Perceptual Image Compression using Relativistic Average Least Squares GANs

Zhengxue Cheng, Ting Fu, Jiapeng Hu, Li Guo,
Shihao Wang, Xiongxin Zhao, Dajiang Zhou, Yang Song
Ant Group, Xihu District, Hangzhou, China.
{zhengxue.czx, ft224739, libby.hjp, li.gl, shihao.wsh,
xiongxin.zxx, dajiang.zdj, zhaoshan.sy}@antgroup.com

Abstract

In this work, we provide a detailed description on our submitted methods ANTxNN and ANTxNN_SSIM to Workshop and Challenge on Learned Image Compression (CLIC) 2021. We propose to incorporate Relativistic average Least Squares GANs (RaLSGANs) into Rate-Distortion Optimization for end-to-end training, to achieve perceptual image compression. We also compare two types of discriminator networks and visualize their reconstructed images. Experimental results have validated our method optimized by RaLSGANs can achieve higher subjective quality compared to PSNR, MS-SSIM or LPIPS-optimized models.

1. Introduction

Image compression is a fundamental topic in the field of image processing for many decades owing to its significant effect to transmission and storage. Many representative compression standards have been widely used in industries, such as JPEG [1], JPEG2000 [2], High Efficiency Video Coding (HEVC/265) [3] and latest Versatile Video Coding (VVC/266) [4]. These standards have achieved impressive coding efficiency with more and more complex hand-engineered coding tools. Although the computational complexity is significantly increasing, further improvement on coding efficiency becomes more and more difficult relative to its predecessor.

Recent learned image compression has made tremendous progress and has achieved impressive coding efficiency regarding well-kown PSNR and MS-SSIM quality metrics in [5, 6, 7, 8, 9, 10] etc. One main advantage of learned image compression is that it is flexible to be adapted to any quality metrics as long as this quality metric is differentiable. Different quality metrics leads to different reconstruction artifacts. Some of artifacts are easily captured by human visual system such as blurring, while some of them are not. For instance, LPIPS [11] is a neural-network based quality metric, which is found to correlate much better with human evaluations in CLIC 2020 low-rate image compression track [12] compared with PSNR and MS-SSIM. Generative adversarial models also exhibit a superior performance in terms on image subjective quality. However, GANs sometimes lead to the unstable training process in rate-distortion optimization problem. In [13] and [14], authors either used a frozen learned encoder or a fixing VVC encoder and only learned the decoder using VGG-19 and ESRGAN losses. In [15], selective detail decoding was realized by an additional selective decoder and stable training could be achieved by using a softplus-based discriminator loss function. In [16], authors present very competitive perceptual qualities, and found MSE is likely to be the driving force for stability, and jointly learning encoder without MSE led to a collapse of GAN training. However, there is a large room to explore the discrminator architectures and GAN training strategies.

This work presents a detailed description on our submitted methods *ANTxNN* and *ANTxNN_SSIM* to Workshop and Challenge on Learned Image Compression (CLIC) 2021. The network architecture combines recent techniques, including split attention and Gaussian mixture models. To achieve a higher perceptual quality, we apply Relativistic Average Leaset Squares GANs to achieve a stable rate-distortion. We also incorporate LPIPS, a learned quality metric as one of distortion metrics. Experimental results demonstrate our models can achieve higher subjective quality than optimizing LPIPS, or MSE or MS-SSIM only.

2. Related Work

Learned Image Compression Related studies have validated image compression can be formulated as a rate-distortion optimzation problem and then solved benefitting recent deep learning techniques in [5, 6, 7, 8, 9, 10]. Some other works explored the approximations of non-differentiable quantization, such as uniform noise [17], straight-through proxy [18], soft quantization [19], universal quantization [20], for the better gradient back-propagation. Some works proposed different network struc-

