Analysis of Directional Distortion on Stripe Regions in Deep Learning Image Coding Based on MS-SSIM

Shahid Abbas, Oh-Jin Kwon*

Department of Electronics Engineering, Sejong University, Seoul, Republic of Korea

shahid@sju.ac.kr, ojkwon@sejong.ac.kr

Abstract—Compression is the most promising and influential research area, especially for images and videos. Many essential well-known traditional compression standards have been developed for many decades. However, recently it is a very crucial and challenging topic in the field of machine and deep learning. Furthermore, most of the researchers focus on AI techniques targeting to achieve very good quality and compression than traditional compression standards. Although, yet more gap exists in demanding. This paper investigates a directionality distortion that may occur on stripped area especially:

I. In the MS-SSIM optimized model.

II. At lower Bitrate (less than 0.5bpp)

To address the critical problem of directionality as shown in Fig. 2., simulation of different AI-based image coding models (optimized by MS-SSIM) is done.

Index Terms **Image Coding, AI-based Image Coding, Directionality, MS-SSIM, Lower Bitrate**

1 Introduction

Image Compression is fundamental and yet very challenging dilemma for the scientific community in the field of signal processing to reduce its size for efficient transmission and storage. JPEG [1] and JPEG2000 [2] are the classical image compression standard. However, nowadays with rapid development of computing power and large data collection technique, Deep convolutional neural networks (CNN) [3, 4], attracted much attention due to their great potential in image compression and powerful representational capacity. CNN can extract features and decompose images due to its hierarchical structure, which contains multiple blocks of convolutional layers as shown in **Fig. 1.** During the decomposition of the image, the dimension is reduced gradually, which is the main role of encoder and makes the data more compact for further compression.

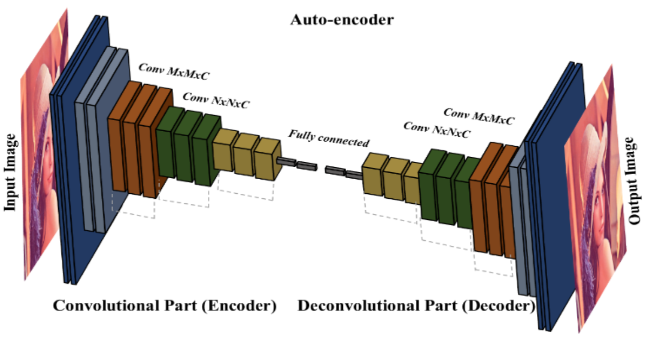


Fig. 1. General Framework of Deep Convolutional Neural Network

Most of the deep image compression models are trained iteratively using cost function of mean square error (MSE) [5] and MS-SSIM [6].

In this paper, we are going to demonstrate a critical directionality problem which may occur in the stripped area of the image. To be specific, the distortion has been found in those models which are optimized using cost function of MS-SSIM matrix and exists mostly for lower bitrate (less than 0.5bpp).

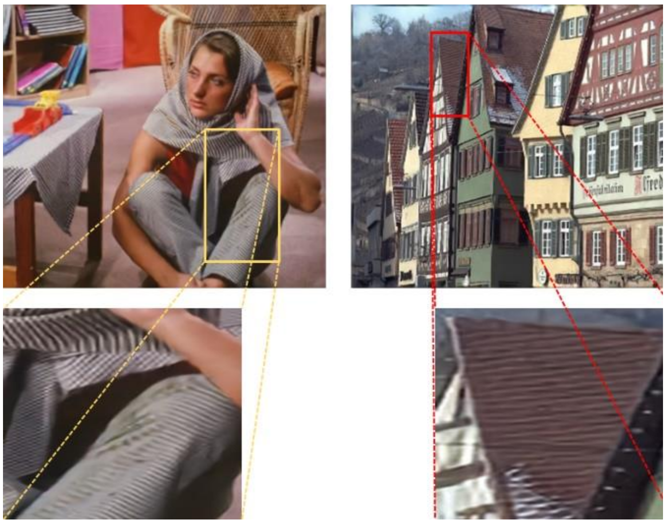


Fig. 2. Directionality distortion in Barbara and House image

2 Related Work

To compress images, classical image coding standards mainly focus to reduce redundancy in the image. Although image coding standards achieve high compression ratio, deep learning codecs attract much attention due to their great potential of image compression. For lossless image compression, deep learning models have achieved state-of-the-art performance [7-9]. For lossy image compression Lucas Theis et al. [10], propose a new approach to the problem of optimizing the auto-encoders. Toderici et al. [11] present a recurrent neural network (RNN) to compress 32x32 images. Furthermore, Toderici et al. [12] introduce a set of full-resolution compression methods for progressive encoding and decoding of images.

