Learned Image Compression with Wavelet Preprocessing for Low Bit Rates

Sofia Iliopoulou [1], Panagiotis Tsinganos [1], Dimitris Ampeliotis [2] and Athanassios Skodras [1]
[1] Department of Electrical and Computer Engineering, University of Patras, Patras, Greece
[2] Department of Digital Media and Communication, Ionian University, Argostoli, Greece
sofia_iliopoulou@upnet.gr, {panagiotis.tsinganos, skodras}@ece.upatras.gr, ampeliot@ceid.upatras.gr

***Abstract*—Deep Learning has revolutionized the field of image processing and image compression in particular. A lot of research has been done in recent years on the subject of learning-based image compression, which has resulted in methods with increased compression performance but high computational complexity. The most successful methods eradicate the redundancies by using entropy modelling. In this paper, we utilize the Discrete Wavelet Transform (DWT) as a preprocessing step for a simple hyperprior model. The proposed method is compared to both traditional and deep learning-based techniques. It proves to be superior in lower bit rates, using both Peak Signal-to-Noise Ratio (PSNR) and Multi-Scale Structural Similarity (MS-SSIM) as metrics for the evaluation of the model.**

Keywords—image compression, hypeprior, Discrete Wavelet Transform, deep learning, entropy modelling, low bit rate

I. Introduction

Image compression is an area of signal and image processing that has been continuously researched in recent years due to its use in efficient storage and transmission of image data. In the past decades, a number of standards has been developed such as JPEG [1], JPEG2000 [2], BPG [3] etc. However, the rapid rise of artificial intelligence has led to the incorporation of machine and deep learning methods into this subject.

The traditional codecs remove redundancies from the input image data mainly by using transform coding followed by quantization, and entropy coding. For example, JPEG and BPG utilize the Discrete Cosine Transform (DCT), while JPEG2000 uses the Discrete Wavelet Transform (DWT) [4]. Many deep learning techniques tend to follow the same sequence, replacing the transform with neural networks, in order to jointly optimize these three steps [5].

Over the past years, deep learning models have shown great success to the problem of image compression. In these approaches, various types of neural networks have been utilized to create a more compact representation of the input data, such as Convolutional Neural Networks (CNNs), Autoencoders (AE), Recurrent Neural Networks (RNNs) and Generative Adversarial Networks (GANs) [6-10]. In end-to-end image compression, the compressed feature map also keeps certain spatial correlation due to the limited receptive field of convolutions [10]. Thus, entropy modelling is used for the elimination of the remaining redundancies. Entropy modelling that aims to estimate the rate of the codes plays a vital role in learned image compression methods. According to Shannon's source coding theorem [10], given a sequence of codes $y = \{y_0,..., y_N\}$, the optimal code length of y should be

$$C = E_y [- \sum^{N}_{i=0} log_2 P(y_i)] \quad (1)$$

Thus, estimating accurate probability distribution functions, i.e., $P(y_i)$, for the codes is essential in determining the compression rate [11]. End-to-end image compression methods with a variational autoencoder (VAE) are popular in this field, which introduce a hyperprior [12] model to transmit the distribution of latent representation.

In this work, we investigate the application of DWT as a preprocessing step to an autoencoder with a hyperprior, for low bit rates. The main contributions presented in this paper are:

- the use of a different representation method for the images when given as inputs to the autoencoder,
- the application of the wavelet transform as a preprocessing step for entropy-modelling techniques.

The rest of the paper is organized as follows. In Section II a review of deep learning-based image compression methods is provided. Section III demonstrates the details of the proposed techniques and the network architectures used in this work. The experiments performed for the evaluation of the model and their results are shown in Section IV. Finally, Section V summarizes the outcomes and outlines future work.

