The following publication Shiyu Hu, Muwei Jian, Guodong Wang, Yanjie Wang, Zhenkuan Pan, and Kin-Man Lam "Deep skip connection and multi-deconvolution network for single image super-resolution", Proc. SPIE 11515, International Workshop on Advanced Imaging Technology (IWAIT) 2020 is available at https://doi.org/10.1117/12.2567030.

Deep Skip Connection and Multi-Deconvolution Network for Single Image Super-Resolution

Shiyu Hu[a], Muwei Jian[b], Guodong Wang*[a], Yanjie Wang[a], Zhenkuan Pan[a], Kin-Man Lam[c]
[a]College of Computer Science and Technology, Qingdao University
PR China, 266071. (e-mail: doctorwgd@gmail.com).
[b]School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, China. (e-mail: jianmuweihk@163.com).
[c]Centre for Signal Processing, Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong. (e-mail: enkmlam@polyu.edu.hk)

ABSTRACT

In this paper, we propose an efficient single image super-resolution (SR) method for multi-scale image texture recovery, based on Deep Skip Connection and Multi-Deconvolution Network. Our proposed method focuses on enhancing the expression capability of the convolutional neural network, so as to significantly improve the accuracy of the reconstructed higher-resolution texture details in images. The use of deep skip connection (DSC) can make full use of low-level information with the rich deep features. The multi-deconvolution layers (MDL) introduced can decrease the feature dimension, so this can reduce the computation required, caused by deepening the number of layers. All these features can reconstruct high-quality SR images. Experiment results show that our proposed method achieves state-of-the-art performance.

Keywords: Super-resolution, convolutional neural network, peak signal-to-noise ratio, deep skip connection, multi-deconvolution layers

1. INTRODUCTION

In real applications, the visual content or objects captured in an image are of low resolution. Therefore, it is necessary to process the image so as to increase its resolution and quality. Many learning-based methods 14-16] have been proposed for Single-Image Super-Resolution [SISR] [1, 11-13]. However, the conventional methods have limited learning capacity, so the super-resolved images have limited quality. Recently, deep neural networks [7, 8, 10] have been proposed, which can achieve much better performance. In this paper, we propose a new SISR model, namely deep Skip connection and multi-Deconvolution Super-Resolution network (SDSR), which is a very deep trainable network. This network adopts Deep Skip Connection (DSC) as building blocks, while multi-deconvolution layers (MDL) as the reconstruction module. Experiment results illustrate that our proposed SDSR achieves much better performance than existing state-of-the-art methods.

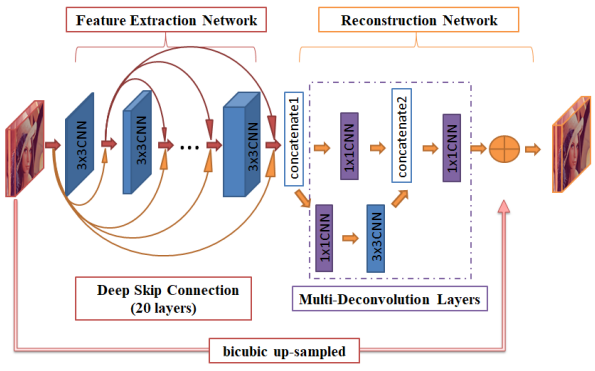
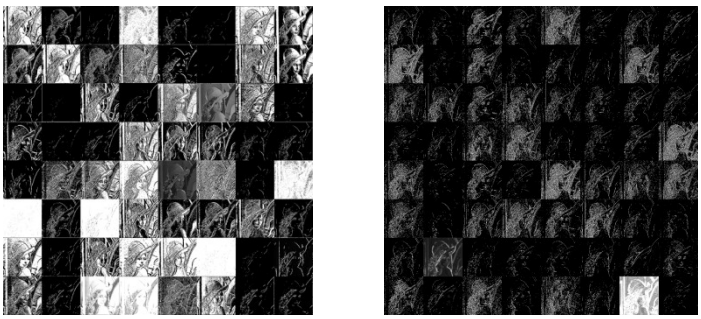


Figure 1. The complete architecture of the proposed model. The network is divided into two parts: a feature-extraction network and a reconstruction network, where boxes with different color represent different convolutional neural networks.

