Super-resolution of compressed image by deep residual network

Yan Jin, *Bumjun Park, **Jechang Jeong
Hanyang University
jinyanbs@gmail.com, *kkbbbj1@naver.com, **jjeong@hanyang.ac.kr

Abstract

Highly compressed images typically not only have low resolution, but are also affected by compression artifacts. Performing image super-resolution (SR) directly on highly compressed image would simultaneously magnify the blocking artifacts. In this paper, a SR method based on deep learning is proposed. The method is an end-to-end trainable deep convolutional neural network which performs SR on compressed images so as to reduce compression artifacts and improve image resolution. The proposed network is divided into compression artifacts removal (CAR) part and SR reconstruction part, and the network is trained by three-step training method to optimize training procedure. Experiments on JPEG compressed images with quality factors of 10, 20, and 30 demonstrate the effectiveness of the proposed method on commonly used test images and image sets.

1. Introduction

Single image super-resolution (SISR) is a class of techniques that reconstructs a high-resolution (HR) image from a single low-resolution (LR) image. Recently, deep neural networks provide significantly improved performance in the field of PSNR in the SR problem. SRCNN [2] is a representative method in SR problem. SRCNN used just three layers to generate the mapping between LR and HR images. VDSR [3] utilized residual learning strategy and has 20 layers. With the single network, VDSR can jointly handle SR of several scales. EDSR [4] employed residual blocks, and removed unnecessary modules: batch normalization layers.

Image information has obvious advantages, intuition, vividness, and more. It is one of the important methods for people to obtain information. But, image information has a large amount of data. Data compression techniques is widely used to solve the problem of data volume and it reduces the time and transmission bandwidth costs. So how to compress image data as much as possible without affecting the acquisition of image information is the goal of many researchers [1].

Image information is generally down sampled and compressed to reduce data volume. Nevertheless, this process usually suffers from burring and the compression artifacts which makes the SR problem more challenging. There are two procedures in compressed images. Removing compression artifacts (CAR) and SR reconstruction [5]. In this paper, we propose a network for compressed images SR reconstruction, which simultaneously reduces compression artifacts and enhances image resolution.

This paper is organized as follows. Section 2 describes proposed method, section 3 presents experimental results, section 4 concludes this paper.

2. Proposed Method

The SISR problem requires a LR image Y to obtain a desired higher resolution image X. Y is a blurred and down sampled version of X.

$$Y = SHX + n \tag{1}$$

where S denotes the down sampling operator, H denotes a blurring filter, and n donates additive noise.

In this work, we try to address the problem of compressed image SR. Therefore, we can change the Eq. (1) to

$$Z = CSHX \tag{2}$$

where C denotes compressed operator. In order to obtain convenient representation, we still let

$$Y = SHX \tag{3}$$

then

$$Z = CY \tag{4}$$

In this paper, we aim to reconstruct $\hat{X}$ as much close as original HR image X from Y and Z.

2.1 Proposed Network

To train our network in an end-to-end manner, the input of the network is the original image and the output of the network is a reconstructed image which has same resolution with the input image. In general, the proposed network consists of two modules: compression artifacts removal (CAR) and SR reconstruction.

The whole process is shown in Figure 1. First, the proposed method down sampled the original image, and encodes the down sampled LR image by the encoder. Then, we transmit the encoded bitstream to the decoder. Next, we remove compression artifacts of the decoded images to generate better input for SR reconstruction. Finally, the original resolution image is obtained by SR reconstruction. Both network modules have their own lost functionality, but it is a trainable end-to-end network.

2.1.1 Compression Artifacts Removal

Joint photographic experts group (JPEG) is a lossy compression method for still images, and the image lose a lot of information after the JPEG compression. The smaller quality factor (QF), the greater the information loss. The goal of CAR is to improve the quality of decoded image. We use denoising method to process CAR which was introduced in [8]. CAR consists of three layers. The first convolution layer, with 64 filters of $3 \times 3 \times c$ kernel size, accepts compressed LR image Z as an input. Here, c represents the number of image channels. Intermediate layers applied residual learning strategy. We set 16 layers where all of intermediate layers generate 64 feature maps with $3 \times 3 \times 64$ filter size. Each layer includes convolution (Conv), batch normalization (BN), and rectified linear unit (ReLU). BN was applied to speed up training time as well as boost the deblocking performance. In the last layer, c filters of size $3 \times 3 \times 64$ is used to reconstruct the output, as shown in Fig.1. Combining Conv and ReLU can gradually separate image structure from the deblocking artifacts through the hidden layers.

