

Learned Image Compression with Discretized Gaussian Mixture Likelihoods and Attention Modules

Dr. G. Ranganathan,

Head of the department,
Department of Electronics & Communication Engineering,
Gnanamani College of Technology,
Namakkal, India.

Dr. Bindhu V,

Professor and Head,
Department of Electronics and Communication Engineering,
PPG Institute of Technology,
Villankurichi Saravanampatti P.O.,
Coimbatore, India.

Abstract: There have been many compression standards developed during the past few decades and technological advances has resulted in introducing many methodologies with promising results. As far as PSNR metric is concerned, there is a performance gap between reigning compression standards and learned compression algorithms. Based on research, we experimented using an accurate entropy model on the learned compression algorithms to determine the rate-distortion performance. In this paper, discretized Gaussian Mixture likelihood is proposed to determine the latent code parameters in order to attain a more flexible and accurate model of entropy. Moreover, we have also enhanced the performance of the work by introducing recent attention modules in the network architecture. Simulation results indicate that when compared with the previously existing techniques using high-resolution and Kodak datasets, the proposed work achieves a higher rate of performance. When MS-SSIM is used for optimization, our work generates a more visually pleasant image.

Keywords: PSNR, discretized Gaussian Mixture, rate-distortion performance, image compression, network architecture

1. Introduction

One of the most predominantly researched topic in signal processing is image compression. Numerous methods have been formulated and designed over the decades to attain efficient image transmission. Some of the commonly used compression standards are HEVC/H.265, JPEG2000, JPEG and VVC. These standards are based on hand-crafted creativity using Kodak dataset with the help of codec at 0.1 bpp [1]. Context adaptive arithmetic coders, quantization, intra prediction, fixed transform matrix and other loop or deblocking filters were used to improve coding efficiency and decrease spatial redundancy. Over the years, there has been much effort put in standardizing the conventional codec. In addition to the quick development in proliferation of high resolution cameras in hand-held devices and new image formats, the already existing standards do not measure up to expectation. To establish end-to-end image compression, a number of approaches have been examined. To improve performance, context-adaptive entropy models were introduced in image compression. These algorithms used hyperprior for adding bits that are in excess. Similarly, in [2], hyperprior was combined with autoregressive mask convolution. In [3], the work takes into consideration two contexts namely bit-free contexts and bit consuming contexts to determine the adaptive entropy model.

In this paper, we have proposed a model with improved performance which built on recent entropy model techniques [4]. We have introduced a discretized Gaussian mixture likelihood based entropy model which is flexible and accurate. Using the compressed techniques, we visualize the compressed codes' spatial redundancy. Based on the output, we have used discretized Gaussian mixture likelihood that can be used to determine the distribution criteria in order to attain an entropy model with high accuracy [5]. This methodology uses only a few encoding bits when compared to the others. In addition, an attention module is also introduced in the network architecture which enables the mode to concentrate on complex regions. This will reflect in the output where training complexity is moderate by improving the performance of our code. Both MS-SSIM and PSNR quality metric values that are obtained through experimental observations are better when compared with classical image

compression standards such as JPEG, JPEG2000 and HEVC [6]. Moreover, the method proposed by us when incorporated with MS-SSIM, produces a visually pleasant image on reconstruction.

2. Related Work

VVC, HEVC, JPEG2000 and JPEG are some of the previously existing image compression standards that are handcrafted. Modules such as quantization, wavelet transform or discrete cosine transform, intra prediction, entropy coders like content adaptive binary arithmetic coder (CABAC) and Huffman coder are included. Each module in the network is designed with many modes and to determine the best node, rate-distortion optimization is performed. Specifically, VVC is more compatible with many transform types, many prediction modes, larger coding units and wide range of coding tools. Taking advantage of the learned super resolution approaches and conventional compression algorithm, we have implemented a hybrid methodology which is a development of classical compression algorithms. In recent years, there has been upliftment in the demand for deep learning based image compression methodologies [7]. In this regard, promising results have been observed when auto-encoder architecture is implemented.