II. Related Work

Many deep learning-based image compression methods follow the logic of traditional codecs which dictates that a transform is followed by quantization and entropy coding, substituting one of the steps by a deep learning algorithm. These techniques assume that all the codes are independent and identically distributed. They also suppose that all of them follow the same probability distribution, in order to have easier-to-handle entropy models [11]. The author of [13] uses

979-8-3503-3959-8/23/$31.00 ©2023 IEEE

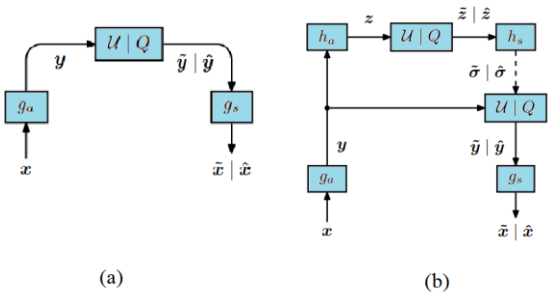

Fig. 1. Operational diagram of the baseline factorized prior (a) and the hyperprior (b). The boxes represent transformations of the data. The boxes U|Q indicate either the addition of uniform noise during the training of the model or quantization and arithmetic coding during testing.

trellis coded quantization in lieu of a more basic form.

However, the most commonly replaced step is the transform. Toderici et al. [14] created an RNN to compress 32×32 images in a progressive manner and they later extended their research to bigger images by introducing a BinaryRNN for context-based entropy modelling [7].

Ballé et al. [8] started off by replacing the typical rectified linear unit function (ReLU) with a generalized divisive normalization (GDN) activation function in an end-to-end manner and modelled the entropy with a factorized prior. In their following work [12], they assume a zero-mean Gaussian distribution for each code, with the deviation estimated by a side information network depending on the hierarchical hyperprior. This method has become a benchmark and base for many subsequent learned image compression techniques by other researchers that utilize entropy modelling [15-18].

Minnen et al. [15] combined the hierarchical hyperprior with a context-based autoregressive prior to enhance the results of the compression. Lee et al. [16] also introduced an autoregressive component to the entropy model. Taking advantage of the high correlation of local dependency, context-adaptive models contribute to a more accurate entropy estimation. However, since these models only capture the spatial information of neighboring latents, there is remaining redundant spatial information across the entire image. Qian et al. [19] built on the architecture introduced by Minnen and proposed the combination of a local context, a global reference and a hyperprior model to overcome this issue and further boost the compression performance.

III. Proposed Image Compression Methods

A. Rate-Distortion Optimization

The goal of our model is to minimize the expected length of the compressed bitstream that results from our processing as well as the expected distortion of the reconstructed image with respect to the original, creating the rate–distortion optimization problem that dominates learned image compression:

$$R+\lambda \cdot D = E_{x\sim px}[-log_2 p_{\hat{y}}(q[f(x)])]+\lambda \cdot E_{x\sim px}[d(x,g(q[f(x)]))] \quad (2)$$

where λ is the Lagrange multiplier that determines the desired rate (R) – distortion (D) trade-off, p_x is the unknown distribution of natural images, q represents rounding to the nearest integer (uniform quantization), $y = f(x)$ is the encoder, $\hat{y} = q[y]$ are the quantized latents and $p_{\hat{y}}$ is a discrete entropy model [19]. The rate is the expected code length (bit rate) of the compact representation of the image, and it can be written as the cross entropy between the marginal distribution of the latents and the learned entropy model [12]. The distortion is the expected difference between the reconstructed and the original image, as measured by a norm or perceptual metric function, e.g., the mean squared error (MSE) or the multi-scale structural similarity (MS-SSIM) [20]. The optimization problem can be represented as a variational autoencoder. Fig. 1 shows the operational diagrams of (a) the factorized prior and (b) the hyperprior models. Our technique is based on the hyperprior method [12].

