ATLANTIS PRESS

Advances in Intelligent Systems Research, volume 146

2nd International Conference on Artificial Intelligence: Technologies and Applications (ICAITA 2018)

Image Super-resolution Reconstruction Based on Deep Residual Network

Bin Sun, Jian Lu* and Xiaopeng Wei

Laboratory of Advanced Design and Intelligent Computing, Dalian University, China

*Corresponding author

***Abstract*—Having powerful ability to learn and represent features, deep convolutional neural networks (CNN) can get better results in image super-resolution reconstruction. However, deep networks also exist some problems. For example, the gradient will gradually vanish through the network, which makes the training difficult to converge. To address these problems, this paper presents a deep network model that is suitable for super-resolution reconstruction. By applying residual network modules, the model realizes the transmission of information across one or more layers. This residual network structure can not only avoid the loss of effective information in the network, but also speed up the training convergence. Compared with the previous classical methods, the proposed model converges faster, and the subjective and objective evaluations have been improved to a certain extent.**

Keywords—super-resolution reconstruction; deep learning; residual network

I. INTRODUCTION

Image Super-resolution (SR) reconstruction is of great significance for scientific research and life. High-resolution images can provide more detailed picture quality and detail information. Therefore, in many fields, such as medical rescue image, safety monitoring, face recognition, high-resolution images are important guarantees for good results. Compared with the direct use of high-cost image acquisition equipment, SR reconstruction does not need to change the existing hardware environment. It is more economical and gradually become a hot spot in recent years.

Dong [1] firstly proposed an image SR method based on convolutional neural network, referred to as SRCNN (SR Convolutional Neural Network). Dong et al. [2] introduces an improved method based on SRCNN. In order to reduce the computational overhead, more convolution layers and smaller convolution kernels are selected, and the interpolation of the low-resolution images is canceled. But straightly to the non-interpolated low image for input, the final selection of transposed convolution layer to enlarge. In 2016, Kim [4] proposed a 20-layer convolutional network structure model VDSR (Image Super-Resolution Using Very Deep Convolutional Networks). In order to speed up the convergence of deep network, the network uses a higher learning rate than SRCNN for network training. On the other hand, global residual learning are also used to be better training. Dong's contribution in the article [5] is to increase the network receptive field (41×41) and use recurrent neural networks(RNN). paper [6] proposed the first framework capable of inferring photo-realistic natural images for 4 upscaling factor.

In this paper, we improve the VDSR network, including the following aspects: (1) Adding global and local residual links in the network can speed up network convergence and better transmission of effective information in the network; (2) In order to reduce the over-fitting, the data set is expanded and the images of different magnitudes are mixed together for training. In addition to the training data, a single network can also support multiple magnification SR reconstruction; (3) We use adaptive moment estimation method to minimize the value of training loss; (4) Searching for appropriate network hyperparameters. Through experiments, we find suitable hyperparameters can effectively improve the quality of reconstruction. In the following chapters, this article details the algorithm in detail.

II. SUPERRESOLUTION-RECONSTRUCTION BASED ON RESIDUAL NETWORKS

A. *Super-resolution Reconstruction Residual Network Model*

A deep network is very necessary for image super-resolution reconstruction. Deep network can predict target pixel information according to more pixels, that is, the deeper network is indeed better. However, we can not increase the depth of the network simply by adding layers, and the existence of the gradient disappears, making the training of deep networks quite difficult. On the premise of keeping the number of network layers unchanged, we try to improve the performance of the network architecture by optimizing the network framework. ResNet has played a major role in hundreds of layers of network training, and the performance is still excellent. The original purpose of ResNet was to reduce the suffering of deep network training.

The following two network frame structure comparison in Figure 2. we use the global residuals and local residuals, residual mapping easier than the original mapping. In order to speed up the convergence of this network, residual is very necessary. Figure 1 shows the VDSR algorithm network structure, the yellow block represents the convolution layer and activation function, a total of twenty groups. The network structure in Figure2 is an explanation of the algorithm for this article. The green cube express a residual block, there are two groups of the same. In the next part of the residual block for a detailed introduction.

