerates convergence rate and eases the difficulty for reconstructing high-resolution (HR) image, these CNN models can achieve higher peak signal to noise ratio (PSNR) values with lower training cost. However, residual image used in present residual network still contains much high frequency information, which increases learning burden and limits learning ability of residual network. Moreover, training a very deep network faces many obstacles and costs too much time. In this paper, we propose a multiple residual learning network (MRLN), which not only further simplifies information complexity of residual image and improves the accuracy of residual network, but also obviously reduces time cost for training a very deep CNN. In MRLN, we use a shallow network formed by 30-layer convolutional layers as basic model and train it for multiple times. The output of previous basic model is used as the HR input of the next one. In this way, an extremely large CNN is converted into a series connection of shallow networks. Fig. 1 shows PSNR of recent state-of-the-art CNN models for scale factor 2 on Set5, our method performs better than other methods and set a new level for SR.**

***Index Terms*—super-resolution, multiple residual learning, convolutional neural network, shallow, series connection**

I. Introduction

SR is a classical problem in computer vision, which aims to generate HR image from given LR image and can be applied in various image processing tasks, such as medical imaging [1], satellite imaging, security and surveillance [2]. For SR, the most significant work is that detail information in truth HR image, which is necessary and important for many applications, can be best recovered.

A. Related Previous Work

Since super-resolution problem has been studied for decades, many useful methods or theories have been proposed to solve it. Recently, CNN based on deep learning has demonstrated superiority over other neighbour embedding methods [3] or sparse coding methods [4] and gradually become a powerful method for SR. Some researchers have done good jobs in this area.

SRCNN [5] was firstly proposed to demonstrate feasibility for applying deep learning into super-resolution, but the convergence rate of it was slow and the depth of it was too shallow to get enough receptive field. In VDSR [6], Kim et al. increased the number of convolutional layers to amplify receptive field and introduced global residual learning to accelerate training process. Tai et al. compared residual architecture in VDSR [6], ResNet [7] and DRCN [8], and then built a DRRN [9] model which combined local and global residual learning and increased the depth of network by recursive architecture. Lim et al. developed an enhanced network EDSR [10] based on residual block, which expands CNN model size and provide more feature information from different level convolutional layers for reconstructing HR image. Zhang et al. proposed residual dense block to extract abundant hierarchical features via dense connected convolutional layers, this residual dense network called as RDN [11] achieved better results than previous methods.

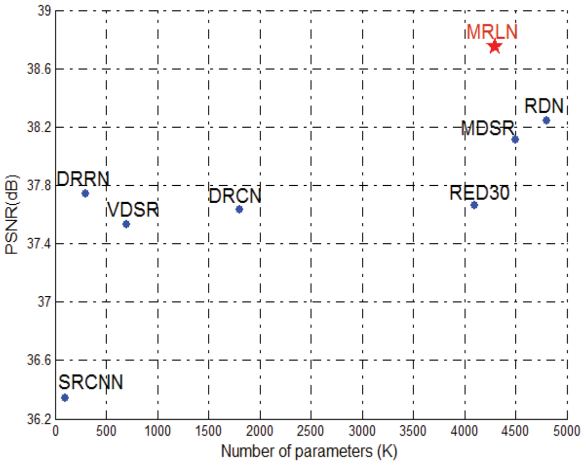


Fig. 1 PSNR of recent state-of-the-art CNN models for scale factor ×2 on Set5. MRLN is our method, in which 4 basic models are connected (the detail of basic model is shown in Fig. 3). With the depth of 120-layer (30×4), MRLN outperforms other very deep networks MDSR (160-layer) and RDN (128-layer) by 0.64 dB and 0.51 dB respectively.

B. Our Contributions

As is seen above, residual learning has been widely applied to favour higher PSNR values. By applying the strategy of residual learning, learning target of CNN is transferred from the ground truth HR image to residual image. However, in present CNN based methods, residual image comes from differences between bicubic interpolated LR image (which is generated by interpolating original LR image to HR size) and the ground truth HR image, it is not a proper