B. Network Architecture

The architecture of our system follows closely that of Ballé [12]. One crucial difference is that in our paper we suggest a different representation of the input images before they are given to the network. First, we split them into the three color-channels (Red-Green-Blue). Our method proposes the use of the Discrete Wavelet Transform as a preprocessing step, after the color-channel splitting. Each of the three new images goes through the DWT and produces four sub-images which correspond to the approximation, horizontal, vertical, and diagonal detail, respectively. Thus, for every image in the original dataset we end up with 12 sub-images that are stacked together to create a 12-channel image which is given as input to the neural network. The chosen wavelet type is the biorthogonal 4.4 wavelet, which is similar to the 9/7 wavelet and is considered suitable for lossy image compression [4]. The proposed framework is illustrated in Fig. 2 and the individual network layers are shown in Table I.

We set up the main and the hyper autoencoder as a series of linear (convolution/deconvolution) and non-linear (GDN/IGDN and ReLU) functions. Similar to other VAE-based image compression methods, we used a learned encoder g_a to map the input image x to a latent representation y, which is further grouped and rearranged as z by the hyper encoder h_a. Both outputs are quantized into discrete values and then given to a lossless arithmetic coder. Finally, a decoder g_s and a hyper decoder h_s transform $\hat{y}$ and $\hat{z}$ to the reconstructed image $\hat{x}$.

TABLE I. NEURAL NETWORK LAYERS

Encoder	*Decoder*	*Hyper Encoder*	*Hyper Decoder*
Conv: 5×5 c192 s2 GDN Conv: 5×5 c192 s2 GDN Conv: 5×5 c192 s2 GDN Conv: 5×5 c12 s2	Deconv: 5×5 c192 s2 IGDN Deconv: 5×5 c192 s2 IGDN Deconv: 5×5 c192 s2 IGDN Deconv: 5×5 c12 s2	Conv: 3×3 c192 s1 ReLU Conv: 5×5 c192 s2 ReLU Conv: 5×5 c192 s2	Deconv: 5×5 c192 s2 ReLU Deconv: 5×5 c192 s2 ReLU Deconv: 3×3 c192 s1
Conv: convolutional layer, Deconv: deconvolutional layer, c:channels, s:stride			

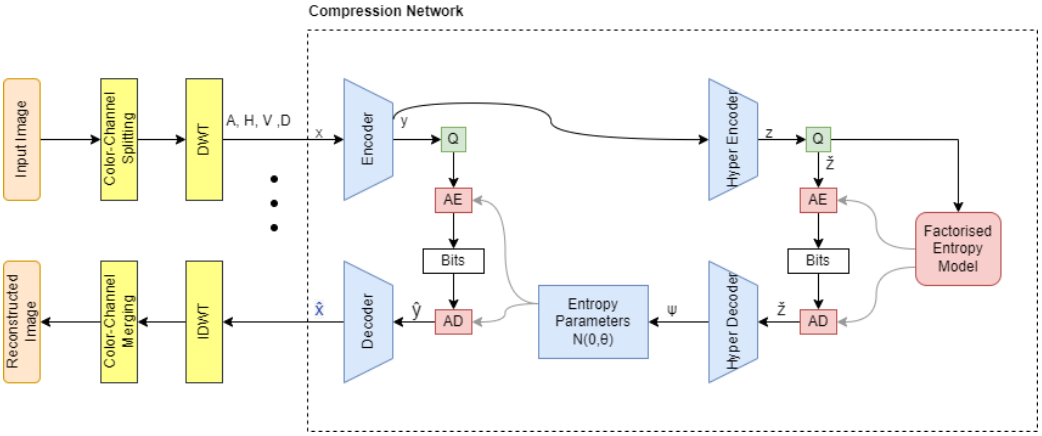

Fig. 2. The architecture of our method. Q denotes quantization and AE/AD is the lossless arithmetic encoder/decoder

IV. EXPERIMENTAL RESULTS

To compare the compression performance of our proposed model, we conducted several experiments using the Tensorflow framework. The proposed models are trained on a subset of the ImageNet dataset [21], as well as the CLIC dataset, which consists of 1258 images.