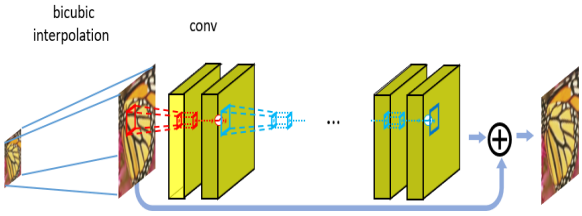


FIGURE I. VDSR

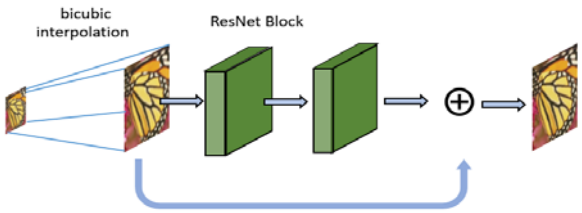


FIGURE II. PROPOSED

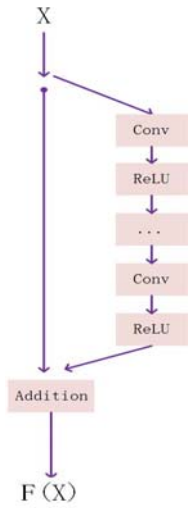


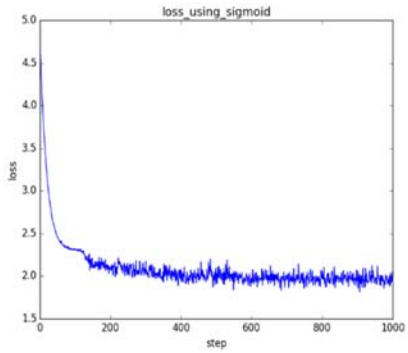
FIGURE III. RESNET BLOCK

B. *Residual Block*

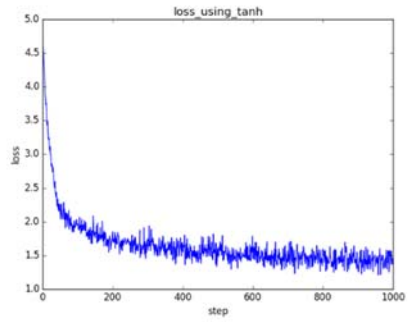
Figure 3 shows the ResNet residual block used by the algorithm in this paper, the structure in the green box in Figure 2. Its function uses the congruent mapping to directly output the previous layer to the following layer. Suppose that a certain neural network input is X, and the expected output is F(X). ResNet is equivalent to changing the learning goal, instead of learning a complete output F(X), but the difference between output and input $F(X) - X$, that is residuals.

Residual learning can be regarded as a good strategy for image super-resolution, since there are many places in the image information that are shared between low-resolution and high-resolution images. In the process of super resolution reconstruction training, CNN learn the difference between high resolution image and low resolution image, the output of CNN is coupled with the original low resolution image to get high resolution image. The reason why high-resolution images can be added and subtracted with low-resolution images is that we have used low-resolution images to be scaled to the same size as high-resolution images during data preprocessing. Its advantage is that the amount of information carried by CNN is small, it is easier to converge and at the same time it can achieve better effect than non-residual network. Super-reconstruction, image noise reduction are image restoration problems, the details are very important. The depth learning method uses the convolution layer instead of the pool layer, so it can retain more details of the image. Activation function layer using the ReLu function [7], its role is to increase the nonlinear relationship between the layers of the neural network to better fit the data.

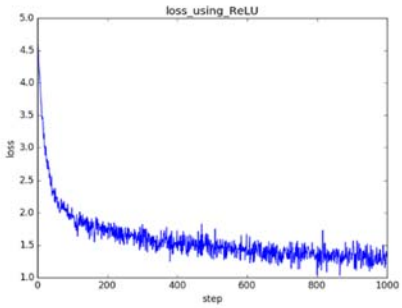
We select three activation functions separately, all adopt Adam learning type. After 1000 iterations, the cost function curve is shown in Figure 4. The loss value of the Sigmoid activation function is reduced to about 2, and the cost function of the Tanh activation function is reduced to a relatively small value, while the ReLu activation function's concussion is smaller. So ReLu-type activation function is more suitable for this network framework. In the next experiment, we use ReLu-type activation function by default.