978-1-5386-4458-4/18/$31.00 ©2018 IEEE

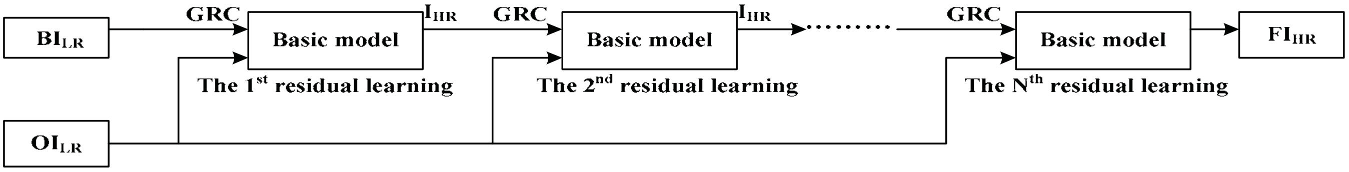

BI_LR
GRC
Basic model
I_HR
GRC
Basic model
I_HR
GRC
Basic model
FI_HR
OI_LR
The 1st residual learning
The 2nd residual learning
The Nth residual learning

Fig. 2 The framework of MRLN. OI_{LR} is the original LR size image down-sampled from the ground truth HR image. BI_{LR} is HR size image formed by bicubic interpolating OI_{LR} to HR size. I_{HR} is HR image reconstructed by the basic model. FI_{HR} is the final HR image reconstructed by MRLN. N is the number of basic models. GRC is global residual channel, which is applied for global residual learning. The detail of GRC and basic model is described in Section II *A*.

way to generate an enough simple residual image because interpolation for LR image loses some high frequency information and much detail information is still reserved in residual image. This type of residual image usually has high information complexity, which restricts the potential of residual learning and the further improvement for reconstruction results. To solve this issue, we propose a multiple residual learning network (MRLN), which gradually decreases the information complexity by generating and learning residual image for multiple times.

The framework of MRLN is shown in Fig. 2. We choose a 30-layer CNN as one basic model and built a MRLN by connecting N basic models. Every basic model is a shallow residual learning network and can learning a mapping from original LR image to HR image by itself. For the first basic model, residual image is from difference between interpolated LR image and the ground truth HR image, but in the next basic model, we use the output HR image of previous basic model as the input of global residual channel and generate residual image by subtracting it from the ground truth HR image. Compared to interpolated LR image, HR image output by a well-trained basic model has a higher similarity with the ground truth HR image, so there will be less high frequency information reserved in residual image and the information complexity of residual image is accordingly reduced. As a final result, the difficulty for learning residual image is gradually reduced and reconstruction performance is also gradually improved as the number of basic model increases.

Moreover, the work for training a MRLN is much easier and faster than training other very deep convolutional networks like EDSR [10] or RDN [11]. There are two reasons for it. Firstly, training task for MRLN is finished by respectively training several basic models. These models keep the same structure, so we can directly fine-tune the next model based on the previous one. Secondly, the information complexity of residual image is gradually reduced after multiple residual learning, lightening the learning burden of network, training time of MRLN is simultaneously reduced. Take a MRLN containing 4 basic models (N=4) for example, our MRLN achieves the best result shown in Fig. 1 and only takes about 9 hours on one GTX 1080 Ti GPU.

II. Proposed Method

In this section, we describe details of our work. Firstly, we analyse the structure of basic model in MRLN. Secondly, we introduce multiple residual learning and show crucial factors for superior performance and fast convergence in our model, and then we argue that the proposed method is an advanced solution to deal with super-resolution problem by presenting more details.

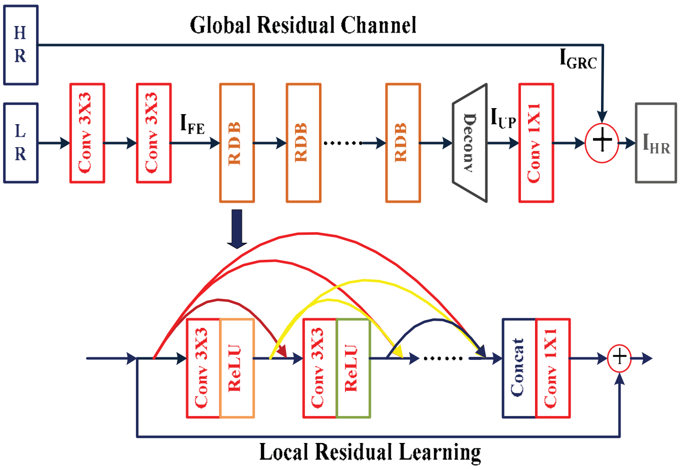


Fig. 3 The structure of basic model in MRLN. Similar to RDN [14], we also used residual dense block (RDB) proposed in it as main components of our basic model. There are 4 RDBs in our basic model. Every RDB contains 6 Conv layers

