Super-Resolution Reconstruction of Single Image based on Multiscale Recursive Residual Network

Mi Dong¹, Jian Huang¹

¹College of communication and information Engineer, Xi’an University of science and technology, Xi’an, Shaanxi, 710054, China

*Corresponding author’s e-mail: dongmi0316@163.com

Abstract. In order to solve the problems of single image super-resolution algorithm based on convolutional neural network, such as shallow network structure, single feature extraction scale and fuzzy texture of reconstructed image, a single image super-resolution reconstruction method based on multi-scale convolutional neural network is proposed. The convolution of several different scales is used to check the original low-resolution image for feature detail extraction, and the recurrent residual network is used to gradually restore the high-frequency information of the image with fewer parameters. Finally, the final reconstructed image is obtained by fusion of the features extracted by different convolutional kernels. The experimental results show that the algorithm proposed in this paper has obtained better image super-resolution result, can obtain more detailed information, and get better visual effect. The classical reconstruction methods such as bicubic, SRCNN, FSRCNN, ESPCN, VDSR are compared, and the peak signal-to-noise ratio (PSNR) and the structure similarity (SSIM) are higher than that of the existing algorithms.

1. Introduction

Image super-resolution reconstruction technology refers to the restoration of a high resolution image (HR) from one or more low resolution images (LR)[1-3]. Currently, the classical image super-resolution research methods are divided into three categories: interpolation-based methods[4], reconstruction-based methods[5] and learning-based methods[6]. In 2016, Dong et al.[7] applied the depth learning method to the image super-resolution for the first time, and proposed a method of directly learning the end-to-end mapping between low-resolution and high-resolution images, called the super-resolution convolution neural network (SRCNN), using the low-resolution image as an input. The experimental results show that the reconstruction effect of this method is superior to other non-deep learning methods. In order to speed up the speed, Dong et al.[8] proposed an improvement to the SRCNN algorithm. A fast image super-resolution reconstruction (FSRCNN), based on deconvolution is proposed to directly take the original low-resolution image as the input of the network, and a deconvolution layer is added to the end of the network to realize the upsampling amplification of the image, which greatly reduces the computational complexity. However, the above is to use shallow network to learn the mapping between low/ high resolution, the structure is simple, and the detail feature information of low resolution image is not fully extracted, so the effect of reconstruction image is not good. In the following research, Kim et al.[9] were inspired by VGG-net for ImageNet classification, used a very deep network (20 layers), and proposed an image super-resolution based on deep convolution neural network (VDSR). The network solves the gradient disappearance problem caused by the depth network by learning the high frequency partial residual between the HR image...
and LR image, and improves the reconstruction quality of the image, but the deepening of the network increases the training parameters of the network, resulting in large memory occupation and high computational complexity. Although all the above image super-resolution reconstruction methods improve the effect of image reconstruction, there are still some shortcomings, and all of them extract features from LR images under a single scale, ignoring the image details at different scales.

In view of the above problems, a single-image super-resolution reconstruction method based on multi-scale convolution neural network is proposed in this paper.

2. Proposed method
Image super-resolution reconstruction is to extract the detail feature information of low-resolution image, and the corresponding high-resolution image is obtained by feature mapping, reconstruction and other steps. However, the previous research is to extract the features of LR images on a single scale, ignoring the image details extracted at different scales, which leads to the reconstruction of image texture blurring and so on. In order to obtain more high frequency features of images and effectively improve the accuracy of super-resolution reconstruction of single image, a single image super-resolution reconstruction method based on multi-scale convolution neural network is proposed in this paper, which consists of three parts: multi-scale feature extraction module, recurrent residual module and high-resolution image reconstruction module. The multi-scale feature extraction module uses convolution kernel sizes of different scales to extract the features of the low-resolution image directly, then uses the recurrent residual module to gradually restore the high-frequency information of the HR image with fewer parameters, and gradually recovers the high-frequency information of the image through the cyclic residual block. Being different from using bicubic interpolation to sample LR images, finally, the high-resolution reconstruction module uses a deconvolution layer to expand the feature map. It not only reduces the computational complexity, but speeds up the convergence speed of the network. The overall network structure diagram is shown in figure 1.

2.1 Multi-scale feature extraction
The feature extraction unit consists of three parts, which use the convolution kernels of 3×3, 5×5 and 7×7 scales to extract the features of the original low-resolution image at the same time. Compared with the previous one, the feature extraction of the LR image is performed by a single scale, the proposed method can extract various features of the original LR image, and fuse the feature information of multiple images to facilitate the learning of the correspondence between the LR image and the HR image, thereby ensuring the detail clarity of the image reconstruction.

2.2 Recursive residual network
The traditional convolution neural network super-resolution method has only three-layer network, and the receptive field is low. By deepening the network structure, it is possible to increase the receptive field for effectively extracting more image detail features, but the deepening
of the network structure may lead to the decrease of learning rate, gradient disappearance or gradient explosion. In order to speed up the convergence of the network, this paper uses the local residual learning of multi-path mode and the combination of global residual learning and multi-weight recurrent learning to improve the learning efficiency. Its structure is shown in figure 2. The local residual element structure is to stack the local residual unit sequentially, and different residual units use different inputs. On this basis, all the residual units have the same input, that is, the output of the first convolution layer in the recurrent block, and each residual unit contains two convolution layers. In a recurrent block, the convolution layer parameters corresponding to the same position in each residual unit are shared. In order to avoid the loss of image information after convolution, zero filling is used to ensure that the output image is equal to the input image size in the convolution process.

