Improved Residual Network for Single Image Super Resolution

Yinxiang Xu, *Seungwoo Wee, **Jechang Jeong

Department of Electronics and Computer Engineering, Hanyang University

xuyinxiang@naver.com, *slike0910@hanyang.ac.kr, **jjeong@hanyang.ac.kr

Abstract

In the classical single-image super-resolution (SISR) reconstruction method using convolutional neural networks, the extracted features are not fully utilized, and the training time is too long. Aiming at the above problems, we proposed an improved SISR method based on a residual network. Our proposed method uses a feature fusion technology based on improved residual blocks. The advantage of this method is the ability to fully and effectively utilize the features extracted from the shallow layers. In addition, we can see that the feature fusion can adaptively preserve the information from current and previous residual blocks and stabilize the training for deeper network. And we use the global residual learning to make network training easier. The experimental results show that the proposed method gets better performance than classic reconstruction methods.

1. Introduction

Single image super–resolution (SISR) recovers a high–resolution image from a single low-resolution image. It is a classical problem in computer vision and many Conventional SISR methods have been proposed. Such as interpolation-based [8] and reconstruction-based [9] methods. At present, these image super-resolution reconstruction methods are widely used in medical imaging, video surveillance, digital television, and other fields. Although these two different methods improve the resolution of the image and image quality, the reconstructed image still has the problem of blurred texture. And the entire reconstruction process consumes a lot of calculations.

Recently, learning-based approaches have become a hot topic for tackling the image restoration problem. The method mainly learns a mapping between high-resolution images and corresponding low-resolution images. Then, the learned mapping relationship is used to reconstruct high resolution images.

Among them, Dong et al [1] firstly proposed the use of a three-layer convolutional neural network for image super-resolution (SRCNN) and achieved excellent improvement over traditional methods. Since the SRCNN is too simple to fully extract the detailed features of the image, it causes the reconstructed image is still a bit fuzzy. Kim et al [2] increased the depth of the network and proposed the use of a 20-layer convolutional neural network structure named very deep convolutional networks for super resolution (VDSR). The network of VDSR is obviously much deeper than SRCNN, but they only connect the first layer to the last layer. It causes image feature loss when learning features in the middle layer.

To solve these drawbacks, in Section 3, we introduced an improved residual network that is mainly composed of multiple residual learning blocks. The residual learning block (RLB) proposed in this paper combines low-level features and high-level features with skip connections [7]. This is beneficial to learn effective features, enhance the delivery of features to provide richer information for image reconstruction. In addition, the connection method makes the network difficult to over-fitting. And it can reduce the problem of vanishing gradient. Then by using the feature fusion [3], the current and previous features are adaptively preserved. And it also improves the stability of the network.

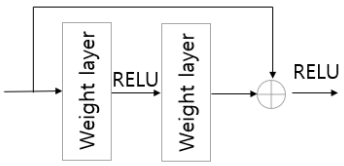

Figure 1. Structure of the residual block

2. Related Work

In recent years, convolutional neural networks have proposed new ideas for solving image super-resolution problems. Dong et al [1] firstly proposed SRCNN consisted of three layers to extract image features. SRCNN is a typical end-to-end network model used to learn a mapping from LR to HR patches. Authors attempted to build the deeper network, but the performance has not improved significantly. The generated image still looks blurry and it takes a long training time. However, the proposed residual network (ResNet) greatly reduces the network training time and achieves the deeper network structure. Since then, residual networks have been widely used in image super-resolution reconstruction problems. The VDSR proposed by Kim et al [2] also applied the residual network to make the network deeper to 20 layers, and performance is much better than SRCNN. But the network only connects the first layer and the last layer to implement residual learning. It still occurs training instability. And Some high-resolution features extracted from each layer are lost, which makes it difficult to improve image performance. In addition, the network handled multiple scales of super-resolution in the network, but it results in more memory usage than architectures with a single scale of super-resolution. We are inspired by the ResNet to deal with the above problem. Our proposed model in this paper improves the structure of residual block and combines the advantage of the VDSR. Next, we introduce the simple structure of the ResNet and VDSR.

