Wide-activated Deep Residual Networks based Restoration for BPG-compressed Images

Yuchen Fan Jiahui Yu Thomas S. Huang

University of Illinois at Urbana-Champaign

{yuchenf4, jyu79, t-huang1}@illinois.edu

Abstract

We investigate a simple pipeline to achieve high-quality image compression under very low bit-rate. The pipeline is a stack of BPG image compression and deep network based restoration. Wide-activated deep residual networks from recent advances in image super-resolution are adopted for image restoration. Experiments demonstrate that the pipeline significantly reduces the quantity loss and remove visual artifacts for compressed images.

1. Introduction

Image compression is a task to convert images into small footprint. Lossy image compression aims a higher compression ratio while allows some distortion of decompressed images. Given specific compression ratio, lossy image compression algorithms are designed to minimize reconstruction distortion in terms of peak signal-to-noise ratio (PSNR), structural similarity (SSIM) or other metrics.

Traditionally, lossy image compression algorithms (e.g. JPEG and BPG) are based on block-wise discrete cosine transform (DCT), quantization and entropy coding. The DCT and quantization steps introduce losses and distortions (e.g. blocking, blurring and ringing). The DCT is based on the spacial continuity of image signals, however, ignores the prior distribution of photographs.

Recently, learned image compression algorithms with deep networks [2, 9, 12, 13, 17] are developed and achieve better image quality than traditional approach. Many of the deep networks based algorithms utilize deep model to transform images instead of DCT. Comparing to DCT transformed representations, the deep networks encoded features are not orthogonal between dimensions, and not discriminative for low and high frequencies, which make it more difficult to ignore high frequency redundancies to further minimize features entropy. Mentzer, *et al.* [12] proposed to use context models (PixelCNN [19]) as entropy coder and achieved better results than others.

In this work, we investigate a simple pipeline cascading BPG image compression and deep networks based image restoration. The deep networks for image restoration are supposed to learn the prior distribution of images, so that can enhance BPG-compressed images both quantitatively in term of peak signal-to-noise ratio (PSNR), and qualitatively for human perception.

2. Related Work

Deep neural networks are widely used for low-level image restoration problems [11, 20] recently. Deep networks based restoration for compressed images is firstly introduced by Dong *et al*. [4]. The work is inspired by Super-Resolution Convolutional Neural Network (SRCNN) [5], in which convolutional networks show potential in low level vision tasks. Since then, several follow-up work [3, 7, 15, 16] further improve the power of deep networks to remove artifacts.

Recently, the Enhanced Deep residual networks for Super-Resolution (EDSR) [10] achieves significant improvement for image super-resolution. In EDSR, the deep networks consist of multiple blocks with linear residuals. The residual blocks have 2 convolutional layers connected with ReLU activation.

We further improve the EDSR with wide activation SR networks (WDSR) (Fig. 1) from three aspects: wide activation, weight normalization in training and simplified global residual pathway. The WDSR are more effective in term of PSNR for image super-resolution. In this work for image restoration, we adopt the WDSR by removing up-sampling pixel-shuffle layers in the final stage.

3. Wide-activated Deep Residual Networks

In this section, we briefly review wide-activated deep residual networks and compare to its baseline EDSR [10].

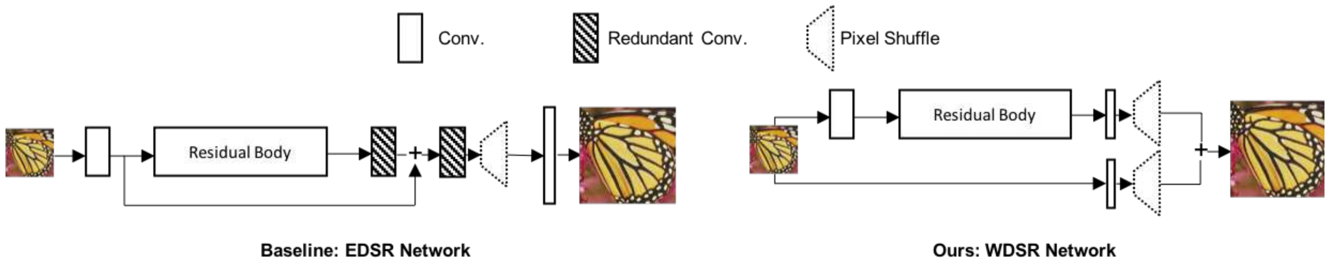

Figure 1. Our WDSR networks compared with EDSR [10].