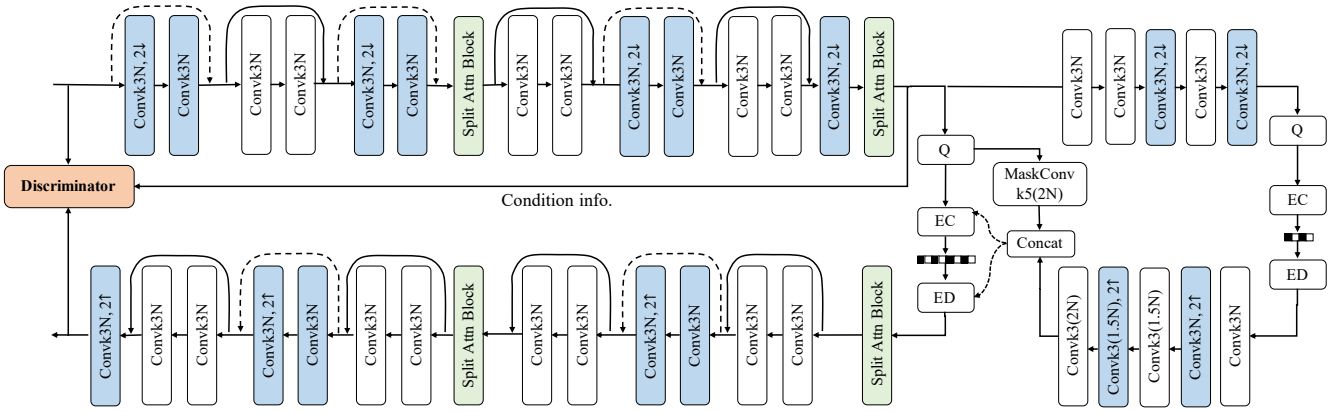

Figure 1: The network structure we used, by referring to [8] as a baseline model. *Orange* color highlights the discriminator, to tell the difference between raw images and compressed images, and it also takes the compressed codes as the conditional information, similar to the design of conditional GANs (CGANs). The backbone is basically same as [8], except the split attention block filled up by *green* proposed in [32]. Besides, *blue* denotes the convolution with the stride of 2 to downscale or upscale feature maps and other convolutions are inserted to maintain large model capacity.

tures, including recurrent neural networks [22], content-weighted map [23] or principle component analysis-based channels de-correlating [24], or energy compaction regularizer [25], deep residual blocks [26], octave convolution [27], channel-level variable quantization network [28]. Variable-rate compression was studied in [29] using a layer-wise scaling based on conditional convolutions.

Perceptual Image Compression To further remove the redundancy from the viewpoint of human perceived quality, many approaches are seeking for solutions from generative adversarial networks to preserve the distributions of raw images, such as some pioneer works [30, 21]. However, rate-distortion-perception tradeoff is mathematically discussed in [31] to show adding perceptual quality constraint leads to a sacrifice of either rate or distortion. It has been validated by the results of [13, 14, 15] in the leaderboard of CLIC 2020 low-rate compression track, and subjective quality optimization usually leads to the decreasing of objective quality metrics, such as PSNR or MS-SSIM with a fixed rate constraint.

3. Proposed Method

3.1. Problem Formulation and Network

Learned image compression is formulated as a Lagrangian multiplier-based rate-distortion optimization problem. This problem can be solved by minimizing its loss function:

$$\mathcal{L} = \lambda \times \mathcal{D}(\boldsymbol{x}, \hat{\boldsymbol{x}}) + \mathcal{R}(\hat{\boldsymbol{y}}) + \mathcal{R}(\hat{\boldsymbol{z}}) \tag{1}$$

where λ controls the rate-distortion trade-off. Different λ values are corresponding to different bit rates. $\mathcal{R}(\hat{\boldsymbol{y}})$ and $\mathcal{R}(\hat{\boldsymbol{z}})$ are rate terms, calculated by $\mathbb{E}[-\log_2(p_{\hat{\boldsymbol{y}}|\hat{\boldsymbol{z}}}(\hat{\boldsymbol{y}}|\hat{\boldsymbol{z}}))]$ and $\mathbb{E}[-\log_2(p_{\hat{\boldsymbol{z}}|\boldsymbol{\psi}}(\hat{\boldsymbol{z}}|\boldsymbol{\psi}))]$. The probabilities are further calculated by