2.1 ResNet

He et al [4] proposed ResNet, which is suitable for the deep network to solve the problems that gradient gradually disappears in backpropagation. This makes it impossible to optimize the weights of the first few layers of the network, so that the deep network cannot converge.

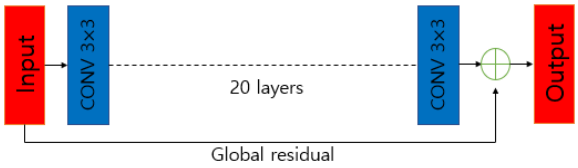

Figure 2. Simple structure of VDSR

But the residual network directly connects the shallow network to the deep network by adding a skip connection (Identity Map) [7]. It allows the gradient to be well transmitted to the shallow layer and avoids the problem of degradation. The residual block is shown in Figure 1.

Weight layer contains a convolutional layer and batch normalization (BN). Firstly, input image x passes weight layers and an activation function. After that, it generates an output. We denote Y as the output of the residual block.

$$Y = max(0, f(x) + x) \tag{1}$$

where $\max(0, \cdot)$ denotes the rectified linear units (ReLU).

2.2 VDSR

Figure 2 shows the simple structure of VDSR. To expand the receptive fields of the network, VDSR deepened the network to 20 layers. it can better extract the correlation between image pixels and improve network performance. VDSR firstly considered that the input image and the output image have great similarities. And the network constructed residual images and use global residual learning to solve the problem that is difficult to converge. Then using gradient cropping strategy to speed up training and make the training easy.

3. Proposed Method

3.1 Proposed Network

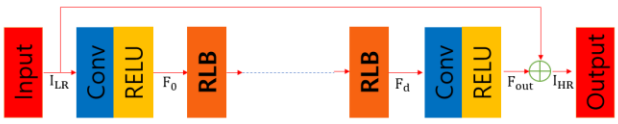

Figure 3. Architecture of our proposed residual network

In order to obtain more high-frequency features of images and effectively improve the single image reconstruction, the network in this paper does not simply stack convolutional layers. Our model consists of four parts as shown in Figure 3

The first part that we use one convolution layer to extract shallow features. Let y denote as the low-resolution image. We firstly use a bicubic interpolation algorithm to preprocess y image and denote interpolated image as I_{LR}. Our first Convolutional layer extracts feature F_0 is

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

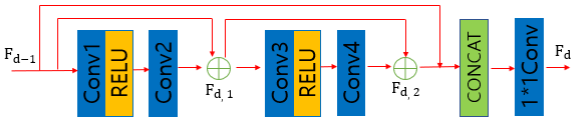

Figure 4. Structure of our residual learning block

where $H_0(\cdot) = \max(0, w * I_{LR})$ and $\max(0, \cdot)$ corresponds to rectified linear units (ReLU). The operator * denotes a convolution, w is weight and F_0 is used as input of the residual learning block.

The second part consists of D residual blocks. The F_d is the output of the d_th residual block and it can be represented as follows:

$$F_d = H_{R,d}(F_{d-1})$$
$$= H_{R,d}\left(H_{R,d-1}\left(\cdots\left(H_{R,1}(F_0)\right)\cdots\right)\right), \tag{3}$$

where $H_{R,d}(\cdot)$denotes the operation of the d− th residual learning block. As shown in Figure 4, our small residual block employs the ResNet with slight modification. We remove the last ReLu part of the residual block to simplify the model. And the purpose of super-resolution reconstruction is to reconstruct the new super-resolution image using the original information of the image. But using BN can change the original information and it also occupies the storage space of the network. BN layer is equivalent to adding a convolutional layer to make network training more complicated. Therefore, we remove the BN layers. Then, each residual learning block contains our two small residual blocks are densely connected by skip connection and feature fusion. By doing so, we can take full advantage of the features extracted by each small residual block and improve image performance. And it makes network training more stable. More details are as follows:

$$F_{d,1} = w_2 * \left(max(0, w_1 * F_{d-1})\right) + F_{d-1} \tag{4}$$

$$F_{d,2} = w_4 * \left(max\left(0, w_3 * F_{d,1}\right)\right) + F_{d,1}, \tag{5}$$

where $F_{d,1}$,$F_{d,2}$ represent the output of each of residual block. Then feature fusion is applied to fuse F_{d-1} and $F_{d,2}$. In addition, the output of the last residual block as the input of the current residual block is concatenated to the feature maps F_d. After that, adding a 1*1 convolution makes the number of output channels consistent with the original.