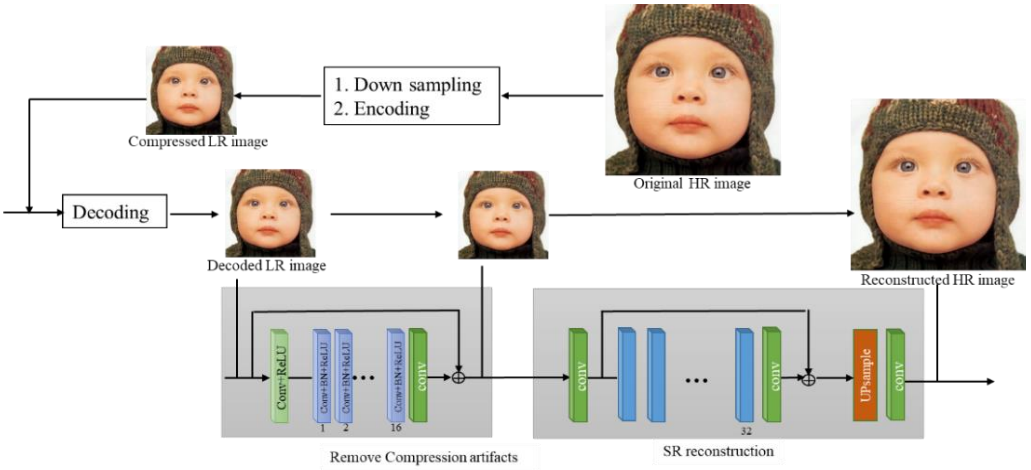


Fig.1: The architecture of the proposed network

2.1.2 Super Resolution Reconstruction

SR reconstruction is also composed of three types of layers. The first convolutional layer has 256 feature maps with 3×3 kernel size. Intermediate part uses residual learning strategy with residual blocks, as described in [4]. Each residual block consists of Conv, ReLU, and Conv. Each convolution has 256 feature maps with 3×3 kernel size. [4] mentioned that when computing resources are limited, increasing the number of feature maps can maximize model capacity. Nevertheless, increasing the number of feature maps to a certain level can make the training process drastically unstable, and they solved this problem by using the residual scaling of factor 0. 1. Inspired by them, we dealt with the unstable problem with the same method. The last convolution operation generates 256 feature maps with 3×3 kernel. We utilized sub pixel convolution [7] to up sample LR image to desired HR image.

3.Training Details

For training, we used RGB input and set the patch size as 48×48. All of these patches are compressed with JPEG. We used general-100 dataset as a training dataset which contains 100 images [6]. We trained our model using Adam optimizer with initial leaning rate 0.0001, and set loss function as L1 loss.

$\{X_i,\ Y_i,\ Z_i\}_{i=1}^{N}$ denote HR, corresponding LR, compressed version of Y, respectively. When training the whole network, it is divided to three steps. First, the set Y_i and Z_i were used to train CAR network. Adopting residual learning strategy, the aim of the network is to learn a residual mapping. The output of the network is $\hat{Y}_i$ which is estimated Y_i from Z_i. Next, we trained the SR reconstruction network. The training set for the SR reconstruction process are X_i, $\hat{Y}_i$. This work aims to learn a mapping from $\hat{Y}_i$ to X_i. Finally, the network is optimized in an end-to-end manner. After we initialize our whole network with the learned parameters of the CAR and SR reconstruction, we optimize the whole network with fine-tuning strategy. Initializing the whole network with the learned parameters made our training process more stable when we process final joint optimization process. Through the above training strategies, the goal of each module could be achieved, and the final joint optimization process minimized the prediction error. We set three different QF values of 10, 20, 30, and we could use the fine-tuning strategy to trained the network instead of starting from scratch.