2. METHODOLOGY

Our model is based on a fully convolutional neural network [2]. As shown in Fig. 1, our model is mainly composed of two parts: a feature-extraction network and a reconstruction network. There are 24 convolutional layers in the proposed model, which combines the local features and the global features through skip connections. Each of our convolutional layers contains weights, bias, and cascades of nonlinear layers. The feature maps in the deep network are illustrated in Fig. 2. In addition, the residual learning in the network can better capture the residual information of images, and alleviate the gradient disappearance caused by increasing the depth of the deep neural network.

In the feature-extraction module, due to the use of skip connections, if x_1 denotes the input layer, the input to the i-th layer will be expressed as $x_i = \max(0, w_i * x_{i-1} + b_i)$, where w_i and b_i represent the weight and bias, respectively, of the convolutional layer or deconvolution layer. Different from other major deep-learning-based large-scale image recognition models, the number of convolution kernels used in the convolutional layers of the feature-extraction network is reduced to 32, compared to 96 in [2].



(a) Conv0 (b) Concatenate1

Figure 2. The non-linear feature maps at different layers.

The reconstruction network is similar to the general convolutional layers. The proposed network applies 1×1 convolution layers multiple times, and results in a new structure, as shown in Fig. 3. The operations in the reconstruction network, namely MDL, are defined as follows:

$$M_1 = \theta(w_{1\times1}^1 * x_{n-1} + b^1), \tag{1}$$

$$N_1 = \theta(w_{1\times1}^1 * x_{n-1} + b^1), \tag{2}$$

$$N_2 = \theta(w_{3\times3}^2 * N_1 + b^2), \tag{3}$$

and

$$O = (w_{1\times1}^3 * \langle M_1, N_2 \rangle + b^3), \tag{4}$$

where * represents the convolution operation, and $\langle M_1, N_2 \rangle$ denotes the concatenation operation.

In general, the contributions of this paper are as follows:

- A new deep network, namely SDSR, is proposed, which can reconstruct high-quality high-resolution (HR) images. The proposed network can be trained efficiently, without requiring any weight initialization method and special training techniques to achieve generalization. The trained network has the ability to handle images of different scales, and generates clear texture in images. This is the first deep model that combines skip connection with deconvolution layers to achieve excellent super-resolution performance.
- A new feature-extraction block, namely deep skip connection (DSC), is proposed, which reduces the loss of information, makes full use of the information from all layers, and achieves strong robustness.
- To achieve high-quality image reconstruction, a new model, called multi-deconvolution layers (MDL), is proposed. This model greatly reduces the training parameters and the feature dimensions. Consequently, the training time can be greatly shortened, while the stability of the model can be maintained.

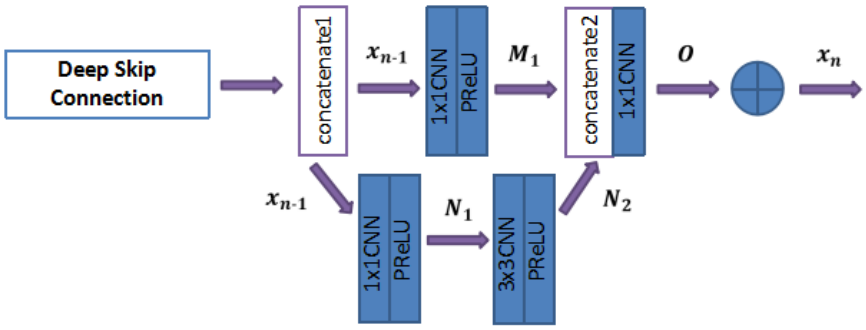


Figure 3. The structure of the multi-deconvolution layers (MDL).

