Application of Wavelet Transform in Image Super-Resolution

Yucong Zhao, Xiaonan Zhu, Ping Wang, Yuanchen Wang and Jiquan Ma*

Computer Science and Technology, Heilongjiang University, Harbin, Heilongjiang, 150080, China
*Corresponding author’s e-mail: majiquan@hlju.edu.cn

Abstract. Most of learning-based super-resolution (SR) methods reconstructed high resolution (HR) image using a very deep convolutional neural network (CNN). CNN-based methods are used to process low resolution images, the predicted image is too smooth and some texture details are ignored. In order to improve the quality of the recovered high resolution image, wavelet transform is applied at the end of our model. Firstly, the prediction image is obtained through the CNN model; then, the predicted image and the ground truth image are processed into four frequency channels by wavelet transform; at last, four frequency channels and original image are considered when calculating loss. Compared with the original CNN model, the results of adding wavelet transform is obviously improved.

1. Introduction
Wavelet transform is an efficient and intuitive multi-resolution image representation and storage tool. It can describe the context and texture information of an image at different levels, which let us to introduce wavelet transform into CNN-based super-resolution network. The result of 1-level wavelet transform is shown in Figure 1, the coefficient approximation (LL) represents the low frequency signal, including different levels of image compression; the detail coefficient (LH, HL and HH) represents the high frequency signal of the image, including the texture detail and structure information of the image.

In order to make full use of the advantage of wavelet transform, we combine wavelet transform with convolutional neural network (CNN) model. Our method reconstructs SR image with rich texture and global information by continuously reducing the wavelet coefficient error of LR image and HR image. To evaluate our approach, we reproduced the original model. The evaluation results showed that the results of our method was improved compared with the original CNN model.
2. Related Work

Image super-resolution methods are mainly divided into three types: statistic-based methods [1, 2], interpolation-based methods and learning-based methods [3, 4, 5]. In the early stage, the statistics-based and interpolation-based methods are popular because of their high computational efficiency, but their results are not obvious on large scale factors. The learning-based approach uses content image to infer missing high frequency information. In recent, deep learning-based methods have been applied in the field of super-resolution because of their strong learning ability. Generally, the MSE loss function is used in existing models, but it makes the output high resolution image too smooth.

With the development of super-resolution technology, Kim et al. proposed that VDSR [6] used residuals to train by increasing the depth of the network. Shi et al. first proposed a real-time super-resolution algorithm ESPCN [7] called sub-pixel convolution, ESPCN magnified the image at the end of the network to reduce the amount of computation. Lim et al. proposed a deeper and wider residual network EDSR [8], which removed the batch normalization layers and used the remaining scaling to speed up training. Zhang et al. proposed that RDN [9] was a residual block, which contained dense residual blocks and dense connections as a whole.

A wavelet-based super-resolution method has been proposed. For super-resolution of a single image, wavelet transform is mainly used for interpolation-based and static-based methods. Bae et al. proposed a wavelet residual network (WavResNet) [10] for image denoising and SISR, they found that learning CNN on wavelet subbands was beneficial to get a better performance network. Guo et al. proposed a deep wavelet super-resolution (DWSR) [11] method to recover the missing details on the subband. In the above methods, the loss function only considered the deviation between the ground truth and predicted image. We decomposed the ground truth and prediction image through 1-level wavelet transform, then we calculated the error of the high frequency and low frequency part.

3. Method

In this section, we will present our model. Here, we applied our method in SRCNN [12] and VDSR [6]. Our method will be described in section 3.1 and section 3.2.

3.1. Wavelet SRCNN

SRCNN consists of patch extraction and representation, non-linear mapping and reconstruction three layers. Our method is to add the wavelet transform at the end of the network; we applied the 1-level wavelet transform to the predicted and the ground truth image, then we used the MSE loss function to calculate the error between the LL frequency channels, the error was used to continuously promote the learning of the network.

3.2. Wavelet VDSR

VDSR introduces a global residual learning, which is the residual learning between the input $t^{HR}$ image and the output $t^{LR}$ image. VDSR uses 20 convolution layers in the residual portion to obtain a larger receptive field (41×41). In VDSR, our error consisted of two parts: one was that the predicted image and the ground truth image by 1-level wavelet transform, then we calculated the error between LL channels by L2 norm; the other calculated the error between the predicted image and the ground truth image. The sum of two parts as a total loss continues to promote network learning.

4. Experiment

4.1. Datasets

To highlight the effectiveness of our approach, we still used dataset from the original papers for training. SRCNN training dataset used 91 images from Yang et al. VDSR training dataset used 291 images with the addition of 200 images from Berkeley Segmentation [13]. For the test data, we used the benchmark dataset: Set5 [14], Set14 [15].
4.2. Training
In SRCNN, the training parameters refer to the parameter settings of the original article. The first layer has a convolution kernel size of 9 x 9 x 64; the second layer uses a 1 x 1 x 32 convolution kernel to control the number of output feature maps; the third layer uses a 3 x 3 x 1 convolution kernel to reconstruct high resolution images. The patch size was set to 33, and the batch size was 128. In SRCNN, the author experimented with the Caffe framework. We retrained it using the TensorFlow framework on Nvidia GeForce GTX 1080 Ti, the initial learning rate was set to 0.01, the decay rate was set to 0.98 and the learning rate was updated every 1480 steps. We set up training within 15000 epochs.