3.1. Wide Activation

We use a deep residual network (two-layer residual blocks) following baseline EDSR [10]. Assume the width of identity mapping pathway (Fig. 1) is w_1 and width before activation inside residual block is w_2. We introduce expansion factor before activation as r, and we have $w_2 = r \times w_1$. Then the vanilla residual networks (e.g., used in EDSR and MDSR) is with $w_2 = w_1$ and have parameters $2 \times w_1^2 \times k^2$ in each residual block. The computational (Mult-Add operations) complexity is a constant scaling of parameter numbers when we fix the input patch size. To have same complexity $w_1^2 = \hat{w_1} \times \hat{w_2} = r \times \hat{w_1}^2$, the residual identity mapping pathway need to be slimmed as a factor of $\sqrt{r}$ and the activation can be expanded with $\sqrt{r}$ times. The simple idea forms our wide-activated deep residual networks. Experiments show that the proposed networks are extremely effective for improving accuracy.

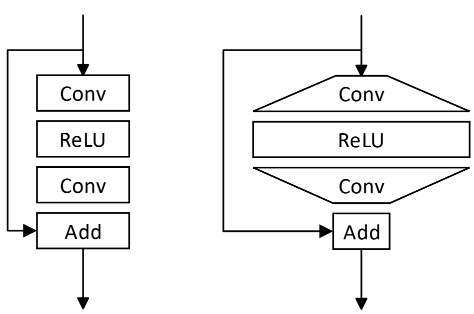

Figure 2. Under same parameter and computation complexity, wider features before ReLU activation has significantly better performance. **Left:** vanilla residual block in EDSR [10]. **Right:** residual block with wider activation in wide-activated deep residual networks.

3.2. Improved Training with Weight Normalization

We introduce weight normalization for training deep and wide restoration networks for the first time, which performs better than batch normalization or no normalization. Batch normalization re-calibrates the mean and variance of intermediate features to solve the problem of *internal covariate shift* [8] in training deep neural networks. It has different formulations in training and testing, which potentially cause

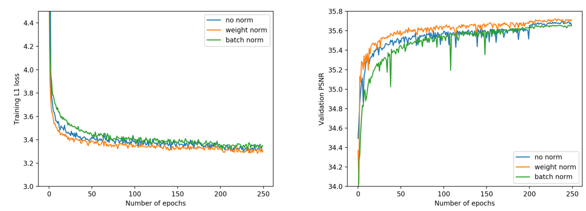

Figure 3. Training L1 loss and validation PSNR with weight normalization, batch normalization or no normalization. The proposed weight normalization yields better performance.

following problems. 1) For image super-resolution, commonly only small image patches (e.g. 48×48) and small mini-batch size (e.g. 16) are used to speedup training, thus the mean and variance of small image patches differ a lot among mini-batches, making theses statistics unstable. It is demonstrated in our experiments. 2) BN is also believed to act as a regularizer and in some cases can eliminate the need for Dropout [8]. However, it is rarely observed that regression networks overfit on training datasets. Instead, many kinds of regularizers, for examples, weight decaying and dropout, are not used in regression networks. 3) Unlike image classification tasks where softmax (scale-invariant) is used at the end of networks to make prediction, for image restoration, the different formulations of training and testing may deteriorate the accuracy for dense pixel value predictions. Thus, in recent image super-resolution networks [6, 10, 18], batch normalization is abandoned.

Weight normalization [14], on the other hand, is a reparameterization of the weight vectors in a neural network that decouples the length of those weight vectors from their direction. It does not introduce dependencies between the examples in a mini-batch, and has the same formulation in training and testing. It is also noteworthy that naively introducing weight normalization in training image restoration networks may not help that much. We find empirically that weight normalization allows higher learning rate (i.e. $10\times$), with which the loss of training normal networks explodes. The advantages of weight normalization are shown in Figure 3.

Figure 4. Qualitative comparison of restoration effects for BPG-compressed images. **Up:** BPG-compressed images. **Bottom:** adding proposed restoration.

3.3. Simplified Global Residual Pathway

We start with EDSR [10] super-resolution network. We find that the global residual pathway is a linear stack of several convolution layers, which is computational expensive. We argue that these linear convolutions are redundant (Fig. 1) and can be absorbed into residual body to some extent. Thus, we slightly modify the network structure and use single convolution layer with kernel size 5×5 that directly take $3 \times H \times W$ LR RGB image/patch as input and output $3S^2 \times H \times W$ HR counterparts, where S is the scale. This results in less parameters and computation. In our experiments we have not found any accuracy drop with our simpler form.