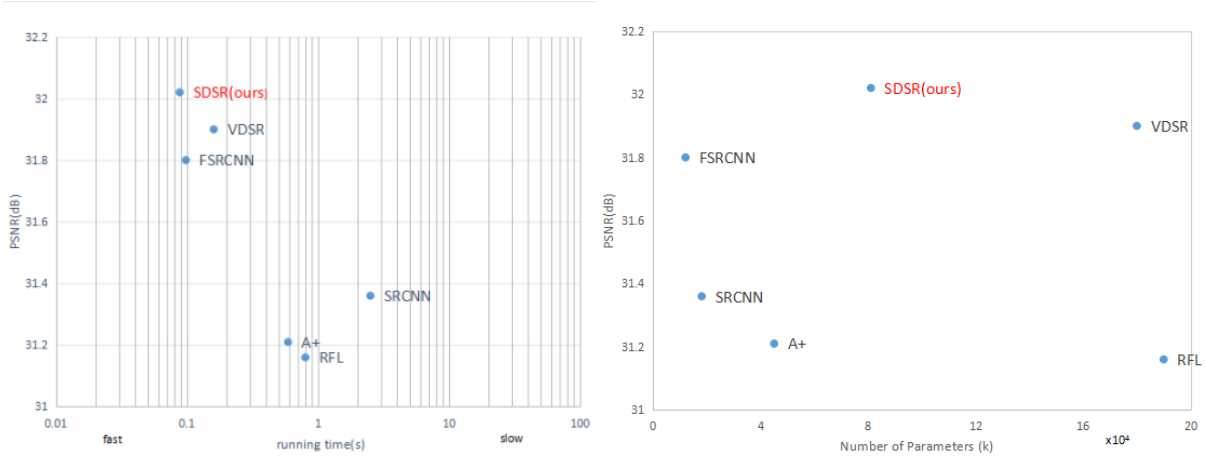


Figure 4. A comparison of the runtimes and the number of parameters required for different SISR methods (upscaling factor = 2, dataset = bsd100).

3. EXPERIMENTS

3.1 Datasets for training and testing

The standard bsd200 dataset consists of 200 natural images, which cover a wide variety and make the training data more representative. Ninety one of the images, proposed by Yang et al. [3], were used to form the training set. The training set is augmented by rotating the images by 90^0, 180^0, 270^0. Four super-resolution standard benchmark datasets, Set5, Set14, bsd100 and Urban100, were used for testing. The peak signal-to-noise ratio (PSNR) and runtime are measured as the evaluation metrics.

3.2 Implementation Details

All the networks are optimized using the Adam [4] method. The initial learning rate is 0.001. If the loss value does not fluctuate in 5 epochs, the learning rate is reduced to half, until the loss is smaller than 2e-5. Parametric ReLU [5] is employed as the activation unit in our proposed network, for tackling the "Dying ReLU" problem.

3.3 Results

Fig. 2 shows the 64 non-linear feature maps of the different layers. In Fig. 4, we give a comparison of the number of parameters and the runtime required for different methods. The evaluation results of our proposed method on the four datasets are shown in Table 1. We can see the superior performance of our method. Five existing state-of-the-art methods are compared, and their performances are also tabulated in Table 1.

Table 1. The average PSNR and runtime, with different upscaling factors (×2, ×3, ×4) on the four benchmark datasets (Set5, Set14, bsd100 and Urban100). The best result and the second-best result are highlighted in red and blue, respectively.