Afterwards, randomly placed 504×504 pixel crops of these images were extracted. The dimensions of these crops were chosen, so that the 12 sub-images produced by the DWT decomposition had a size of 256×256 pixels. These sub-images were stacked together to create the inputs of the

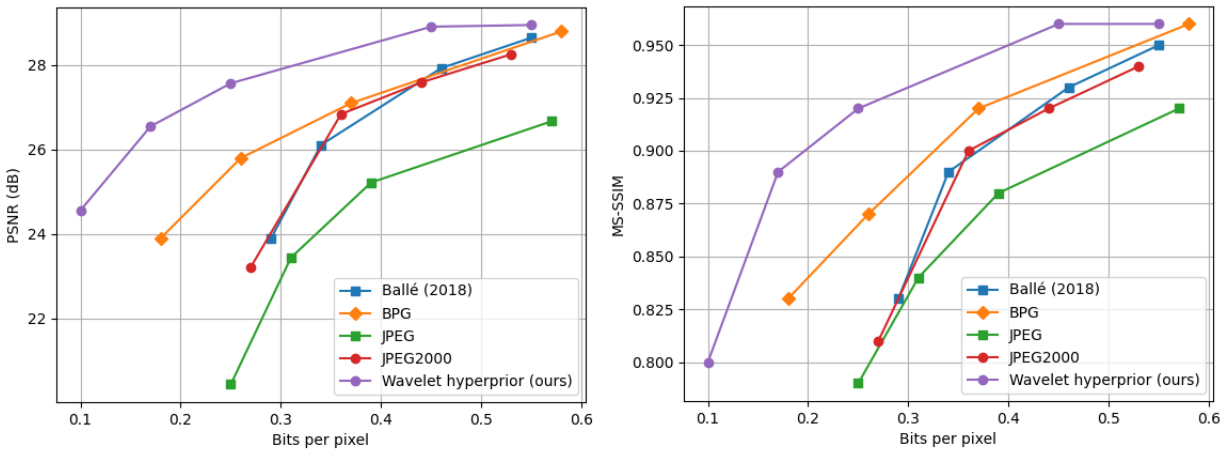

Fig. 3. Model evaluation over the Kodak dataset

TABLE II. COMPARISON OF RESULTS

Method	Results		
	Bit rate	*PSNR (dB)*	*MS-SSIM*
Ballé (2018)	0.39	26.85	0.91
BPG	0.37	27.11	0.92
JPEG	0.36	24.67	0.86
JPEG2000	0.37	26.93	0.91
Wavelet hyperprior (ours)	**0.35**	**28.42**	**0.94**

compression network. Minibatches of 8 of the input images at a time were used to perform stochastic gradient descent using the Adam optimization algorithm with a learning rate of 10^{-4}. With this setup, we trained 5 separate models. The mean squared error (MSE) was used as the distortion metric in all of them and 5 different values of λ (0.001, 0.005, 0.01, 0.05 and 0.1) were used, in order to produce a wide range of rate-distortion tradeoffs.

The evaluation of the proposed models was performed on the publicly available Kodak dataset [22]. The rate-distortion curves are shown in Fig.3. The distortion metrics that were used in this paper are the PSNR and the MS-SSIM.

In the experiments, we investigate a way of improving the results of the classic hyperprior image compression algorithm for low bit rates, by using the DWT decomposition of the inputs. The average results of the evaluation experiments are presented in Table II and an example image is shown in Fig.4.

As we can see, the wavelet hyperprior method has much better results than JPEG, and even surpasses other traditional techniques such as JPEG2000 and BPG. It also outperforms Ballé's method, which was the base upon which our method was built. This proves that the utilization of wavelets as a preprocessing step is extremely advantageous for low bit rates. Additionally, another benefit of our method is its low computational complexity. The average test runtime for Ballé on GPU was 670.14ms, while for the wavelet hyperprior was 107.46ms.

