Single Image Super Resolution Reconstruction Based on Recursive Residual Convolutional Neural Network

Shuyi Cao, *Seungwoo Wee, **Jechang Jeong

Department of Electronics and Computer Engineering, Hanyang University

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

Abstract

At present, deep convolutional neural networks have made a very important contribution in single-image super-resolution. Through the learning of the neural networks, the features of input images are transformed and combined to establish a nonlinear mapping of low-resolution images to high-resolution images. Some previous methods are difficult to train and take up a lot of memory. In this paper, we proposed a simple and compact deep recursive residual network learning the features for single image super resolution. Global residual learning and local residual learning are used to reduce the problems of training deep neural networks. And the recursive structure controls the number of parameters to save memory. Experimental results show that the proposed method improved image qualities that occur in previous methods.

1. Introduction

Single-image super-resolution is one of the classic computer vision problems, meaning to restore low-resolution images to high-resolution images. At present, this technology is widely used in various fields such as medical ultrasound imaging, surveillance video, and satellite remote imaging.

In recent years, the existing super-resolution technologies are mainly divided into interpolation-based [2], reconstruction-based [4], and learning-based super-resolution [1]. The most popular super-resolution algorithm is learning-based super-resolution, and the image is reconstructed by training relationships of mapping between low-resolution images and high-resolution images.

Among them, Dong et al. [3] have demonstrated that a convolutional neural network (CNN) can be used to super-resolution (SRCNN). They use a three-layer convolutional neural network to learn the nonlinear mapping between low-resolution images and high-resolution images. The results of this method are significantly better than the traditional non-deep learning algorithms. Since then, CNN has been widely used in super-resolution methods. However, since the model uses the shallow neural network to directly learn the original mapping function, the image restoration quality is still not satisfied. Then, in 2016, Kim et al. [5] use 20 convolution layers and the global residual method to learn the feature, and the method accelerates the convergence speed of very deep networks (VDSR). The image reconstruction quality is greatly improved. Since then, many researchers use deeper layers to improve performance, but as the number of network layers deepens, not only does the occupied memory increase, but the calculation is more complicated.

We construct a recursive residual network model to build a more compact network. Specifically, the main advantages of this paper are: 1) Introduce local residual network and global residual network to learn and transmit more image information. 2) Use local residual learning to avoid information disappearing in deep network architectures 3) Make model more compact by using the recursive structure to share parameters. 4) Generate multi-scale super-resolution images using multi-scale image training.

2. Related work

2.1 SRCNN

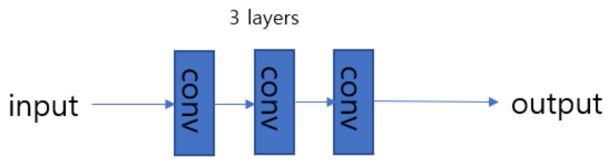

Figure 1. Model structure of SRCNN.

In [3], Model SRCNN consists of three layers: patch extraction and representation, nonlinear mapping, and reconstruction. Figure 1 shows the model structure of SRCNN. Compared with the traditional non-deep learning algorithm, the final image has been greatly improved. But because the SRCNN layer is very shallow, the learning ability is very limited and the effect is not satisfactory.

2.2 ResNet

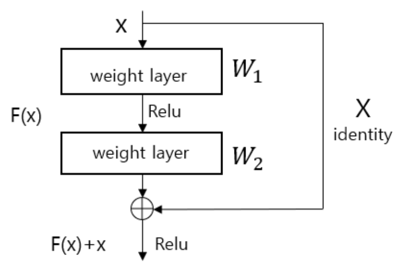

Figure 2. Model structure of the residual block.

Figure 2 shows the model structure of residual block. The residual block has two weight layers. Each weight layer contains a convolution layer and a batch normalization layer (BN). These are represented as follows:

$$F(x, W) = W_2 \delta(W_1 x). \tag{1}$$

The residual network (ResNet) [7] learns the residual function according to the input, making the training of the deep network simpler. The residual block structure is

$$\hat{y} = U(x) = \delta(F(x, W) + x), \tag{2}$$

where $\hat{y}$ is the output of the residual block, $U(x)$ is the residual block function, W is the network weight, $F(x, W)$ is the learned residual map, δ is the activation function Relu function. Residual block directs shallow layer data directly into the deep layer by adding a connection. It can avoid data disappearing during the transfer process.