Algorithm	Scale	Set5	Set14	bsd100	Urban100
		PSNR/time	PSNR/time	PSNR/time	PSNR/time
Bicubic	x2	33.66	30.24	29.56	26.88
A+	x2	36.54/0.58	32.26/0.86	31.21/0.59	29.20/2.96
RFL	x2	36.54/0.63	32.26/1.13	31.16/0.80	29.11/3.62
SRCNN	x2	36.66/2.19	32.42/4.32	31.36/2.51	29.50/22.12
FSRCNN	x2	37.00/0.068	32.63/0.160	31.80/0.098	29.66/0.87
VDSR	x2	**37.53**/0.13	**33.03**/0.25	**31.90**/0.16	**30.76**/0.98
SDSR (Ours)	x2	**37.65**/0.010	**33.15**/0.154	**31.97**/0.088	**30.82**/0.780
Bicubic	x3	30.39	27.55	27.21	24.46
A+	x3	32.58/0.32	29.13/0.56	28.29/0.33	26.03/1.67
RFL	x3	32.43/0.49	29.05/0.85	28.22/0.62	25.86/2.48
SRCNN	x3	32.75/2.23	29.28/4.40	28.41/2.58	26.24/19.35
FSRCNN	x3	33.16/0.027	29.43/0.061	28.60/0.035	26.33/0.65
VDSR	x3	**33.66**/0.13	**29.77**/0.26	**28.82**/0.21	**27.14**/1.08
SDSR (Ours)	x3	**33.70**/0.147	**29.83**/0.139	**28.89**/0.065	**27.17**/0.878
Bicubic	x4	28.42	26.00	25.96	23.14
A+	x4	30.28/0.24	27.32/0.38	26.82/0.26	24.32/1.21
RFL	x4	30.14/0.38	27.24/0.65	26.75/0.48	24.19/1.88
SRCNN	x4	30.48/2.19	27.49/4.39	26.90/2.51	24.52/18.46
FSRCNN	x4	30.71/0.015	27.59/0.029	26.98/0.019	24.60/0.74
VDSR	x4	**31.35**/0.12	**28.01**/0.25	**27.29**/0.21	**25.18**/1.06
SDSR (Ours)	x4	**31.57**/0.145	**28.05**/0.12	**27.37**/0.059	**25.24**/0.952

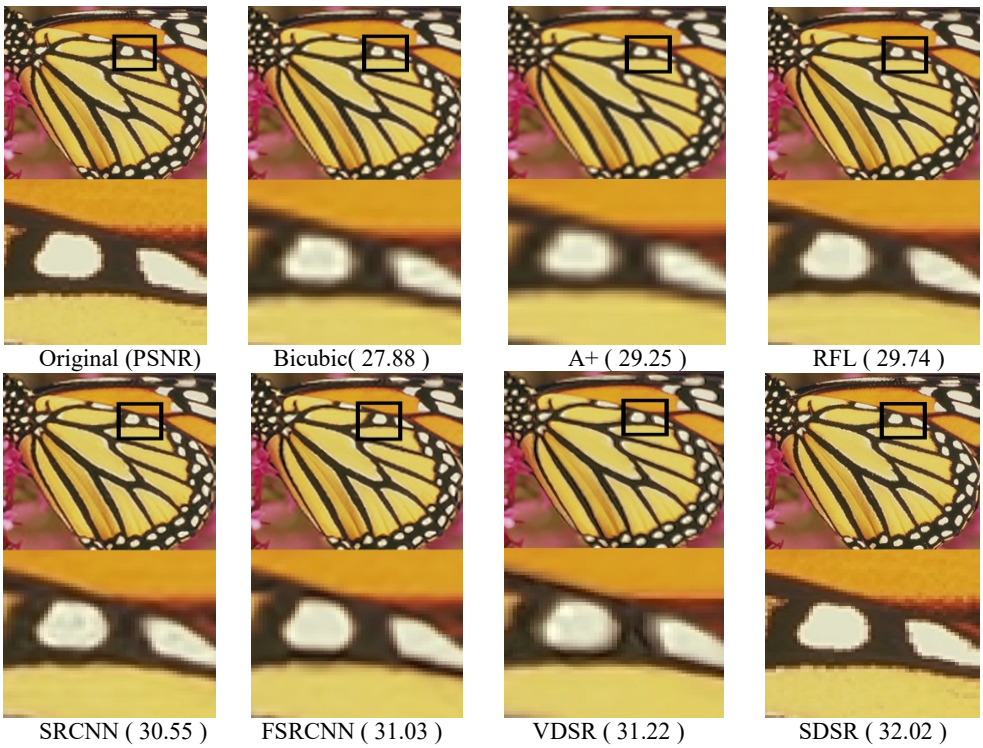


Figure 5. Super-resolution results of the image “img_003 ” in Set5, with upscale factor ×2.

