Deep Convolutional Network based Image Quality Enhancement for Low Bit Rate Image Compression

Chuanmin Jia [#], Xiang Zhang [#], Jian Zhang [#], Shiqi Wang [*], Siwei Ma [#**]

[#] *Institute of Digital Media, Peking University, Beijing 100871, China*

[**] *Peking University Shenzhen Graduate School, Shenzhen, China*

$\{cmjia, x\_zhang, jian.zhang, swma\}$@pku.edu.cn

[*] *Nanyang Technological University, Singapore 639798*

wangshiqi@ntu.edu.sg

***Abstract*—In this contribution, a novel image quality enhancement algorithm based on convolutional network is proposed for low bit rate image compression. Specifically, a downsample procedure is performed to generate lower resolution image for low bit rate compression. While the decoder side, upsample is to be performed firstly to the original resolution. Image quality is further enhanced by the proposed convolutional deep network. In particular, an optional image quality improvement network can be utilized for further enhancement after the first network. With the help of deep network, more detailed and high-frequency information can be recovered while maintaining the consistency of contour area, leading to better visual quality. Another benefit of this approach lies in that the proposed approach is fully compatible with all third-party image codec pipeline. Experimental result shows that the proposed scheme significantly outperforms JPEG in low bit rate image compression.**

***Index Terms*—Low Bit Rate, Image Compression, Deep Convolutional Network**

I. Introduction

Popular image compression standards such asin JPEG [1] and JPEG2000 [2] can not handle the coding performance in low bit rate image compression well. However, image compression at low bit rate has always been considered as a tough research topic due to undisciplinable quality of transmission channel and compatibility of display scalability in devices with different resolutions. Meanwhile, limited number of bits for each pixel would lead visually unacceptable compression artifacts [3], [4], [5]. A sampling-based image compression strategy has been proposed for low bit rate coding [6], [7], [8], where an image was downsampled before compression for reducing spatial redundancies among neighboring pixels which would be codec-friendly during compression. Lower resolution can offer flexibility when displaying on devices with low resolution screens, e.g. mobile devices.

However, the image quality may also suffer from the information loss during simple sampling. Therefore, extensive algorithms have been proposed for enhancing the compressed image quality, which can be divided to three categories, i.e. iteration-based, dictionary learning and deep learning methods. A wavelet inpainting driven image compression method was proposed by Zhao *et al.* [9] to overcome unsatisfactory quality of low bit rates image coding, in which a wavelet inpainting technique via collaborative sparsity was utilized by merging wavelet transform when downloading. Dictionary learning based super resolution (SR) approaches typically built upon sparse coding (SC) theory [10]. Yang et al. [11] used a SC formulation to learn low-resolution (LR) and high-resolution (HR) dictionaries by assuming that LR and HR features share the same reconstruction coefficients. More recently, deep learning has shown its power in image SR by learning hierarchical representations of high-dimensional data, and various successful applications have been observed for both low-level image processing tasks [12], [13] and high-level computer vision tasks [14].

We establish a deep convolutional network based image enhancement mechanism for low bit rate image coding in this contribution. The benefits are manifold. First, besides improving the image quality, the proposed framework can also provide essential features from the deep network for other vision applications. Second, the framework is compatible with any third-party image codec so that scalability and flexibility can be preserved. Third, comparing with iteration-based algorithms [15], the proposed method can achieve faster running speed while maintaining the same level of performance. Experimental results have shown that our framework yields significant performance improvements compared to JPEG at low bit rate conditions.

The remainder of the paper is organized as follows. Section II describes the details of the proposed framework based on the deep convolutional network. Section III shows the compression performance in terms of both objective and subjective comparisons. Section IV will draw the conclusion of this work.

II. Proposed Framework

In this section, we will first overview our framework and formulate the model as an optimization problem. Subsequently, the details of the deep network are introduced.