2.3 Feature Fusion and Reconstruction
Before feature reconstruction, the feature map of multi-scale extracted image feature after recurrent residual learning is fusion, and the original low-resolution image obtained by bicubic interpolation is merged with the feature map after recurrent residual again to form global residual learning. This can supplement details of the image and improve training efficiency of the network.

The reconstruction module uses the deconvolution layer to expand the feature map and does not perform up-sampling operations using bicubic like SRCNN. Deconvolution reconstruction of super-resolution image has the following advantages:

- Accelerate the reconstruction process of super-resolution image, after adding deconvolution layer at the end of the network, the whole image feature extraction is completed on the original LR image, which greatly reduces the computational complexity.
• Different from the traditional interpolation method, the deconvolution method can learn the upsampling kernel of image features and can be regarded as the inverse computation of convolution. For convolution operation, when the filter moves step, the output is input, on the contrary, the output of deconvolution operation is twice as much as that of input. When the amplification factor is equal to the magnification factor, the super-resolution image of the required size can be directly output by the deconvolution layer, and the computational complexity of the network can be reduced.

3. Experiment results and analysis

3.1 Datasets
In order to compare with the existing methods, this paper uses the same training set and test set, and selects 91 standard images widely used in learning-based super-resolution algorithm for training. In order to avoid overfitting, the data enhancement technique is used to expand the training set of the image. All the images are rotated according to 90°, 180° and 270°, and then the images are zoomed to 0.5 times, 0.6 times, 0.7 times, 0.8 times and 0.9 times of the original image. The training image is expanded 24 times to a total of images \( 91 \times 24 = 2184 \), of which 1820 images are used as the training set, the remaining 364 images are used as the verification set of the training process, and Set5 and Set14 are used as the test sets. Before the network training, the image is preprocessed, all the images of the training set are down-sampled twice to get the corresponding low-resolution image, and then the image is clipped to \( 25 \times 25 \) sub-image as the input of the network according to the step size of 14.

3.2 Experimental results
In order to test the performance of the algorithm proposed in this paper, the experiments are compared with the existing classical super-resolution reconstruction methods such as Bicubic, SRCNN, FSRCNN, VDSR. All the algorithms carry out super-resolution reconstruction under the condition of magnification of 2 times, 3 times and 4 times. Set5 and Set14 datasets are used as test sets, and three images with rich edge details are selected for testing. Figure 3, figure 4 and figure 5 are the effect images reconstructed by different super-resolution reconstruction methods with magnification of 2, 3 and 4, respectively. From the subjective effect, the image effect of bicubic reconstruction is the most blurred, the overall visual effect of SRCNN reconstruction image is clearer, but the image edge is still blurry, the image reconstructed by FSRCNN and VDSR is relatively clear, and the details at the edge are restored more completely, but it is easy to produce pseudo shadow linearity. Compared with the above algorithms, the reconstructed image proposed in this paper can restore more image details, sharper edges, higher recognition.

4. Conclusion
In order to solve the problems that the existing image super-resolution reconstruction algorithms image texture region, a single image super-resolution reconstruction method based on multi-scale convolution neural network is proposed in this paper. Multiple convolution kernels of different scales are used to extract features directly from LR images, and the details of more LR images are extracted, and the model combining local residual learning of multi-path mode with multi-weight recurrent learning is used to learn a variety of high-frequency information of the image. Finally, the image super-resolution reconstruction is realized by using deconvolution at the end of the network. The experimental results show that the proposed algorithm is superior to other existing algorithms in subjective visual effect and objective evaluation index, and can extract the features of the original LR image and restore the texture information of the image more fully. The next work can further optimize the network model while ensuring the speed to improve the accuracy of image super-resolution, and the algorithm can be applied to other fields, such as medical image super-resolution reconstruction.
Figure 3. “butterfly” image with an upscaling factor 2.

Figure 4. “bird” image with an upscaling factor 3.
Figure 5. “baby” image with an upscaling factor 4.

References
[1] Zhang H, Zhang L, Shen H. (2015) A blind super-resolution reconstruction method considering image registration errors. International Journal of Fuzzy Systems,17(2): 353-364.
[2] Chen H, He X, Teng Q. (2016) Single image super resolution using local smoothness and nonlocal self similarity priors. Signal Processing Image Communication,43: 68-81.
[3] Li S M, Lei G Q, Fan R. (2017) Depth map super-resolution reconstruction based on convolutional neural network. Acta Optica Sinica,37(12):1210004.
[4] Ji C T, He X H, Fu Y Q. (2014) An edge-oriented interpolation algorithm based on regularization. Journal of Electronic Information,36(2):293-297.
[5] Panyan V, Elad M. (2016) Multi-scale patch-based image restoration. IEEE Transactions on Image Processing,25(1): 249-261.
[6] Peleg T, Elad M. (2014) A statistical prediction model based on sparse representations for single image super-resolution. IEEE Transactions on Image Processing, 223(6): 2569-2582.
[7] Dong C, Chen C L, He K M, Tang X O. (2014) Learning a deep convolutional network for image super-resolution. In: Proceedings of the 13th European Conference on Computer Vision, Zurich Switzerland. Springer, Cham. 184-199.
[8] Dong C, Chen C L, Tang X O. (2016) Accelerating the supersolution convolutional neural network. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands. Springer, Cham. 391-407.
[9] Kim J, Lee J K, Lee K M. (2016) Accurate Image Super-Resolution Using Very Deep Convolutional Networks. IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society,1646-1654.