2.3 VDSR

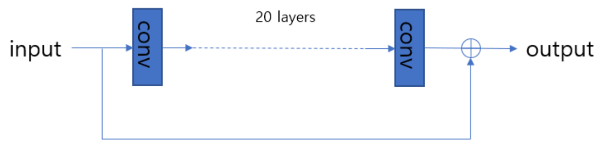

Figure 3. Model structure of VDSR.

In [5], VDSR uses 20 convolution layers to significantly improve image training accuracy. Referring to the ResNet [7] network, VDSR uses a global residual learning structure in the network. One of the advantages is that the gradient disappearance and explosion can be avoided by adjustable gradient clipping. However, deep networks require a lot of parameters, which requires more storage space than the recursive parameter model structure.

3. Method

The deep network structure and multiple secondary associations generate a lot of parameters to occupy the storage space. For solving this problem, a recursive residual network is proposed in this paper. The recursive structure is to avoid introducing new parameters while stacking more layers. And it saves memory space. It's like using the same layer again and again. Local residuals and global residuals are more conducive to the transmission and learning of information.

As shown in Figure 4, first, the network interpolates and extracts features from low resolution images. After using a recursive residual structure for nonlinear mapping, it performs super-resolution image reconstruction.

We refer to the main idea of ResNet [7] and construct a residual block structure with three convolutional layers. First, the image features are extracted by a convolution layer. Then after passing each residual block, we add the feature image extracted from the first convolutional layer. by doing so, we keep the branch inputs of the residuals consistent. More image information is transmitted to the deep network.

The function of the residual network is

$$H^m = R(H^{m-1}) = F(H^{m-1}, W) + H^0, \tag{3}$$

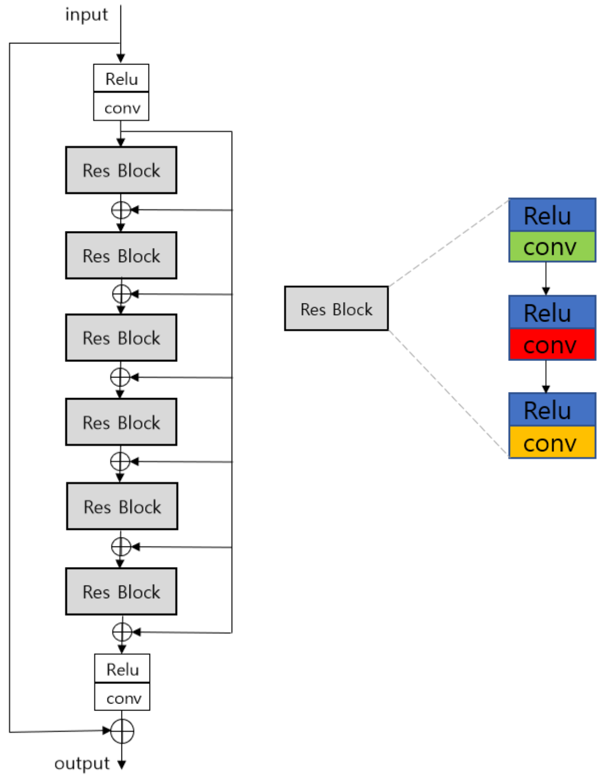

Figure 4. Proposed algorithm flowchart.

where H^m is the output of the m-th residual block, R is the residual block function, $F(H^{m-1}, W)$ is the learned residual map, and H^0 is the feature image that is output through the first convolution layer.

We use a recursive structure to control the number of parameters, which means that the weight parameters between the residual blocks are shared. The above function formula can also be written as:

$$H^m = R^{(m)}(H^0) = R\big(R \ldots R\big(R(H^0)\big) \ldots\big). \tag{4}$$

4. Experiment

The number of convolution layers is the same as that of VDSR [5], which is 20 layers. The first layer operates on the input bicubic image, and the last layer is used for high-resolution image reconstruction by a filter of size 3×3. Except the last layer, every layer has 64 filters of size 3×3. Each residual block consists of three convolution layers in each residual block, for a total of six residual blocks.

4.1 Experimental details

Training images are 91 images of Yang et al [8], and we rotate these images by $90°$, $180°$, $270°$ and flip them to make data enhanced. In theory, the larger the data set, the better the training effect. Then, we also use scale augmentation to train, the image with different scales ($\times 2, \times 3, \times 4$) are added to the training set. We change the recursion and residuals, and test with the Set5 [9] test set. And the results are shown in the following table. It can be seen from Table 1 that residual learning can transmit more image information to the deep layer of the network, and the recursive structure can improve the image performance while reducing the number of parameters.