A. Overview

The overview diagram of our proposed framework is shown in Fig. 1. Firstly, a bicubic downsampling process is performed

978-1-5090-5316-2/16/$31.00

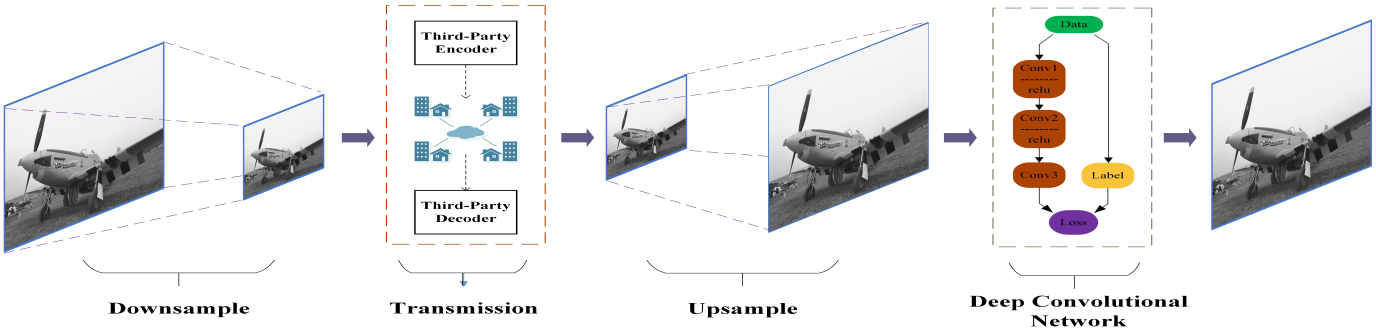


Fig. 1. Overview of the Proposed Framework.

on the original image to generate a lower resolution version. Subsequently, the downsampled image is compressed by a third-party codec, and the coded bitstream will be transmitted through the channel. Regarding the decoder side, the decoded then upsampled image will be processed by a three-layer deep convolutional network for further image quality improvement.

B. Problem Formulation

The formation is defined as follows,

$$Y = HX + N, \tag{1}$$

where X denotes the original image and Y represents the directly reconstructed low resolution image at decoder side. H indicates the downsampling operator, which is the bicubic interpolation in this work. N is the quantization and Gaussian noises introduced during compression and transmission. At the decoder side, an up-sampling process is performed to obtain the image in original resolution, which can be considered as an inverse progress of downsampling. According to (1), the restoration problem can be formulated as follows,

$$\hat{X} = \arg\min_{X} \parallel HX - Y \parallel_2^2 . \tag{2}$$

As for image enhancement using deep network, we define the learning process as follows. Given a set of ground truth images x_i and their corresponding upsampled decoded images y_i , the Mean Squared Error (MSE) is adopted as the loss function for optimizing the F process,

$$Ł(\theta) = \frac{1}{N}\sum_{i=1}^{N} \parallel F(y_i;\theta) - x_i \parallel_2^2, \tag{3}$$

where θ consists of all convolution kernel coefficient, N is the number of training examples.

C. Deep Network

In recent years, deep network have demonstrate its power in image processing. Dong et al. [16] proposed an algorithm using deep convolutional nets for super resolution and compression artifact reduction. In our framework, the decoded and upsampled image will be processed by a three-layer convolutional network, see Fig. 2. Formally, the convolution layers can be expressed as operation F in Eq. (6):

$$F_i(Y_i) = \max(0, W_i * Y_i + B_i), i = 1, 2 \tag{4}$$

$$F_3(Y_2) = W_3 * F_2(Y_2) + B_3. \tag{5}$$

$$F(Y) = F_3(F_2(F_1(Y))). \tag{6}$$

where in Eq. (4) W_i and B_i represent the filters and biases of each layer respectively, and $*$ denotes the convolution operation. Here, W_i corresponds to n_i filters of support $c \times f_i \times f_i$, where c is the number of channels of input image, f_i is the filter kernel size. That is, W_i applies n_i convolutions with each kernel size $c \times f_i \times f_i$ on the image. Therefore, the output of each layer-i is composed of n_i feature maps, and B_i is the n_i-dimensional bias vector. All filter responses are then applied by a ReLU unit [17].

D. Implementation Details

We implement our model using the latest version of *Caffe* package [18]. To generate the training and test samples, we prepare our ground truth image X_i as $f_{train} \times f_{train} \times c$ pixel sub-images, which are randomly cropped from the training image dataset [12]. To synthesize the low-resolution samples Y_i, we implemented the downsampling and upsampling process via bicubic interpolation. The *Caffe* file format we choose is $HDF5$.