The five methods are denoted as A+ [6], RFL [7], SRCNN [8], FSRCNN [9], and VDSR [10], respectively. Our proposed method achieves the best performance in terms of both accuracy and runtime. Fig. 5 shows the visual performance of the different algorithms, when the upscaling factor is ×2. We can see that our method can reconstruct more clear outlines for the butterfly, as shown in the zoomed box. In other words, the experimental results show that SDSR has superior performance in recovering contours in natural images. Fig. 6 shows the reconstruction results of the different methods, when the upscaling factor is 3. Our method can generate a sharper face contour, when compared to other methods. Fig. 7 shows the results when the upscaling factor is 4. We can see that our method can still produce sharper and clearer visual results. The training time required to train our model is nearly 5 hours, on GTX 1080Ti.

Figure 6. Super-resolution results of the image “img_024” in bsd100, with upscale factor ×3.

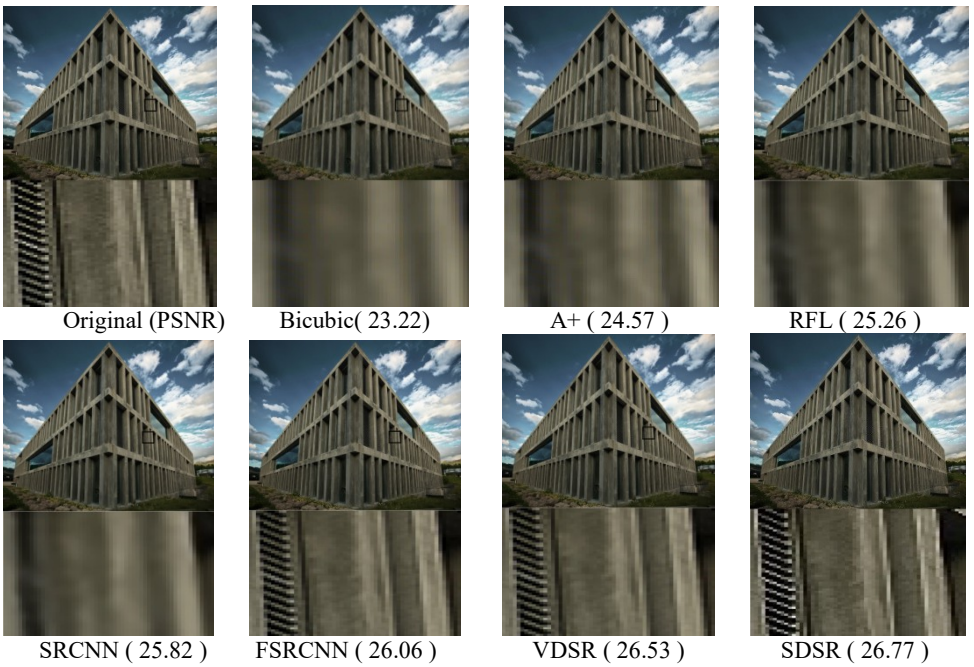


Figure.7 Super-resolution results of the image “img_027” in Urban100, with upscaling factor ×4.