Table 1. Effect of residual and recursive structure on experimental results.

structure		PSNR		
Residual	Recursive	Scale=2	Scale=3	Scale=4
X	X	35.51	32.53	30.42
O	X	36.68	32.95	30.69
O	O	37.42	33.66	31.38

In the experiment, the images are split into 41×41 patches, the stride step size is 41, the image batch size is 16, the momentum parameter is 0.9, the initial learning rate is 0.1, the weight decay is 0.0001, the learning rate decreased to 1/10 every 10 epochs, and the learning is stopped after 50 epochs.

The activation function is Relu and the loss function is mean square error (MSE).

$$L = \frac{1}{N}\sum_{i=1}^{N}\|r - f(x)\|^2, \tag{5}$$

where x is interpolated low-resolution image and y is the original high-resolution image. We proposed to learn f that predicts values $\hat{y} = f(x)$. And the residual image $r = y - x$. Because the output image and the input image are largely similar, we only need to calculate the residual image, which will take up less memory.

4.2 Experimental result

The training set is the same as the above experiment, both are 91-image training sets [8]. The test used five images of Set5 [9] test set, fourteen images of the Set14 [10] test set.

Table 2. Average PSNR of various SISR DL methods for scale factor 2,3 and 4 on Set5, Set14.

Dataset	Scale	Bicubic	SRCNN	VDSR	Ours
Set5	× 2	33.66	36.66	37.39	37.42
	× 3	30.39	32.75	33.58	33.66
	× 4	28.42	30.48	31.26	31.38
Set14	× 2	30.24	32.42	32.89	32.89
	× 3	27.55	29.28	29.72	29.74
	× 4	26.00	27.49	27.92	27.97

Training data preparation in the experiment is obtained from the Matlab implementation. We used PyTorch1.0.1 and Python3.6 for training and testing. Training takes about 10 hours on an NVIDIA GeForce GTX 960 GPU.

Table 2 provides a comparison of the PSNR values of the proposed method with the Bicubic, SRCNN [3], VDSR [5] methods. Among them, Bicubic and the SRCNN results from the reference paper [5]. And the VDSR result is our retest based on the 91-image training set [8]. It can be seen from Table 2 that the quality of our reconstructed image is significantly improved compared with other deep-learning (DL) methods.

Figure 5. Results of “butterfly” (Set5) with scale factor ×2.

Figure 6. Results of “comic” (Set14) with scale factor ×2.

Figure 5 and Figure 6 show the image comparison results with the above DL methods. By magnifying the details of the image, we can easily compare the image reconstruction quality.

5. Conclusion

We propose a deep recursive residual network structure, which uses both global residual learning and local residual learning. And we add recursive structure to control the number of parameters. It can make the occupied space less and the calculation difficulty is relatively reduced. The experimental results show that compared with several typical super-resolution image reconstruction methods, the performance of the proposed structural model is better.

Acknowledgements

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

References

[1] W. Freeman, T. Jones, and E. Pasztor, "Example-based super-resolution," *IEEE Computer graphics and Applications* 2, pp. 56–65, 2002.

[2] K. Jensen and D. Anastassiou, "Subpixel edge localization and the interpolation of still images," *IEEE transactions on Image Processing* 4.3, pp. 285–295, 1995.

[3] C. Dong, C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," *European conference on computer vision,* pp. 184–199, 2014.

[4] S. Borman and R. Stevenson, "Spatial resolution enhancement of low-resolution image sequences-a comprehensive review with directions for future research," *Lab. Image and Signal Analysis, University of Notre Dame, Tech. Rep*, 1998.

[5] 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.

[6] J. Kim, J. Lee, and K. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1637–1645, 2016.

[7] 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.

[8] J. Yang, J. Wright, and T. Huang, "Image super-resolution via sparse representation," *IEEE transactions on image processing*, pp. 2861–2873, 2010.

[9] M. Bevilacqua, A. Roum, and C. Guillemot, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Brit. Mach. Vis. Conf*, pp. 1–10, 2012.

[10] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," *International conference on curves and surfaces*. Springer, Berlin, Heidelberg, pp. 711–730, 2010.