4.Experimental results

For testing the performance of the proposed method, we used four data sets commonly used in the SR problems: Set5, Set14, B100, and Urban100. In our experiments, all test images were compressed by JPEG. Table 1 shows the PSNR results, and Table 2 shows the PSNR results of the Set5 dataset with different QFs. Table 3 shows the PSNR comparison between the proposed method and conventional methods setting the QF as 10, and Figure 4 shows the result images. As shown in Figure 4, Bicubic, sparse coding [9], and VDSR still suffers from the serious compression artifacts. [10] is a learning-based method that jointly trained SR with deblocking for compressed image (LBJSRD), it blurred unexpected details. However, proposed method preserved details of the images.

Table 1. comparisons of average PSNR (dB) on different dataset

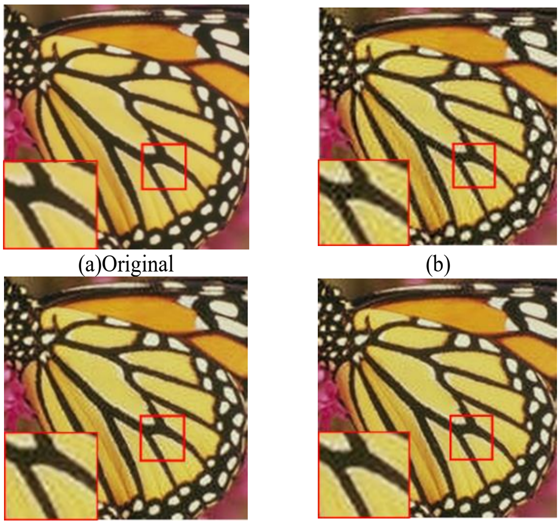
Dataset	**Set5**	**Set14**	**B100**	**Urban100**
Q=10	27.991	26.004	24.998	24.172
Q=20	29.268	27.503	26.384	25.857
Q=30	30.193	28.275	27.353	26.713

Table 2. comparison of PSNR（dB）on Set5 dataset

Test images	**Baby**	**Bird**	**Butterfly**	**Head**	**Woman**	**Average**
Q=10	30.356	28.425	24.304	28.574	28.297	27.991
Q=20	31.498	29.755	25.236	29.966	29.884	29.268
Q=30	32.623	30.812	26.490	30.341	30.699	30.193

Table 3. comparisons with different method(Q:10) on Set5 dataset

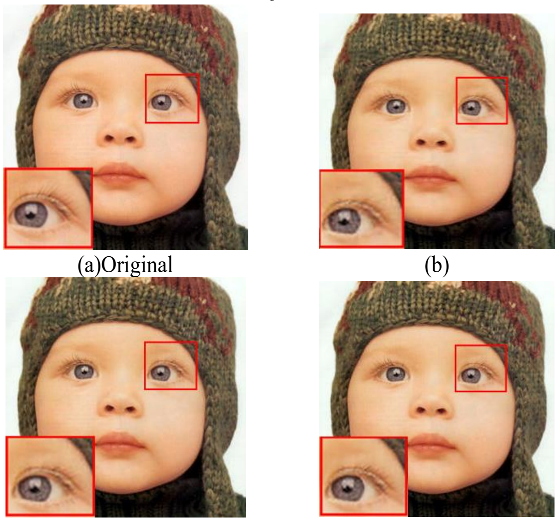
bicubic	**SCSR**	**VDSR**	**LBJSRD**	**Proposed**
26.378	26.476	26.479	26.613	27.991



(a)Original (b)

(c) (d)

Fig.2 super resolution results of proposed method on butterfly (from Set5) with different QFs. (a)Original (b) QF=10 (c) QF=20 (d) QF=30



(a)Original (b)

(c) (d)