The networks are typically trained to minimize the MSE or using perceptual metrics such as MS-SSIM between the original and reconstructed image. Despite of using MS-SSIM cost function, to the best of authors knowledge, no one reported a crucial problem which occurs mostly at low bit rates. So, in this paper we describe the directionality distortion which may occur in the stripped area of the image.

3 Experimental Setup

3.1 Papers Considered

The papers which provide their source code are analyzed for directionality problem, which is the following.

i. End-to-end optimized image compression [13]

ii. Variational image compression with a scale hyperprior [14]

iii. Joint autoregressive and hierarchical priors for learned image compression [15]

iv. Context-adaptive entropy model for end-to-end optimized image compression [16]

3.2 Training Details

For training of [14, 15] models, they used JPEG color images having heights/width between 3000 and 5000 pixels, comprising approximately 1 million images scraped from the world wide web. The images were further down-sampled with height and width equaled between 640 and 1200 pixels. Then randomly 256 x 256 pixels were extracted from these down-sampled images. They optimized their model using Adam [17] algorithm with a minibatch size of 8 and learning rate of 10^{-4}. We use their pre-trained models for our experiment.

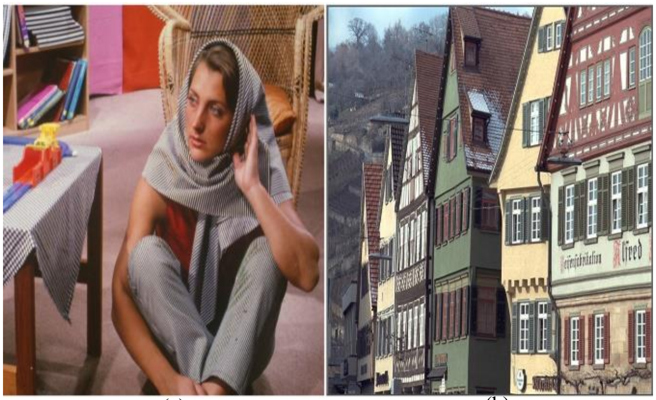
In case of [13], we train the respective model using default configuration parameters by using JPEG AI dataset [18], the datasets consist of training, validation, and test dataset individually. However, for training the model, we use training dataset in which all the images are in PNG format with spatial resolution from 256×256 to 8K.

For model [16] training, they used randomly selected images from the YFCC100m [19] with patch size of 256×256 and each batch consists of eight images. While training, they optimized the model using Adam optimizer for 1M iterations of the training steps with the initial learning rate of 5×10^{-5}, in which the rate is reduced by half for every 50,000 iterations for the last 200,000 iterations.

3.3 Experimental Results

For our experimental results, we used two images, Barbara and House image shown in **Fig. 3.**

To generate the results of the models [14-16], we used pre-trained model provided by their respective authors. However, in case of model [13], we trained the model using default configuration. The results for our experiments are shown in **Fig. 4.**



(a) (b)