The third part is image reconstruction. This step is familiar with the first part. The specific formula is as follows:

$$F_{out} = H_0(F_d), \tag{6}$$

where F_{out} denotes an image reconstruction.

Final, we use the global residual learning, which can reduce the problem of vanishing gradient caused by training a deep network. The formula is

$$I_{HR} = I_{LR} + F_{out}, \tag{7}$$

where I_{HR} represent the last generated image.

Figure 5. Super resolution results for image butterfly with scale factor ×2 from set5

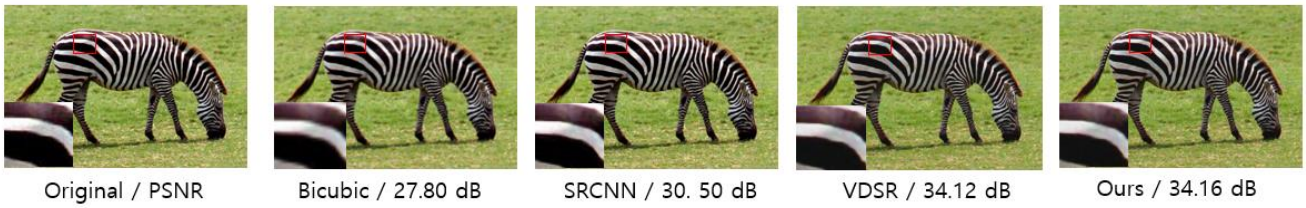

Figure 6. Super resolution results for image zebra with scale factor ×2 from set14

4. Experiment

4.1 Training settings

For training, we use 91 images from the dataset provided by Yong et al [5]. In addition, to get more datasets, we randomly augment the patches by rotating 90, 180, 270, flipping horizontally or vertically and downsizing 05, 0.7. For testing, we use Set5, Set14 that are useful for the benchmark.

The model in this paper is an 18-layer network consisting of four residual learning blocks. All convolutional layers use a 3*3 filter except a convolutional layer after feature fusion. We set the momentum to 0.9 and weight decay to $1e^{-4}$. Batch size is set to 16. The initial learning rate is set to 0.1, which decreases by a factor of 10 every 10 epochs and our total epoch is 50. Training images are divided into 41*41 patches. And we use it as a mini batch for stochastic gradient descent. We implement our proposed model with the Pytorch. Training our model takes about 22 hours with an NVIDIA GeForce GTX 960.

4.2 Loss function

The purpose of super-resolution reconstruction is to make the generated image I_{HR} and the original image as similar as possible. In the paper, the mean square error (MSE) is used as the loss function of the network to estimate and optimize the parameter θ=w. The mathematical formula is as follows:

$$L(\theta) = MSE = \frac{1}{n}\sum_{n=1}^{n}\left\|I_{HR,i} - X_i\right\|^2. \quad (8)$$

In this paper, we use the peak signal-to-noise ratio (PSNR) to evaluate the quality of reconstructed images. The higher the PSNR, the better the predicted image.

$$PSNR = 10 \times log\left(\frac{255^2}{MSE}\right). \quad (9)$$

Table 1. Benchmark results. Average PSNR for scale factor ×2 , ×3 , ×4 o n datasets Set5, Set14.