A. Basic Model for MRLN

Our basic model, outlined in Fig 3, is mainly inspired by VDSR [6], RDN [11] and ResNet [7], it includes two input channels and is different with above other CNN methods containing single input channel. One channel is global residual channel (GRC), the other channel can be divided into three parts: shallow feature extract net (SEN), residual learning net (RLN), and reconstruction part (RP). At last, we combine the outputs of the two channels to get the reconstructed HR. SEN contains two convolutional layers. Supposing I_{LR} is original LR image, I_{HR} is the HR image reconstructed by the basic model, I_{FE} is feature map extracted by SEN. The relationship between I_{LR} and I_{FE} can be represented as:

$$I_{FE} = f_{SE}\,(I_{LR}) \qquad (1)$$

Where f_{SE} denotes 2-layer convolution operation in SEN. I_{FE} equates to the input of RLN, which consists of 4 residual dense blocks (RDB) and 1 deconvolution layer. If we denote active function of the n-th residual dense block (RDB) as f_{Rn}, the number of residual dense blocks as n, deconvolution operation as f_{DE}, the output of RLN I_{UP} can be represented as:

$$I_{UP} = f_{DE}\,(f_{Rn}\,(f_{R(n-1)}\,(\ldots f_{R1}(I_{FE})\ldots))) \qquad (2)$$

I_{HR} is the final output of the basic model. And it can be expressed as follows:

$$I_{HR} = f_{RP}(I_{UP}) + I_{GRC} \qquad (3)$$

Where f_{RP} denotes 1-layer convolution operation in RP. Generally speaking , convolution operation in our paper is

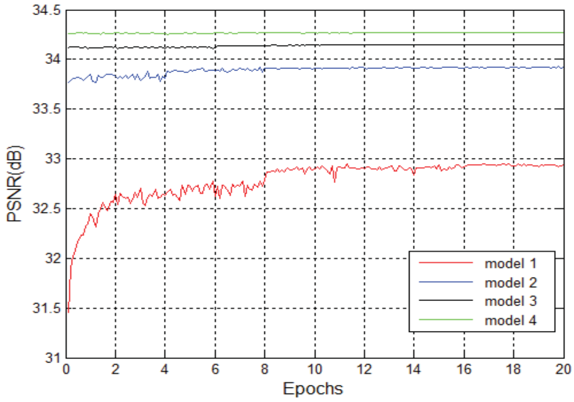


Fig. 4 The convergence curve of MRLN containing 4 basic models for scale ×2 on dataset14.

considered as composite function, such as ordinary convolution and rectified linear unit (ReLU) [12], so f_{SE}, f_{Rn}, f_{RP} is a kind of complicated process for input data.

B. Multiple Residual Learning

For multiple residual learning, the work of building a mapping between LR and HR image pairs is performed by multiple same basic models. Every basic model is called as shallow convolutional network. And every basic model is a simple residual learning network and can recovers some high frequency features of the ground truth image. The output HR image of previous shallow network is fed into global residual channel (GRC) of next basic model. And the similarity between it and the ground truth HR image accordingly increases, then the information complexity of residual image generated in next basic network is further reduced, making the difficulty for learning residual image is eased. As a result, convergence rate of MRLN is apparently accelerated and performance of image reconstruction is improved.

To better show the convergence tendency of different basic models, we built a MRLN formed by 4 basic models and show their convergence curve for scale ×2 on Set14 dataset in Fig. 4. We train all 4 models with the same epochs, but the convergence rate and final reconstruction performance change greatly among different models. Every model converges much faster and achieves higher PSNR values than the previous one, fully demonstrating the advantage of multiple residual learning. Besides, the promotion of PSNR values between the previous and present models is obvious, but this effect is weakened as the number of residual learning increases. This is because SR is a typical ill-posed problem, MRLN can gradually reduce the learning difficulty for target image, but the detail features which can be recovered by basic model become less and less during multiple residual learning, making that simplifying information complexity of residual image becomes more and more difficult.

In fact, the essence of deep learning based super-resolution is to achieve a mapping from LR size to HR size. Mapping accuracy is mainly decided by two factors: learning ability of network and complexity of target image. Before MRLN, many superior residual learning models have been introduced into the area of super-resolution, but all of them only contribute to improve architecture of network itself, neglect a fact that information complexity in residual image also decides the upper limit of residual learning. In contrast with them, in MRLN, basic model provides fundamental learning ability for building an end-to-end mapping, and at the same time, target residual image we aim to learn is progressively simplified by applying multiple residual learning.