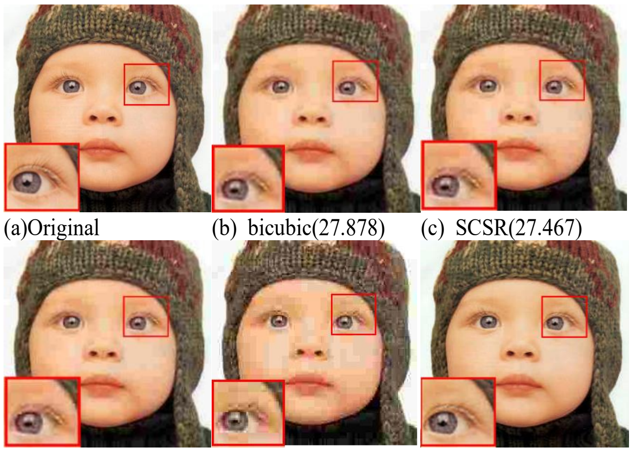
Fig.3 super resolution results of proposed method on baby (from Set5) with different QFs. (a)Original (b) QF=10 (c) QF=20 (d) QF=30

5.conclusion

In this paper, we proposed a SR solution for compressed images. Unlike other existing compressed image SR methods, we consider this task as two related sub-problems, CAR and SR. Through three-step training, the CAR part and the SR reconstruction part of the network are separately trained and jointly trained to optimize the whole network. It is shown that the proposed network can generate SR results of the compressed images effectively. However, due to the limitation of computational complexity, the model proposed in this paper has not been further verified by training with a larger dataset. In the future, we will use larger data sets for verification, and study to reduce the complexity of the model to get a better solution.

Acknowledgement

This work was supported by the ICT R&D program of MSIP/IITP.[2014-0-00670, Software Platform for ICT Equipment]



(a)Original (b) bicubic(27.878) (c) SCSR(27.467)

(d) LBJSRD(27.902) (e) VDSR (28.960) (f) proposed (30.356)

Fig.4 super resolution results of proposed method on baby (from Set5) with different QFs.

References

[1] Miyazawa R, Kudoh S, Nakachi T, et al. “Coding efficiency improvement on distributed video coding utilizing super resolution approach,” *in proc IEEE Conf. 2011 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS)*, December 2011.

[2] C. Dong, C. Change Loy, K. He, X. Tang. “Image Super-Resolution Using Deep Convolutional Networks,” *in proc IEEE Conf. Transactions on Pattern Analysis and Machine Intelligence*, pp. 295-307, Jun. 2015.

[3] Jiwon Kim, Jung Kwon Lee, Kyoung Mu Lee, “Accurate Image Super-Resolution Using Very Deep Convolutional Networks,” *in Proc. The IEEE Conference on Computer Vision and Pattern Recognition*., pp. 1646-1654, Jun. 2016

[4] B. Lim, S. Son, S. Nah, and K. Lee, “Enhanced deep residual networks for single image super-resolution,” *in Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, June 2017.

[5] H Chen, X He, C Ren, L Qing, and Q Teng, “CISRDCNN: Super-resolution of compressed images using deep convolutional neural networks,” *Journal of Neurocomputing*, pp. 204-219, April 2018.

[6] C. Dong, C. Change Loy, K. He, X. Tang. “Accelerating the Super-Resolution Convolutional Neural Network,” *Computer Vision – ECCV 2016*, pp. 391-407, September 2016.

[7] Wenzhe Shi, Jose Caballero, Ferenc Huszar, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, Zehan Wang, “Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network,” *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1874-1883, 2016.

[8] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang, “Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising,” *IEEE Transactions on Image Processing*, July 2017.

[9] Yang Jianchao, John Huang, Thomas S and Ma Yi “Image super-resolution via sparse representation,” *IEEE Transactions on Image Processing*. vol.19, pp. 2861-2873, 2010.

[10] Li-Wei Kang, Chih-Chung Hsu, Boqi Zhuang, Chia-Wen Linand Chia-Hung Yeh, “Learning-Based Joint Super-Resolution and Deblocking for a Highly Compressed Image,” *IEEE Transactions on Multimedia*, pp. 921 – 934, May 2015.