As for training parameters, to avoid overfitting during training, we set the learning rate as 0.0001. For better convergence, the momentum is set as 0.9, while ignoring the weight decay term which is mainly utilized to avoid local minimum in deep learning training. The training set consists of 200 images. For better visual quality, we also add artifact reduction convolutional neural network (ARCNN) [16] as an optional stage after bicubic upsampling at the decoder side since low bit rate coding always brings heavy artifacts and bad subjective quality. During training process, no dropout is adopted due to our network only consists of several convolution layers.

The overall diagram of our deep convolutional network is shown in Fig. 2.

III. Experiment Results

This section provides experimental results including both objective and subjective quality as well as complexity analysis. Specifically, seven commonly used images are utilized for testing, including *Baby*, *Bird*, *Butterfly*, *Head*, *Woman*, Zebra and *Barbara*. The popular single image compression standard JPEG coding is adopted as the comparison method (baseline).

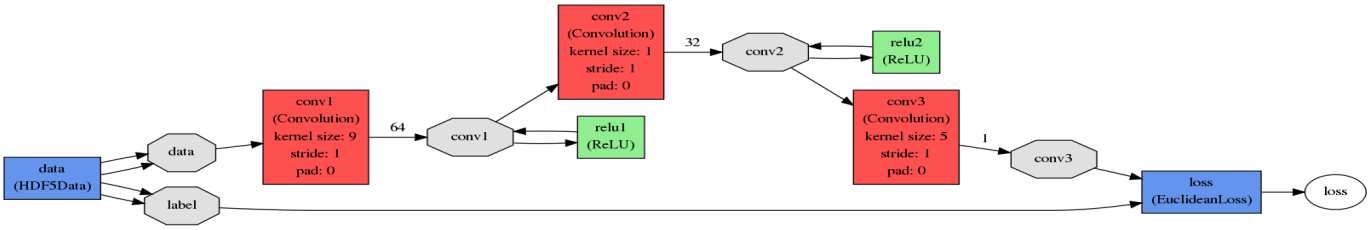
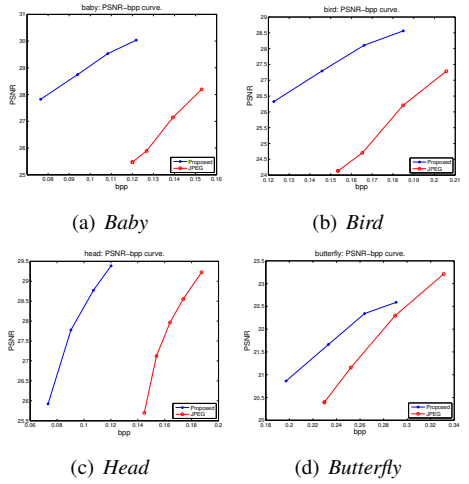


Fig. 2. Diagram of Deep Convolutional Network.

TABLE I
PSNR PERFORMANCE FOR DIFFERENT CODING RATE.

Rate(bpp)		0.15	0.20	0.25	0.30	0.35
Baby	JPEG	25.47	30.99	32.57	33.86	34.74
	Proposed	**31.69**	**32.92**	**33.98**	**34.51**	**34.97**
Bird	JPEG	24.13	27.29	29.65	31.02	32.59
	Proposed	**28.43**	**30.48**	**31.61**	**32.72**	**33.42**
Butterfly	JPEG	20.05	20.39	21.16	22.98	23.88
	Proposed	**21.07**	**22.81**	**23.27**	**23.87**	**24.57**
Head	JPEG	27.12	29.78	31.22	32.07	32.81
	Proposed	**30.81**	**31.65**	**32.21**	**32.61**	**33.02**
Woman	JPEG	23.15	24.68	26.88	28.07	29.41
	Proposed	**25.92**	**27.44**	**28.39**	**29.11**	**29.92**
Zebra	JPEG	22.32	24.08	25.69	26.89	27.36
	Proposed	**24.23**	**26.19**	**26.88**	**27.66**	**28.23**
Barbara	JPEG	22.84	25.01	26.44	27.13	27.82
	Proposed	**25.75**	**26.41**	**26.84**	**27.22**	**28.02**
Average	JPEG	23.58	26.03	27.66	28.86	29.80
	Proposed	**26.84**	**28.27**	**29.03**	**29.67**	**30.31**