4. Experimental Results

4.1. BPG Image Compression

For BPG image compression, we use the libbpg implementation. To meet the requirements of 0.15 bits per pixel (bpp), we use the JCT-VC encoder with quantizer parameter at 38 and convert images in YCoCg color space with bit depth at 9.

4.2. Networks Training

4.2.1 Dataset

The image restoration networks are trained on DIV2K dataset [1]. The input images are reconstructed from BPG compressed ones. The ground-truth targets are original uncompressed images.

4.2.2 Networks Structure

To compromise between speed and accuracy, the networks are designed to have 16 residual blocks, with 32 nodes in identity mapping pathway and 4x expansion for wide activation (128 nodes). The networks input is RGB images in 3 channels normalized to 0-1 and subtracted the global mean of each channel, and the output is also corresponding to RGB channels and de-normalized with inverse process.

4.2.3 Hyper parameters

The restoration networks are trained based on square image patches whose height and width are 224. The training objective function is L1 loss, and optimized with a fixed learning rate of 1e-3.

4.3. Results

The proposed approach is evaluated on the validation dataset in challenge on learned image compression (CLIC) 2018.

4.3.1 Quantitative Results

Compared with BPG compression, in Table 1. , the proposed restoration can significantly improve PSNR for reconstructed images.

Method	PSNR (dB)
BPG	31.09
+ Restoration	31.71

Table 1. Comparison of restoration effects for BPG-compressed images

4.3.2 Qualitative Results

In Figure 4, some representative patches are selected to demonstrate improvements over BPG-compressed images. The proposed approach can remarkably remove the artifacts for BPG-compressed images and provide more pleasing photographs.

5. Conclusion

We successfully adopt the recent advances of wide-activated deep residual networks in image super-resolution. The deep networks can remarkably restore compressed images to reduce the loss and remove artifacts in image compression.

References

[1] E. Agustsson and R. Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, volume 3, page 2, 2017.

[2] J. Ballé, V. Laparra, and E. P. Simoncelli. End-to-end optimized image compression. *arXiv preprint arXiv:1611.01704*, 2016.

[3] L. Cavigelli, P. Hager, and L. Benini. Cas-cnn: A deep convolutional neural network for image compression artifact suppression. In *Neural Networks (IJCNN), 2017 International Joint Conference on*, pages 752–759. IEEE, 2017.

[4] 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*, pages 576–584, 2015.

[5] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016.

[6] Y. Fan, H. Shi, J. Yu, D. Liu, W. Han, H. Yu, Z. Wang, X. Wang, and T. S. Huang. Balanced two-stage residual networks for image super-resolution. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, pages 1157–1164. IEEE, 2017.

[7] L. Galteri, L. Seidenari, M. Bertini, and A. Del Bimbo. Deep generative adversarial compression artifact removal.

[8] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[9] M. Li, W. Zuo, S. Gu, D. Zhao, and D. Zhang. Learning convolutional networks for content-weighted image compression. *arXiv preprint arXiv:1703.10553*, 2017.

[10] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 136–144, 2017.

[11] D. Liu, Z. Wang, Y. Fan, X. Liu, Z. Wang, S. Chang, and T. Huang. Robust video super-resolution with learned temporal dynamics. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 2526–2534. IEEE, 2017.

[12] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. Van Gool. Conditional probability models for deep image compression. *arXiv preprint arXiv:1801.04260*, 2018.

[13] O. Rippel and L. Bourdev. Real-time adaptive image compression. In *International Conference on Machine Learning*, 2017.

[14] T. Salimans and D. P. Kingma. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In *Advances in Neural Information Processing Systems*, pages 901–909, 2016.

[15] P. Svoboda, M. Hradis, D. Barina, and P. Zemcik. Compression artifacts removal using convolutional neural networks. *arXiv preprint arXiv:1605.00366*, 2016.

[16] Y. Tai, J. Yang, X. Liu, and C. Xu. Memnet: A persistent memory network for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4539–4547, 2017.

[17] L. Theis, W. Shi, A. Cunningham, and F. Huszár. Lossy image compression with compressive autoencoders. *arXiv preprint arXiv:1703.00395*, 2017.

[18] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, L. Zhang, B. Lim, S. Son, H. Kim, S. Nah, K. M. Lee, et al. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, pages 1110–1121. IEEE, 2017.

[19] A. van den Oord, N. Kalchbrenner, L. Espeholt, O. Vinyals, A. Graves, et al. Conditional image generation with pixelcnn decoders. In *Advances in Neural Information Processing Systems*, pages 4790–4798, 2016.

[20] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang. Generative image inpainting with contextual attention. *arXiv preprint arXiv:1801.07892*, 2018.