V. CONCLUSIONS

This work presented a new method of learned image compression for lower bit rates, based on the hyperprior model. The main contribution of our paper was the use of wavelets as a preprocessing step for the entropy modelling method. Our approach surpassed both traditional codecs such as JPEG and BPG, and other learned compression techniques. An added benefit of this methodology was its low computational complexity and test runtime. The addition of the DWT resulted in smaller bit rates than the simple hyperprior for the same value of λ and improved the performance of the neural networks.

Our next step is the further improvement of the results via the utilization of additional entropy models, as well the use of other metric functions for the distortion, i.e., the MS-SSIM. More experiments will be necessary to drastically improve the compression results. This technique is targeted towards low bit rates and does not perform as well as other learned image compression techniques for higher bit rates. It is our aim to implement a more general method as well.

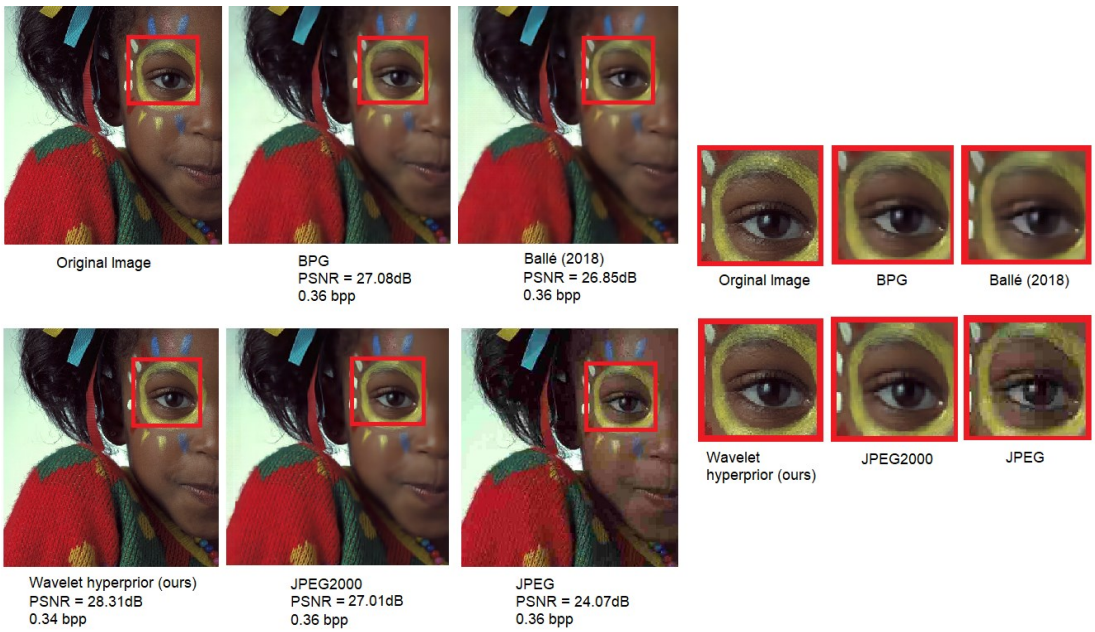

Fig. 4. At similar bit rates, our wavelet hyperprior method provides the highest visual quality on the Kodak dataset

REFERENCES

[1] G. K. Wallace, “The jpeg still picture compression standard,” IEEE Transactions on Consumer Electronics, vol. 38, 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] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, “Overview of the high efficiency video coding (hevc) standard,” IEEE Transactions on circuits and systems for video technology, vol. 22, no. 12, pp. 1649-1668, 2012.

[4] A. Skodras, C. Christopoulos and T. Ebrahimi, "The JPEG 2000 still image compression standard," in IEEE Signal Processing Magazine, vol. 18, no. 5, pp. 36-58, Sept. 2001, doi: 10.1109/79.952804.

[5] L. Yuan, J. Luo, S. Li, W. Dai, C. Li, J. Zou and H. Xiong, "Learned Image Compression with Channel-Wise Grouped Context Modeling," 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 2021, pp. 2099-2103, doi: 10.1109/ICIP42928.2021.9506076.