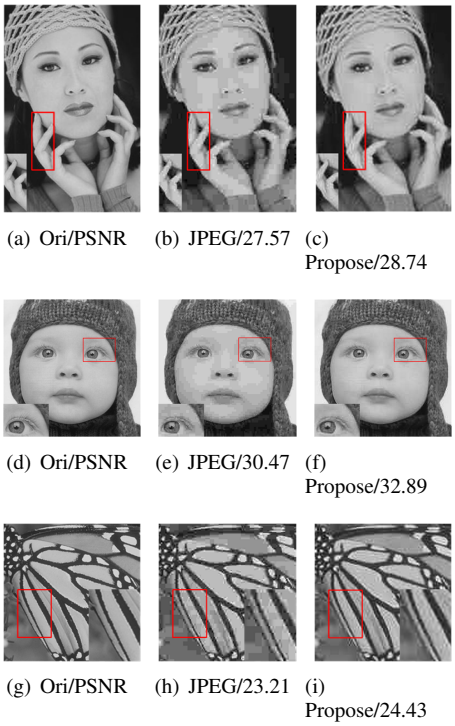
(a) *Baby* (b) *Bird*

(c) *Head* (d) *Butterfly*

Fig. 3. PSNR-bpp Curves Comparing with JPEG Coding.

A. Objective Quality

The PSNR comparisons under similar rate are reported in Table.I. At each bit rate, the best PSNR results are marked in bold for each image. The corresponding PSNR-bpp curves are also drawn in Fig. 3. We can clearly see that our proposed framework can provide significant coding gains against JPEG for all the cases under extremely low bit rates from 0.15 bpp to 0.30 bpp, even when the bit rate reaches up to 0.35 bpp.



(a) Ori/PSNR (b) JPEG/27.57 (c) Propose/28.74

(d) Ori/PSNR (e) JPEG/30.47 (f) Propose/32.89

(g) Ori/PSNR (h) JPEG/23.21 (i) Propose/24.43

Fig. 4. Visual Quality Comparisons.

B. Subjective Quality

We then exploit the subjective quality of image *woman* in Fig. 4. One can observe that the edges and contours of woman's hands can be retained by the proposed algorithm while the JPEG brings obvious ringing artifacts. Moreover, the details in strips area especially around boundaries can be also reserved, leading to better subjective quality. According to these observations, we can conclude that the proposed algorithm not only can process the pixels inside objects, but also can filter object boundary for suppressing compression artifacts, leading to more pleasing visual quality.

C. Quality Improvement

To further improve image quality, we adopt compression artifact reduction convolutional neural network (ARCNN) [16] after bicubic upsampling at the decoder side as described in Section II-D. Since the philosophy of our proposed framework mainly emphasizes on a post-processing stage of the upsampled image for better image quality. By appending

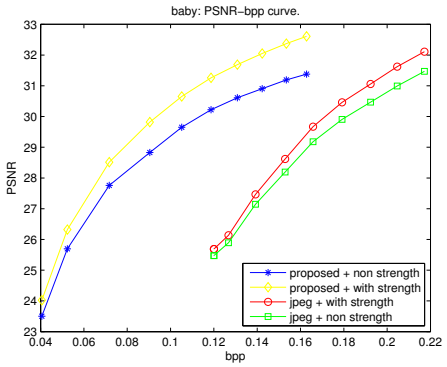


Fig. 5. PSNR-bpp Curve: Appending ARCNN Network before Upsampling at Decoder Side

the strengthen network ARCNN, the objective and subjective quality of upsampled image can be promoted to a higher level.

To evaluate the performance of the ARCNN, the PSNR-bpp curves before and after ARCNN strengthen are plotted in Fig. 5. Obviously, the ARCNN can bring further performance improvements after the first stage of deep network for both the JPEG coding and the proposed scheme.

D. Complexity Analysis

Many existing sampling-based algorithms need iteration operation for better performance which discounts their ability for practical application usage [19], [9], [20]. Our proposed CNN-based algorithm can benefit from the fast processing, while the major complexity lies in the offline training procedure in which we choose the standard version of *Caffe*. After training, we adopted the *Caffe* MATLAB wrapper to run our model. The computing environment is Windows 10 64-bit operating system with 8GB RAM. It takes 10 hours for training to get all network model parameters. Fortunately, only 2.39 seconds are used for processing a 288×344 image, indicating the proposed framework can be applied in practical applications.