Over the years, learning based compression models have been implemented in a number of applications and have various related works. However, at the start of certain work, end-to-end training is made possible using rate estimation and non-differential quantization. In the next stage, there is much concentration on the network structure which is essential in presenting efficient latency and compatibility. It is also used to produce images of high-quality from the compressed input. Authors in [8] have used generative models to determine image distribution with the aid of adversarial training which enables higher quality images. Some other works [9] use recent neural networks that work on binary data representation and data compression which can be retrieved using scalable coding. Some other approaches include deep residual units, energy compaction, principal component analysis, de-correlating channels or content weighted strategy to improve the network architecture. In recent years, many studies have examined the adaptive context model which can be used to determine the parameters of neural network that will help attain best trade-off between required bits and construction errors. In learned compression algorithm, the incorporation of entropy estimation shows good improvement [10]. The best representation of the two methods is using joint model and hyper prior model. There is some variation between the true marginal distribution and estimated distribution of the latent representation. Several parameterized distribution models have been investigated over the years such as PixelCNN [11].

3. Proposed Methodology

3.1. Learned Compression Model Formulation

The following are the formula used for image compression:

$$y = g_a(x, \emptyset) \quad (1)$$

$$\hat{y} = Q(y) \quad (2)$$

$$\hat{x} = g_s(x, \theta) \quad (3)$$

Where \hat{x} , \hat{y} , y and x are the reconstructed images, compressed codes, prior to quantization latent presentation and raw images respectively. θ and \emptyset are the optimized synthesis transform and analysis parameters respectively. Entropy coding and quantization can be represented as O|Q. In order to determine the noisy code \tilde{y} , uniform noise, $O\left(-\frac{1}{2}, \frac{1}{2}\right)$ is used for quantization during training. On the other hand, during inference, \tilde{y} is generated through quantization based on real round. This can be further modified such that $\tilde{y}|\hat{y}$ can be represented as \hat{y} . Using a probability model, it is possible to compress the quantized codes in a lossless manner using entropy coding technique [12-13].

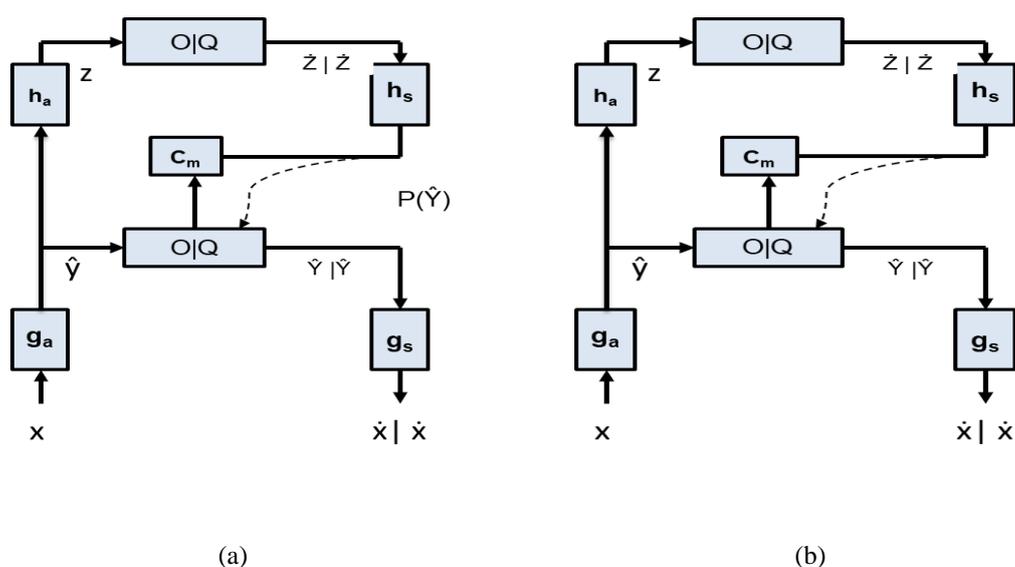


Fig.1.a) Joint b) Proposed Model

A side information is introduced in order to search the elements to determine the spatial dependencies using the formula:

$$\hat{z} = Q(z) \quad (4)$$

$$z = h_a(y; \phi_h) \quad (5)$$

$$P_{\hat{y}|\hat{z}}(\hat{y}|\hat{z}) \leftarrow h_s(\hat{z}; \theta_h) \quad (6)$$

where θ_h and ϕ_h are optimized parameters, h_s and h_a represents synthesis transforms and analysis in auxiliary auto-encode and the distributed condition can be represented by $P_{\hat{y}|\hat{z}}(\hat{y}|\hat{z})$.