In VDSR, the training parameters refer to the parameter settings of the original article. VDSR is mainly composed of 20 layers of convolution. Except for the last layer of convolution, the convolution kernel of each layer is 3 x 3 x 64. We retrained VDSR using the TensorFlow framework, learning rate was set to 0.0001. The patch size was set to 41, and the batch size was set to 64. We trained the experiment 50 epochs.

In all experiments, we converted the RBG color channel to the YCbCr color channel. We only considered the Y channel during training and evaluation. We mainly used PSNR and SSIM [16] for evaluation.

5. Result
The SRCNN author implemented the project under the caffe framework. We used the TensorFlow framework to retrain SRCNN and VDSR. We tested them with Set5 and Set14 dataset. The comparison of PSNR and SSIM values of our methods in VDSR and SRCNN is shown in Table 1, Table 2. The visual effect is shown in Figure 2 and Figure 3.

Table 1. In SRCNN, average PSNR/SSIM for scale factor ×2, ×3 and ×4 on datasets Set5, Set14. Red color indicates the best performance.

| Dataset | Scale | bicubic  | srcnn     | our       |
|---------|-------|----------|-----------|-----------|
| Set5    | ×2    | 27.88/0.8620 | 29.20/0.8591 | 29.32/0.8605  |
|         | ×3    | 25.62/0.7881 | 26.96/0.8057 | 26.81/0.7947  |
|         | ×4    | 23.67/0.7236 | 24.82/0.7440 | 25.13/0.7422  |
| Set14   | ×2    | 25.51/0.7763 | 24.87/0.7819 | 24.65/0.8017  |
|         | ×3    | 23.56/0.6782 | 23.01/0.7042 | 23.20/0.7124  |
|         | ×4    | 22.27/0.6987 | 21.73/0.6336 | 22.28/0.6409  |

Table 2. In VDSR, average PSNR/SSIM for scale factor ×2, ×3 and ×4 on datasets Set5, Set14. Red color indicates the best performance.

| Dataset | Scale | Bicubic  | VDSR       | our       |
|---------|-------|----------|------------|-----------|
| Set5    | ×2    | 33.65/0.8886 | 37.18/0.9138 | 37.19/0.9229  |
|         | ×3    | 30.37/0.8125 | 33.32/0.8737 | 33.33/0.8781  |
|         | ×4    | 28.38/0.7474 | 30.97/0.8261 | 31.04/0.8250  |
| Set14   | ×2    | 30.22/0.8215 | 32.68/0.8665 | 31.04/0.8703  |
|         | ×3    | 27.53/0.7202 | 29.58/0.7787 | 29.64/0.7792  |
|         | ×4    | 25.99/0.6433 | 27.76/0.7037 | 27.78/0.7076  |
6. Discussion
In SRCNN and VDSR, we conducted four groups of experiments: the predicted image and the ground truth image were processed into four frequency channels by wavelet transform, 1. four channels were considered when calculating loss; 2. four channels and original image were considered when calculating loss; 3. only LL channel was considered when calculating loss; 4. LL channel and original image were considered when calculating loss. The experimental results showed that the experimental results of SRCNN in Experiment 3 and VDSR in Experiment 4 were better than original model. Due to the image is divided into high frequency part and low frequency part by wavelet transform, high frequency information (LL, LH and HH) has less image sharing. We believe that wavelet transform can also be added to the network framework, which can better promote network learning.

7. Conclusion
We applied wavelet transform to SRCNN, VDSR for image reconstruction. SRCNN consists of 3-layer convolution. VDSR used a very deep network of residual learning. Because of the wavelet transform has the function of restoring texture details and sharp parts, we applied wavelet transform to the end of each network structure to achieve reconstruction. Experimental result show that the results got by our method are better than by the original model.
References

[1] Sun, J., Xu, Z., & Shum, H. Y. (2008) Image super-resolution using gradient profile prior. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1-8.

[2] Yang, C. Y., & Yang, M. H. (2013) Fast direct super-resolution by simple functions. In: Proceedings of the IEEE international conference on computer vision. pp. 561-568.

[3] Chang, H., Yeung, D. Y., & Xiong, Y. (2004) Super-resolution through neighbor embedding. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2004. CVPR 2004. IEEE. Vol. 1, pp. I-I.

[4] Huang, J. B., Singh, A., & Ahuja, N. (2015). Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5197-5206.

[5] Singh, A., Porikli, F., & Ahuja, N. (2014) Super-resolving noisy images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2846-2853.

[6] Kim, J., Kwon Lee, J., & Mu Lee, K. (2016) Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1646-1654.

[7] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., ... & Wang, Z. (2016) Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1874-1883.

[8] Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017) Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 136-144.

[9] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018) Residual dense network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2472-2481.

[10] Bae, W., Yoo, J., & Chul Ye, J. (2017) Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 145-153.

[11] Guo, T., Seyed Mousavi, H., Huu Vu, T., & Monga, V. (2017) Deep wavelet prediction for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 104-113.

[12] Dong, C., Loy, C. C., He, K., & Tang, X. (2014) Learning a deep convolutional network for image super-resolution. In: European conference on computer vision. Springer, Cham. pp. 184-199.

[13] Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Vancouver: Iccv.

[14] Bevilaqua, M., Roumy, A., Guillemot, C., & Alberi-Morel, M. L. (2012) Low-complexity single-image super-resolution based on nonnegative neighbor embedding.

[15] Zeyde, R., Elad, M., & Protter, M. (2010) On single image scale-up using sparse-representations. In: International conference on curves and surfaces. Springer, Berlin, Heidelberg. pp. 711-730.

[16] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004) Image quality assessment: from error visibility to structural similarity. In IEEE transactions on image processing., 13(4): 600-612.