C. Implementation Details

In proposed MRLN, almost all convolutional layers have 64 filters of size 3×3. Specially, the last convolutional layer in reconstruction part have only one filter of size 1×1. For convolutional layer with 3×3 kernel size, we pad zero to each side of the input image to keep size fixed.

For loss function, the mean squared error (MSE) is typically used to favour high PSNR in super resolution. We also use this criteria to minimize prediction loss of MRLN. The loss function can be represented as:

$$l(\theta)=\frac{1}{2N}\sum||y^{(i)}-F(x^{(i)})||^2 \tag{4}$$

Where θ denotes the parameter set, N is the number of training patches. $y^{(i)}$ is the ground truth HR patch of the LR image patch $x^{(i)}$. F denotes the mapping from input LR image to output HR image. In addition, considering both the training time and memory capacity of GPU, training images are split into 25×25 patches, and batch size is set as 32 in our method.

III EXPERIMENT RESULTS

A. Datasets

For training, we use a dataset of 291 images, where 91 images are from Yang et al. [13], and other 200 images are from Berkeley Segmentation Dataset [14]. For testing, we used three standard datasets Set5, Set14 and BSD100.

B. Settings

The method proposed in He et al. [15] is applied to initialize weights for our network. For training setting on every basic model, learning rate is initially set to 0.001, and then decreased by a factor of 10 every 8 epochs. If learning rate is less than 10^{-5}, the training is terminated.

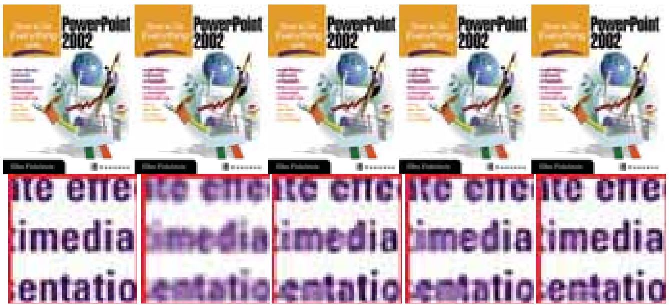
C. Comparisons with State-of-the-Art Methods

We now provide quantitative and qualitative comparisons. We compared our results with other state-of-the-art methods including A+ [4], SRCNN [5], DRRN [9], MemNet [16], MDSR [10], and RDN [11]. Because we crop image boundary for multiple times to make image size suitable for every basic model in MRLN, for fair comparison, similar to [6,9,17], we also crop pixels to the same amount. For evaluation measures, we mainly focus PSNR and structural similarity index (SSIM) on the Y channel of Ycbcr space, which are widely used in most super-resolution paper.

Table I provides a summary of quantitative results on the three standard testing datasets. We do these tests on scale factor ×2, ×3 respectively and MRLN we use to achieve report results in Table I contains 4 basic models (N=4). For PSNR, our method outperforms other methods in most tests.

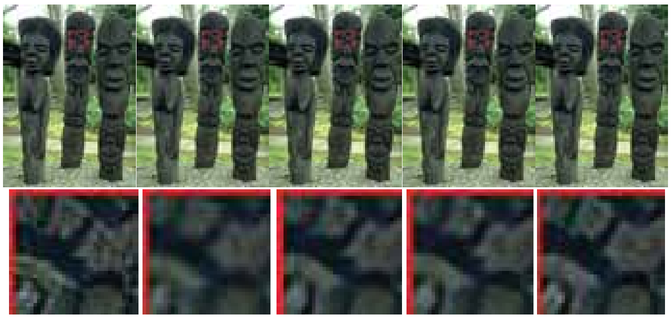
TABLE I
AVERAGE PSNR/SSIM for SCALE FACTOR ×2, ×3 on Set5, Set14, BSD100. RED COLOR INDICATES THE BEST PERFORMANCE AND BLUE COLOR INDICATES THE SECOND BEST PERFORMANCE.