3.2. Discretized Gaussian Mixture Likelihoods

In an entropy model, it is essential to determine if the marginal distribution of \hat{y} is satisfied by the parameterized distribution model as it affects the rate-distortion performance. To model $P_{\hat{y}}(\hat{y}|\hat{z})$ on decoding, much effort is required. Hence consider a uni-variate Gaussian distribution model such that

$$P_{\hat{y}|\hat{z}}(\hat{y}|\hat{z}) \sim \mathcal{N}(0, \sigma^2) \quad (7)$$

Based on observations, it is identified that the value of \hat{y} is closer to μ . When the value of σ is small, it will result in smooth regions whereas when the value is large, it will require more number of encoding bits. When Joint entropy model is compared with HyperPrior model, we can find that on adding autoregressive mode, we can delete more structures. Still, there are some spatial redundancies that are noticed when using Joint entropy model. Based on these analyses, we propose the use of Gaussian mixture model which can be represented as:

$$P_{\hat{y}|\hat{z}}(\hat{y}|\hat{z}) \sim \sum_{k=1}^K \omega^{(k)} \mathcal{N}(\mu^{(k)}, \sigma^{2(k)}) \quad (8)$$

The proposed model will be able to achieve higher efficiency since the value of K chosen indicates a smaller scale. This work is similar to that of CABAC (Context-based Adaptive Binary Arithmetic Coding)

4. Network Architecture

In our network architecture, we have incorporated a residual block to improve the rate-distortion performance by increasing large receptive field. Instead of transposed convolution, sub-pixel convolution is used in the decoder side in order to store more data. The model capacity can be expressed using the letter 'N' indicating the number of channels. In this proposed work, the Gaussian mixture model will need $K \times N \times 3$ channels to denote the capacity of the model. In image compression and restoration, attention modules have been introduced in recent days to improve its performance. Fig.2.a. represents the attention module which is not preferred because of the time requirement it involves. However, we can further simplify this by deleting the non-local block since it is possible to hold a great receptive field in the network architecture. A comparison of training time and comparison

between the different attention modules indicate that there is a significant reduction by 16 epoches on training. Fig. 2.b represents the attention model simplified in order to decrease the loss that is incurred.