Dataset	Scale	Bicubic PSNR	SRCNN PSNR	VDSR PSNR	Ours PSNR
Set5	× 2	33.66	36.66	37.39	37.43
	× 3	30.39	32.75	33.58	33.62
	× 4	28.42	30.48	31.26	31.29
Set14	× 2	30.24	32.42	32.89	32.93
	× 3	27.55	29.28	29.55	29.73
	× 4	26.00	27.49	27.92	27.94

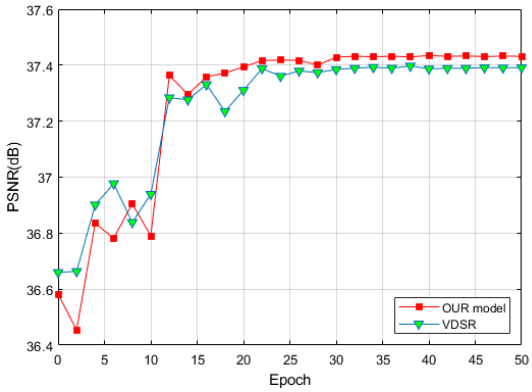

Figure 7. Convergence of our model and VDSR

4.3 Results

For the results of the Table 1, we used 91images to train our proposed network and VDSR. Our experimental results show that the average PSNR is improved by 0.03dB, 0.08dB compared with VDSR on Set5, Set14, respectively. In Figure 7, it is obvious that the performance of our model is higher than that of VDSR and both models begin to converge when epoch is equal to 30. In addition, we found that the proposed method improves the stability of the model when training epoch is equal to 15, while VDSR optimizes parameters to achieve optimal performance when epoch is equal to 25. It means that our proposed residual learning block is more stable training network than the directly stacked convolutional layer.

Table 2. Depth of the network results. Average PSNR for scale factor ×2 , ×3 , ×4 on datasets Set5

Dataset	Scale	No=4 PSNR	No=6 PSNR	No=8 PSNR	VDSR PSNR
Set5	× 2	37.43	37.47	37.53	37.53
	× 3	33.62	33.71	33.80	33.66
	× 4	31.29	31.43	31.50	31.35

We also study the effect of increasing residual network depth, we test three models with different numbers of residual learning blocks (4,6 and 8). Table 2 shows the super resolution performance of our networks on Set5 with scale factor ×2. We denote No as the number of residual learning blocks. It verifies deepening the depth of the residual network, performance becomes better and PSNR is increased. We also found that the average PSNR is 0.1dB higher than the VDSR (20 layers) when we use 8 residual learning blocks (34 layers). Then, our model reduces training time compared with VDSR. And in Figure 5 and 6, results of images are given. Our method outperforms other classical methods in Set5, Set14. In addition, texture and high frequency details of reconstructed images are better.

5. Conclusion

In this paper, we proposed an improved residual network to achieve single image super-resolution reconstruction. In our residual learning block, two residual blocks are closely combined by a skip connection method. And by using feature fusion method, we designed our network which can fully and effectively utilize the features of shallow layers. Experimental results show that the proposed method has better performance than two conventional methods. And the image quality is improved by our proposed network.

Acknowledgements

This work was supported by the Brain Korea 21 plus Project in 2014.

References

[1] C. Dong, C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” *IEEE transactions on pattern analysis and machine intelligence*, no. 38, vol. 2, pp. 295–307, 2016.

[2] J. Kim, J. Lee, and K. Lee, “Accurate image super-resolution using very deep convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition,* pp. 1646–1654, 2016.

[3] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual dense network for image super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2472–2481, 2018.

[4] 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*, pp. 770–778, 2016.

[5] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolution via sparse representation,” *IEEE transactions on image processing*, no. 19, vol. 11, pp. 2861–2873, 2010.

[6] X. Mao, C. Shen, and Y. Yang, “Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections,” *Advances in neural information processing systems*, pp. 2802–2810, 2016.

[7] M. Bätz, A. Eichenseer, J. Seiler M. Jonscher, and A. Kaup, “Hybrid super-resolution combining example-based single-image and interpolation-based multi-image reconstruction approaches,” *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 58–62, 2015.

[8] K. Kim and Y. Kwon, “Single-image super-resolution using sparse regression and natural image prior,” *IEEE transactions on pattern analysis and machine intelligence*, no. 32, vol. 6, pp. 1127–1133, 2010.