(a) Sigmoid



(b) TanH



(c) ReLu

FIGURE IV. DIFFERENT ACTIVATION FUNCTION LOSS FUNCTION CURVE

III. EXPERIMENTS

A. Dataset

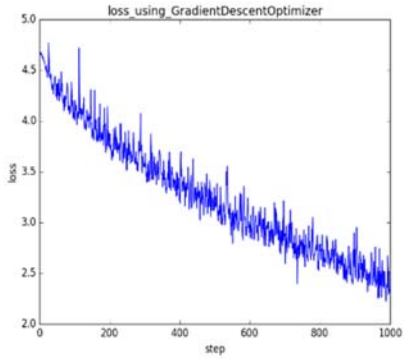
Learning from [4], we use the training set of 291 images. 91 images come from Yang et al. [8] and other 200 images are from Berkeley Segmentation Dataset [9]. Inspired by Timofte [6], it is necessary to flip and rotate the training images. The extensive training library not only provides more extensive and comprehensive features but also prevents over-fitting. Be enlightened by VDSR, we also use scale augmentation to apply our model. Different scales (×2, ×3, and ×4) images are included in the training dataset. So we can only train one model to test different scales images. About the choice of test dataset, Set5[9], Set14[10], and Urban100[11] are widely used, including 5, 14 and 100 images each .

B. Implementation Details

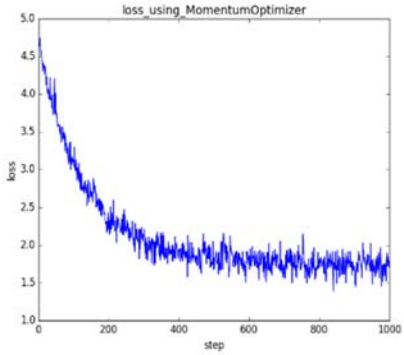
The Euclidean distance of the network output image and original image is taken as the loss function. During training, Caffe calculates the gradient of the network parameters and updates the network parameters to minimize the loss function. In deep learning, the loss function is usually non-convex and should be solved by the optimization method. For a dataset, the objective function that needs to be optimized is the average of all data loss across the dataset. At the same time the data set is divided into several batches, the loss function at this time:

$$\mathrm{L}(w) = \frac{1}{|D|}\sum_{i}^{|D|} fw(X^i) + \lambda \mathrm{r}(w) \qquad (1)$$

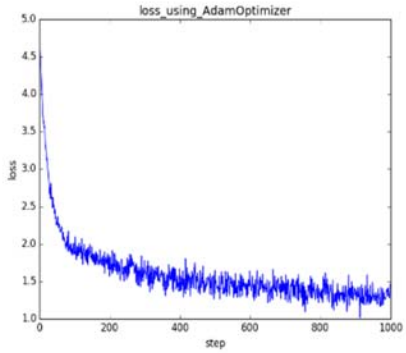
Here, $\mathrm{fw}(X^i)$ calculates the loss on the X^i. Then sum, and find the mean of them. r(W) is regular item, in order to weaken the phenomenon of over fitting. Adam optimization algorithm is more suitable for large data sets and high-dimensional space, less memory requirements, most of non-convex optimization. In order to choose the optimization method that is more suitable for the network model, experiments on the above three algorithms. While maintaining the basic parameter settings, three different learning algorithm cost function curves are shown in Figure 5. SGD convergence is relatively slow, while the upper and lower shocks more obvious; SGD + Momentum (Stochastic Gradient Degradation plus Momentum Factor) Oscillation decreases, converge in a faster time than SGD; The Adam converges fastest, and smaller oscillation. Can speculate more iterations, and the ideal results can be obtained. VDSR uses SGD + Momentum (Stochastic gradient descending plus momentum factor) for optimization solution, this article uses the adaptive moment estimation (Adam).