Fig. 3. Original Image of (a) Barbara and (b) House

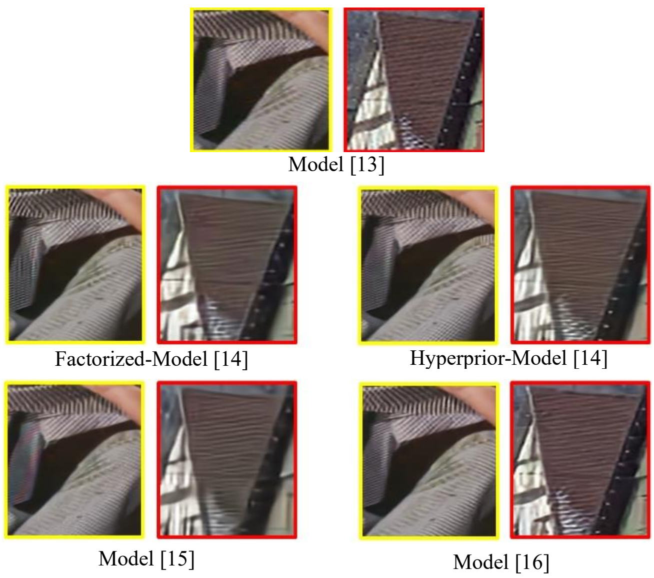


Fig. 4. Results of ROIs for their respective model

4 Conclusion & Future Direction

In this paper, we investigated different AI based image coding techniques for the purpose to address one of the critical directionality problems, which may occur in the stripped area of the image. As all the model networks usually use two different types of distortion metrices: MSE and MS-SSIM to minimize the loss between the original and reconstructed image, accordingly, this kind of distortion is found mostly in the model using MS-SSIM loss function for the optimization and, to be more specific, at lower bitrates (less than 0.5bpp).

In the future, we need to find a novel solution for this kind of problem in AI based image coding by investigating MS-SSIM matric. As the problem may occur in stripped area in the MS-SSIM optimized models regardless of the training datasets, it is important to consider this distortion while evaluating the performance of AI-based image coding.

Acknowledgements

This work was supported by the Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.2020-0-00347, Development of JPEG Systems standard for snack culture contents).

References

[1] G. K. Wallace, "The JPEG still picture compression standard," *IEEE transactions on consumer electronics,* vol. 38, no. 1, pp. xviii-xxxiv, 1992.

[2] M. Rabbani and R. Joshi, "An overview of the JPEG 2000 still image compression standard," *Signal processing: Image communication,* vol. 17, no. 1, pp. 3-48, 2002.

[3] N. Ketkar, "Convolutional neural networks," in *Deep Learning with Python*: Springer, 2017, pp. 63-78.

[4] H. H. Aghdam and E. J. Heravi, "Guide to convolutional neural networks," *New York, NY: Springer,* vol. 10, pp. 978-973, 2017.

[5] M. Binieli. "an introduction to mean squared error and regression lines." https://www.freecodecamp.org/news/machine-learning-mean-squared-error-regression-line-c7dde9a26b93/ (accessed.

[6] W. Malpica and A. Bovik, "SSIM based range image quality assessment," in *4th Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2009: Citeseer.

[7] A. Van Oord, N. Kalchbrenner, and K. Kavukcuoglu, "Pixel recurrent neural networks," in *International Conference on Machine Learning*, 2016: PMLR, pp. 1747-1756.

[8] O. Rippel and L. Bourdev, "Real-time adaptive image compression," in *International Conference on Machine Learning*, 2017: PMLR, pp. 2922-2930.

[9] L. Theis and M. Bethge, "Generative image modeling using spatial lstms," *arXiv preprint arXiv:1506.03478,* 2015.

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

[11] G. Toderici *et al.*, "Variable rate image compression with recurrent neural networks," *arXiv preprint arXiv:1511.06085,* 2015.

[12] G. Toderici *et al.*, "Full resolution image compression with recurrent neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5306-5314.

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

[14] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, "Variational image compression with a scale hyperprior," *arXiv preprint arXiv:1802.01436,* 2018.

[15] D. Minnen, J. Ballé, and G. Toderici, "Joint autoregressive and hierarchical priors for learned image compression," *arXiv preprint arXiv:1809.02736,* 2018.

[16] J. Lee, S. Cho, and S.-K. Beack, "Context-adaptive entropy model for end-to-end optimized image compression," *arXiv preprint arXiv:1809.10452,* 2018.

[17] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.

[18] J. Ascenso and P. Akayzi, "JPEG AI image coding common test conditions," *ISO/IEC JTC1/SC29/WG1 N,* vol. 84035, 2019.

[19] B. Thomee *et al.*, "YFCC100M: The new data in multimedia research," *Communications of the ACM,* vol. 59, no. 2, pp. 64-73, 2016.