[6] O. Rippel and L. Bourdev, “Real-Time Adaptive Image Compression,” in Proc. 34th Int. Conf. Mach. Learn., Sydney, NSW, Australia, Aug. 2017, pp. 2922-2930.

[7] G. Toderici, D. Vincent, N. Johnston, S. Hwang, D. Minnen, J. Shor and M Covell, “Full resolution image compression with recurrent neural networks,” in 2017 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, July 2017, pp. 5435-5443.

[8] J. Balle, V. Laparra, and E. P. Simoncelli, “End-to-end optimized image compression,” in 5th Int. Conf. Learn. Rep., Toulon, France, Apr. 2017.

[9] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. V. Gool, “Conditional probability models for deep image compression,” in 2018 IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Salt Lake City, UT, USA, June 2018, pp. 4394-4402.

[10] T.M Cover and J.A. Thomas, “Data Compression” in “Elements of Information Theory”, 2nd ed., John Wiley & Sons, 2006, pp. 103–142.

[11] M. Li, K. Zhang, J. Li, W. Zuo, R. Timofte and D. Zhang, "Learning Context-Based Nonlocal Entropy Modeling for Image Compression," in IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 3, pp. 1132-1145, March 2023, doi: 10.1109/TNNLS.2021.3104974.

[12] J. Balle, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, “Variational image compression with a scale hyperprior,” in 6th Int. Conf. Learn. Rep., Vancouver, BC, Canada, Apr. 2018.

[13] B. Li, M. Akbari, J. Liang and Y. Wang, "Deep Learning-Based Image Compression with Trellis Coded Quantization," 2020 Data Compression Conference (DCC), Snowbird, UT, USA, 2020, pp. 13-22, doi: 10.1109/DCC47342.2020.00009.

[14] G. Toderici, S. O’Malley, S. Hwang, D. Vincent, D. Minnen., S. Baluga, M Covell and R. Sukthankar, “Variable rate image compression with recurrent neural networks”, 2015, arXiv:1511.06085. [Online]. Available: http://arxiv.org/abs/1511.06085

[15] D. Minnen, J. Balle, and G. D. Toderici, “Joint autoregressive and hierarchical priors for learned image compression,” in Proc. Neural Inf. Process. Syst., 2018, pp. 10794–10803.

[16] J. Lee, S. Cho, S-K Beack, “Context-Adaptive Entropy Model for End-to-end optimized Image Compression”, Intl. Conf. on Learning Representations (ICLR) 2019.

[17] T. Chen, H. Liu, Z. Ma, Q. Shen, X. Cao and Y. Wang, "End-to-End Learnt Image Compression via Non-Local Attention Optimization and Improved Context Modeling," in IEEE Transactions on Image Processing, vol. 30, pp. 3179-3191, 2021, doi: 10.1109/TIP.2021.3058615.

[18] Z. Cheng, H. Sun, M. Takeuchi and J. Katto, “Learned image compression with discretized Gaussian mixture likelihoods and attention modules,” in 2020 IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Seattle, WA, USA, June 2020, pp. 7939-7948.

[19] Y. Qian, Z. Tan, X. Sun, M. Lin, D. Li, Z. Sun, H. Li and R. Jin. “Learning accurate entropy model with global reference for image compression”, Intl. Conf. on Learning Representations (ICLR) 2021.

[20] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multiscale structural similarity for image quality assessment," The 37th Asilomar Conference on Signals, Systems & Computers, 2003, Pacific Grove, CA, USA, 2003, pp. 1398-1402 Vol.2, doi: 10.1109/ACSSC.2003.1292216.

[21] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, L. Fei-Fei., “Imagenet: A large-scale hierarchical image database”, In 2009 IEEE conference on computer vision and pattern recognition, 2009. p. 248–55.

[22] Eastman Kodak, “Kodak Lossless True Color Image Suite (PhotoCD PCD0992)”, URL: http://r0k.us/graphics/kodak/, 1993.