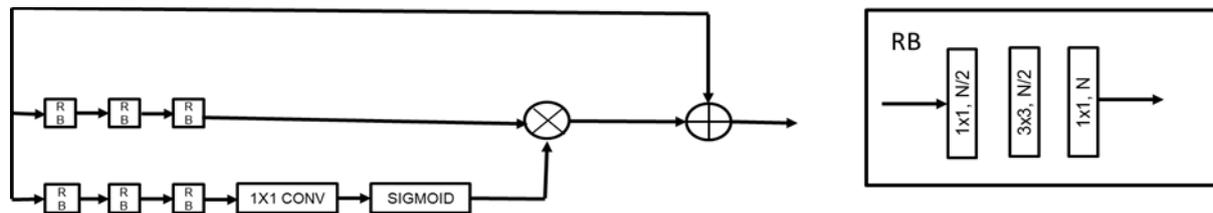


Fig. 2.(a) Simplified Attention Module

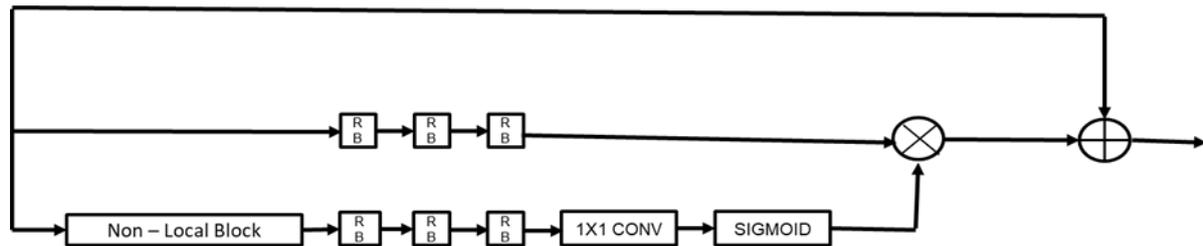


Fig.2.(b) Attention Module

4. Results and Discussion

The rate-distortion performance is represented in Fig.3 where the difference can be observed in a clear manner by converting MS-SSIM to decibels. Using the PSNR, the values of RD points can be observed. As far as PSNR is concerned, the method proposed by us shows positive results when VVC is used. Similarly, in Fig. 8 a comparison of the different validation results such as HEVC, JPEG2000 and JPEG are observed. The proposed work also provides favourable response for high resolution images.