Dataset	Scale	Bcubic	A+[4]	SRCNN[5]	DRRN[9]	MemNet[16]	MDSR[10]	RDN[11]	MRLN
Set5	×2	33.66/0.9299	36.54/0.9544	36.66/0.9542	37.74/0.9591	37.78/0.9597	38.11/0.9602	38.24/0.9614	38.80/0.9682
	×3	30.39/0.8682	32.58/0.9088	32.75/0.9090	34.03/0.9244	34.09/0.9248	34.66/0.9280	34.71/0.9296	34.72/0.9366
Set14	×2	30.24/0.8688	32.28/0.9056	32.45/0.9067	33.23/0.9136	32.28/0.9142	33.85/0.9198	34.01/0.9212	34.33/0.9386
	×3	27.55/0.7742	29.13/0.8188	29.30/0.8215	29.96/0.8349	30.00/0.8350	30.44/0.8452	30.57/0.8468	30.37/0.8660
BSD100	×2	29.56/0.8431	31.21/0.8863	31.36/0.8879	32.05/0.8973	32.08/0.8978	32.29/0.9007	32.34/0.9017	33.67/0.9319
	×3	27.21/0.7385	28.29/0.7835	28.41/0.7863	28.95/0.8004	28.96/0.8001	29.25/0.8091	29.26/0.8093	29.82/0.8407



Original (PSNR/SSIM)	Bicubic 27.01/0.9481	LapSRN[18] 32.86/0.9881	DRRN[9] 33.88/0.9903	MRLN(ours) 36.25/0.9933

Fig. 5 Super resolution results of image "ppt3" from Set14 with scale factor ×2. The words are clear and sharpen in the result of our method.



Original (PSNR/SSIM)	Bicubic 23.88/0.5817	LapSRN[18] 24.82/0.6568	DRRN[9] 24.86/0.6600	MRLN(ours) 25.59/0.7169

Fig. 6 Super resolution results of image "101085 " from BSD100 with scale factor ×3, The edge of stone head is sharpen in the result of our method.

Especially for scale factor ×2 on Set5, MRLN outperforms MDSR [10] and RDN [11] by 0.64dB and 0.51dB respectively, demonstrating the strategy of multiple residual learning is an accurate method for SR. For SSIM, MRLN exhibits outstanding results and contributes to a great promotion even when the corresponding PSNR value is not the highest one, which indicates MRLN can furthest keep the structure similarity between the ground truth image and reconstructed image.

To visually show the results of MRLN, some example images on scale factor ×2, ×3 are given in Fig. 5 and Fig. 6. The proposed method produces detailed textures and edges in the reconstructed HR images and exhibit better-looking super resolution outputs compared with the previous methods.

IV. Conclusion

In this paper, we propose an advanced multiple residual learning network (MRLN) for SR. By multiple residual learning , the information complexity of target residual image is gradually reduced and the difficulty of learning residual image is simultaneously alleviated. On the other hand, the task for establishing a mapping from LR to HR image is completed by multiple basic models cascaded each other, training process is eased and convergence rate of network is greatly accelerated. We demonstrate that multiple residual learning provides a feasible and effective way to solve super resolution problem, extensive benchmark experiments and analysis also show MRLN is a concise and superior method for SR.

References

[1] A.Marvao, T. Dawes, D. ORegan, and D. Rueckert. Cardiac, "Image super-resolution with global correspondence using multi-atlas patchmatch," in *MICCAI*, 2013.

[2] W. Zou and P. C. Yuen, "Very low resolution face recognition Problem," *IEEE Transactions on image processing*, vol. 21, pp. 327–340, Feb. 2012.

[3] C. G. Marco Bevilacqua, Aline Roumy and M.-L. A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *BMVC*, 2012.

[4] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *ACCV*, 2014.

[5] C. Dong, C. C. Loy, X. Tang, "Learning a deep convolutional network for image super-resolution," in *ECCV*, 2014.

[6] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super resolution using very deep convolutional networks," in *CVPR*, 2016.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016.

[8] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," In *CVPR*, 2016.

[9] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *CVPR*, 2017.

[10] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in CVPRW, 2017.

[11] Y. Zhang, Y. Kong, B. Zhong, Y. Tian, Y. Fu, "Residual dense network for Image super-resolution," in *CVPR*, 2018.

[12] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *AISTATS*, 2011.

[13] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super resolution via sparse representation," *IEEE Transactions on image processing*, vol. 19, pp. 2861–2873, Nov. 2010.

[14] D. Martin, C. Fowlkes, and J. Malik,"A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *ICCV*, 2001.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *CoRR*, abs/1502.01852, 2015.

[16] Y. Tai, J. Yang, X. Liu, and C. Xu, "Memnet: A persistent memory network for image restoration," in *ICCV*, 2017.

[17] C. Dong, C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38. pp. 295–307, Jan. 2016.

[18] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super resolution," in *CVPR*, 2017.