IV. Conclusion

In this paper, an image enhancing scheme based on deep convolutional network is proposed for low bit rate image compression, which follows the sampling-based coding strategy. The visual quality of the decoded and upsampled image can be improved through the trained deep network. The proposed framework is also compatible with any third-party image codec so that the scalability and flexibility can be maintained. Experimental results have shown that the coding performance of the proposed scheme can be significantly improved over JPEG at low bit rate conditions.

Acknowledgment

This work was supported in part by the National Basic Research Program of China (973 Program, 2015CB351800), and National Natural Science Foundation of China (61322106, 61421062), and Shenzhen Peacock Plan, which are gratefully acknowledged.

References

[1] M. Ghanbari, *Standard codecs: image compression to advanced video coding*. Iet, 2003, no. 49.

[2] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The jpeg 2000 still image compression standard," *Signal Processing Magazine, IEEE*, vol. 18, no. 5, pp. 36–58, 2001.

[3] S. Ma, X. Zhang, J. Zhang, C. Jia, S. Wang, and W. Gao, "Nonlocal in-loop filter: The way toward next-generation video coding?" *IEEE MultiMedia*, vol. 23, no. 2, pp. 16–26, 2016.

[4] J. Zhang, R. Xiong, C. Zhao, Y. Zhang, S. Ma, and W. Gao, "Concolor: constrained non-convex low-rank model for image deblocking," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1246–1259, 2016.

[5] C. Zhao, J. Zhang, S. Ma, X. Fan, Y. Zhang, and W. Gao, "Reducing image compression artifacts by structural sparse representation and quantization constraint prior," *IEEE Transactions on Circuits Systems and Video Technology*, 2016.

[6] Y. Tsaig, M. Elad, G. H. Golub, and P. Milanfar, "Optimal framework for low bit-rate block coders," in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, vol. 2. IEEE, 2003, pp. II–219.

[7] B. Zeng and A. N. Venetsanopoulos, "A jpeg-based interpolative image coding scheme," in *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, vol. 5. IEEE, 1993, pp. 393–396.

[8] A. M. Bruckstein, M. Elad, and R. Kimmel, "Down-scaling for better transform compression," *Image Processing, IEEE Transactions on*, vol. 12, no. 9, pp. 1132–1144, 2003.

[9] C. Zhao, J. Zhang, S. Ma, and W. Gao, "Wavelet inpainting driven image compression via collaborative sparsity at low bit rates," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*. IEEE, 2013, pp. 1685–1689.

[10] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?" *Vision research*, vol. 37, no. 23, pp. 3311–3325, 1997.

[11] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[12] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 184–199.

[13] H. Tian, Y. Fang, Y. Zhao, W. Lin, R. Ni, and Z. Zhu, "Salient region detection by fusing bottom-up and top-down features extracted from a single image," *Image Processing, IEEE Transactions on*, vol. 23, no. 10, pp. 4389–4398, 2014.

[14] Y. Fang, W. Lin, Z. Chen, C.-M. Tsai, and C.-W. Lin, "A video saliency detection model in compressed domain," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 24, no. 1, pp. 27–38, 2014.

[15] L. Ma, F. Wu, D. Zhao, W. Gao, and S. Ma, "Learning-based image restoration for compressed image through neighboring embedding," in *Advances in Multimedia Information Processing-PCM 2008*. Springer, 2008, pp. 279–286.

[16] C. Dong, Y. Deng, C. Change Loy, and X. Tang, "Compression artifacts reduction by a deep convolutional network," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 576–584.

[17] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 807–814.

[18] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the ACM International Conference on Multimedia*. ACM, 2014, pp. 675–678.

[19] W. Lin and L. Dong, "Adaptive downsampling to improve image compression at low bit rates," *Image Processing, IEEE Transactions on*, vol. 15, no. 9, pp. 2513–2521, 2006.

[20] X. Wu, X. Zhang, and X. Wang, "Low bit-rate image compression via adaptive down-sampling and constrained least squares upconversion," *Image Processing, IEEE Transactions on*, vol. 18, no. 3, pp. 552–561, 2009.