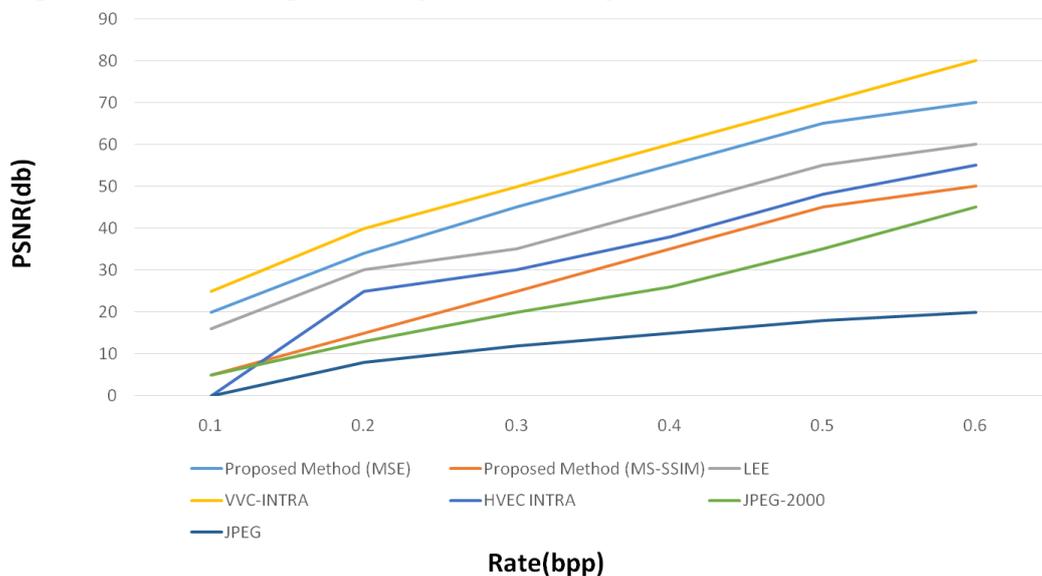


Fig.3 Performance Evaluation

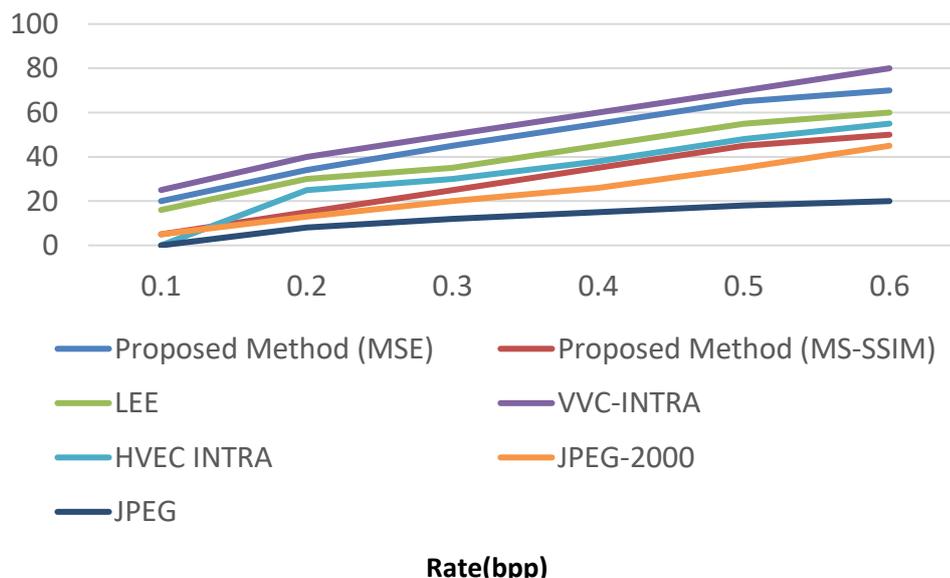


Fig.4. Evaluation on Validation Dataset

5. Conclusion

In this paper, a discretized Gaussian mixture likelihood with attention modules is executed for learned image compression. Using the redundancy left in the learned compression technique, one can determine that it is not possible to attain parameterized model and restriction on entropy model accuracy. In order to achieve accurate and flexible entropy model, we have used discretized Gaussian Mixture Likelihood. We have modified the attention module, simplifying it further such that it is possible to attain high coding efficiency, with moderate complexity. Simulation results show that the proposed methodology shows improved performance when compared with other coding standards and compression methods. Moreover, we have established a more sturdy performance with respect to VVC coding standard in PSNR. The output trained model outperforms the other previous methods existing, using MS-SIM technique.

References

- [1] . O'Hagan, A., Murphy, T. B., Scrucca, L., & Gormley, I. C. (2019). Investigation of parameter uncertainty in clustering using a Gaussian mixture model via jackknife, bootstrap and weighted likelihood bootstrap. *Computational Statistics*, 34(4), 1779-1813.
- [2] . Ai, J., Luo, Q., Yang, X., Yin, Z., & Xu, H. (2019). Outliers-robust CFAR detector of Gaussian clutter based on the truncated-maximum-likelihood-estimator in SAR imagery. *IEEE Transactions on Intelligent Transportation Systems*, 21(5), 2039-2049.
- [3] . Coons, J. I., Marigliano, O., & Ruddy, M. (2020). Maximum likelihood degree of the two-dimensional linear Gaussian covariance model. *Algebraic Statistics*, 11(2), 107-123.
- [4] Weber, J. H., & Immink, K. A. S. (2018). Maximum likelihood decoding for Gaussian noise channels with gain or offset mismatch. *IEEE Communications Letters*, 22(6), 1128-1131.
- [5] Kharrati-Kopaei, M. (2021). On the exact distribution of the likelihood ratio test statistic for testing the homogeneity of the scale parameters of several inverse Gaussian distributions. *Computational Statistics*, 1-16.
- [6] D.,Ruth Anita Shirley, A., Ranjani, K., Arunachalam, G., & Janeera, D. A. (2021). Automatic Distributed Gardening System Using Object Recognition and Visual Servoing. In *Inventive Communication and Computational Technologies* (pp. 359-369). Springer, Singapore.
- [7] Sellentin, E., & Heavens, A. F. (2018). On the insufficiency of arbitrarily precise covariance matrices: non-Gaussian weak-lensing likelihoods. *Monthly Notices of the Royal Astronomical Society*, 473(2), 2355-2363.
- [8] Ippoliti, L., Martin, R. J., & Romagnoli, L. (2018). Efficient likelihood computations for some multivariate Gaussian Markov random fields. *Journal of Multivariate Analysis*, 168, 185-200.
- [9] Rios, G., & Tobar, F. (2018, July). Learning non-Gaussian time series using the Box-Cox Gaussian process. In *2018 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

- [10] Subba Rao, S., & Yang, J. (2020). Reconciling the Gaussian and Whittle Likelihood with an application to estimation in the frequency domain. *arXiv e-prints*, arXiv-2001.
- [11] Xie, X., Zhang, Y., Wang, X., & Peng, D. (2019). A Mixture Likelihood Model of the Anisotropic Gaussian and Uniform Distributions for Accurate Oblique Image Point Matching. *IEEE Geoscience and Remote Sensing Letters*, 16(9), 1437-1441.
- [12] Manoharan, S. (2020). Population Based Meta Heuristics Algorithm for Performance Improvement of Feed Forward Neural Network. *Journal of Soft Computing Paradigm (JSCP)*, 2(01), 36-46.
- [13] Dhaya, R. (2020). Improved Image Processing Techniques for User Immersion Problem Alleviation in Virtual Reality Environments. *Journal of Innovative Image Processing (JIIP)*, 2(02), 77-84.