4. CONCLUSIONS

In this paper, we have proposed a novel deep model for single-image super-resolution, which can reconstruct more accurate high-resolution images, through the use of deep skip connection and multi-deconvolution layers. The combined features are then passed to the reconstruction network, which is composed of 1×1 convolutional layers. After using the 1×1 CNNs thrice, the dimension of the input feature is reduced and becomes more accurate, so the high-resolution images can be generated more accurately, without substantially increasing the computation. In the reconstruction network, we combine the global feature after dimensionality reduction of the feature maps and residual learning, and then the final high-resolution image is generated. In addition, our proposed deep model is able to process images of different scales, and can achieve state-of-the-art performance in terms of PSNR and visual quality. In particular, our model achieves the best PSNR among all the methods compared in our experiments.

5. ACKNOWLEDGEMENTS

This work was supported by the Natural Science Foundation of Shandong Province (ZR2019MF050) and the National "Twelfth Five-Year" development plan of science and technology (No.2014BAG03B05). This work was also supported by National Natural Science Foundation of China (NSFC) (61601427, 61976123); Royal Society - K. C. Wong International Fellowship (NIF\R1\180909); Mount Tai Youth Talents Plan.

REFERENCES

[1] M. Irani, S. Peleg: Improving resolution by image registration. In Graphical models and image processing, pp. 231–239, (1991)

[2] M. Thom, R. Schweiger, G. Palm, A. Rothermel: Learning convolutional neural networks from few samples. In International Symposium on Neural Networks, pp. 1884–1890, (2013).

[3] Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image Super Resolution via Sparse Representation. In: IEEE Transactions on Image Processing, vol. 19, no. 11, pp. 2861-2873, (2010).

[4] G. E. Hinton, N. Srivastava, A., Krizhevsky, I. Sutskever, R.R. Salakhutdinov: Improving Neural Networks by Preventing Coad aptation of Feature Detectors, In Neural and Evolutionary Computing, (2019).

[5] He, K., Zhang, X., Ren, S., Sun, J.: Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: IEEE International Conference on Computer Vision, pp. 1026-1034, (2015).

[6] Timofte R, De Smet V, Van Gool L. A+: Adjusted anchored neighborhood regression for fast super-resolution. Asian Conference on Computer Vision, pp.111–126, (2014).

[7] C. Dong, C. Loy, K. He, X. Tang: Accelerating the Super-Resolution Convolutional Neural Network. In European Conference on Computer Vision, pp. 391–407, (2016).

[8] C. Dong, C. Loy, K. He, X. Tang: Learning a Deep Convolutional Network for Image Super-Resolution. In European Conference on Computer Vision, pp. 184–199, (2014).

[9] R. Nayak, D. Patra: New single-image super-resolution reconstruction using MRF model. In: Neurocomputing, Volume 293, pp. 108-129, (2018).

[10] J. Kim, J. Kwon Lee, and K. Mu Lee: Accurate image super-resolution using very deep convolutional networks. In CVPR, (2016).

[11] Muwei Jian, Kin-Man Lam, Simultaneous Hallucination and Recognition of Low-Resolution Faces Based on Singular Value Decomposition, IEEE Trans. on CSVT, 25(11), pp. 1761-1772, (2015).

[12] Muwei Jian, Kin-Man Lam, Junyu Dong, A Novel Face-Hallucination Scheme Based on Singular Value Decomposition, Pattern Recognition, 46 (11), 3091-3102, (2013).

[13] Muwei Jian, Huaxiang Zhang, Liqiang Nie, Yilong Yin, Multi-view face hallucination using SVD and a mapping model, Information Sciences, 488, 181-189, (2019).

[14] S. Baker and T. Kanade, “Limits on super-resolution and how to break them,” IEEE Trans. on Pattern Analysis and Machine Intelligence, 24(9), pp. 1167-1183, (2002).

[15] H. Chang, D. Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," Proc. IEEE Int. Conf. on CVPR, (2004).

[16] Yu Hu, Kin Man Lam, Guoping Qiu, Tingzhi Shen, "From Local Pixel Structure to Global Image Super-resolution: A New Face Hallucination Framework," IEEE Trans. on Image Processing, 20 (2), pp. 433-445, (2011).