(a) SGD



(b) SGD+Momentum



(c) Adam

FIGURE V. DIFFERENT LEARNING ALGORITHM LOSS FUNCTION CURVE

C. Subjective and Objective Evaluation Results

Figure 6 show the results of five algorithms including interpolation, A+ [3], SRCNN [1], VDSR [4]and the improved algorithm in this paper, which scale ×3 of 4 images. The algorithms of SRCNN and VDSR, which are based on deep convolutional networks. They are obviously superior to the A + algorithm and the double-three interpolation algorithm. The

VDSR algorithm and the algorithm in this paper are clearer, and the edge preservation is better, which is close to the original image effect. However, the results of other algorithms are not satisfactory.

FIGURE VI. IMAGE SCALE FACTOR ×3

The objective evaluation indicators adopted in this paper are still PSNR and SSIM. It can be seen in Table 1 that the results of this paper are better than SRCNN and interpolation both in terms of PSNR and SSIM. And compared with the VDSR algorithm, objective evaluation has been slightly improved.

TABLE I. AVERAGE PSNR/SSIM FOR SCALE FACTOR ×2, ×3 AND ×4 ON DATASETS SET5, SET14AND URBAN100

Dataset	Scale	Bicubic PSNR/SSIM	A+ PSNR/SSIM	SRCNN PSNR/SSIM	VDSR PSNR/SSIM	Proposed PSNR/SSIM
Set5	×2	33.66/0.9299	36.54/0.9544	36.66/0.9542	37.53/0.9587	37.62/0.9659
	×3	30.39/0.8682	32.58/0.9088	32.75/0.9090	33.66/0.9213	33.77/0.9225
	×4	28.42/0.8104	30.28/0.8603	30.48/0.8628	31.35/0.8838	31.44/0.8842
Set14	×2	30.24/0.8688	32.28/0.9056	32.42/0.9063	33.03/0.9124	33.14/0.9123
	×3	27.55/0.7742	29.13/0.8188	29.28/0.8209	29.77/0.8314	29.88/0.8232
	×4	26.00/0.7027	27.32/0.7491	27.49/0.7503	28.01/0.7674	28.12/0.7701
Urban 100	×2	26.88/0.8403	29.20/0.8938	29.50/0.8946	30.76/0.9140	30.74/0.9013
	×3	24.46/0.7349	26.03/0.7973	26.24/0.7989	27.14/0.8279	27.21/0.8156
	×4	23.14/0.6577	27.32/0.7183	24.52/0.7221	25.18/0.7524	26.24/0.7516

IV. CONCLUSION

In this paper, we propose a method of using global residuals and local residuals in deep networks. The network learns the residual images, the computation is small, and the network is easy to converge. In order to get the parameters that match the network, a series of comparative experiments were conducted. At last, we test the test dataset by the model, and the subjective and objective effects are compared with those of the classic algorithm. The result is better. The next step is to study how to better train the ultra-deep network and get the ideal reconstruction performance.

ACKNOWLEDGMENT

This work is supported by Liaoning Provincial Natural Science Foundation of China (20170540039).

REFERENCES

[1] C. Dong, C. C. Loy, and K. He. Image super-resolution using deep convolutional networks[J]. IEEE transactions on pattern analysis and machine intelligence. 2016.

[2] C. Dong, C. Loy, and X. Tang. Accelerating the super resolution convolutional neural network. In ECCV, 2016.

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

[4] J. Kim, J. K. Lee, and K. M. Lee. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In CVPR, 2016.

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

[6] C. Ledig, L.Theis and F. Huszar. Photo-realistic single image super-resolution using a generative adversarial network. 2016.

[7] V. Nair, and G. Hinton. Rectified linear units improve restricted boltzmann machines. In ICML.

[8] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang. Deep networks for image super-resolution with sparse prior. In ICCV, 2015.

[9] C. M. Bevilacqua, A. Roumy, and M. - L. A. Morel. Lowcomplexity single-image super-resolution based on nonnegative neighbor embedding. In BMVC, 2012.

[10] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. In Curves and Surfaces, pages 711-730. Springer, 2012.

[11] Jia-Bin. Huang, A. Singh, and N. Ahuja. Single image superresolution from transformed self-exemplars. In CVPR, 2015.