$$p_{\hat{\boldsymbol{y}}|\hat{\boldsymbol{z}}}(\hat{\boldsymbol{y}}|\hat{\boldsymbol{z}}) = \prod_i (\sum_{k=1}^{K} w_i^{(k)} \mathcal{N}(\mu_i^{(k)}, \sigma_i^{2(k)}) * \mathcal{U}(-\frac{1}{2}, \frac{1}{2}))(\hat{y}_i) \tag{2}$$

$$p_{\hat{\boldsymbol{z}}|\boldsymbol{\mu_z},\boldsymbol{\sigma_z}}(\hat{\boldsymbol{z}}|\boldsymbol{\mu_z},\boldsymbol{\sigma_z}) = \prod_{j \in N} (\mathcal{N}(\mu_{z_j}, \sigma_{z_j}) * \mathcal{U}(-\frac{1}{2}, \frac{1}{2}))(\hat{z}_j) \tag{3}$$

where y_i, z_i denote the i-th element of $\boldsymbol{y}$ and $\boldsymbol{z}$. The likelihoods of y_i are estimated by predicted $w_i^{(k)}$, $\mu_i^{(k)}$ and $\sigma_i^{2(k)}$, which are parameter of Gaussian Mixture Models (GMM). Typical K is 3. The entropy models for $\hat{\boldsymbol{z}}$ is a channel-wise single Gaussian model given by a pair of mean $\boldsymbol{\mu_z}$ and scale $\boldsymbol{\sigma_z}$ with the size of $[N, 1]$, where j denotes the index of channels, N denotes the number of total channels. $\mathcal{D}(\boldsymbol{x}, \hat{\boldsymbol{x}})$ denotes the distortion term and will be discussed in next section.

The network architecture we used is shown in Fig. 1. We follows the latest learned image compression approach with a hyperprior and autoregressive model, and uses a Gaussian mixture models (GMMs) with K mixtures [8, 9], which are reported to have state-of-the-art performance. But, we replace the attention module by grouped-separated attention block in [32] to further enhance the performance. This block can enable separate groups attention and improve the PSNR by about 0.1dB at the same bit rate. All the parameters in Fig. 1 are end-to-end learned to achieve the best rate-distortion optimization.

Table 1: Two different types of discriminators, where all the layers use leakyReLU with alpha of 0.2 as activation functions except for the last layer with no activation.

$\textbf{disc}_{4\times4}$	$\textbf{disc}_{\text{RB}}$
$\boldsymbol{y} \in \mathbb{R}^{12\times12\times192}$	
Conv 3×3, stride 1, 12	
UpSampling2D (size=(16, 16))	
Concat($\boldsymbol{x}$, features), $\boldsymbol{x} \in \mathbb{R}^{192\times192\times3}$	
Conv 4×4, stride 2, 64	RB 3×3, stride 2, 64
Conv 4×4, stride 2, 128	RB 3×3, stride 2, 128
Conv 4×4, stride 2, 256	RB 3×3, stride 2, 256
Conv 4×4, stride 2, 256	RB 3×3, stride 2, 256
Conv 4×4, stride 2, 512	RB 3×3, stride 2, 512
Conv 4×4, stride 2, 512	RB 3×3, stride 2, 512
Conv 4×4, stride 1, 512	RB 3×3, stride 2, 512
Flatten	Flatten
Dense layer, 1024	Dense layer, 1024
Dense layer, 1	Dense layer, 1

3.2. Least Squares Relativistic Generative Adversarial Learning

As Fig. 1, we add a discriminator network to tell the difference between raw images $\boldsymbol{x}$ and compressed images $\boldsymbol{y}$ are also fed into the discriminator network as the conditional information, similar to the design of conditional GANs (CGANs). We tried two designs of discriminators, as listed in Table 1. Different discriminator structures lead to different artifacts, which will be visualized in Section 5.

Then we further explore different types of GAN losses, including standard GAN, HingeGAN, WGAN and LSGAN. Experimental results show LSGAN [33] shows the most stable performance for end-to-end training. So we use LSGAN. Beside, we use the relative average GAN [34] by assuming the real samples should be better than average fake ones, and vice versa. Denote the $C(x)$ and $C(\hat{x})$ are the output of discriminator with the input of raw images x and compressed images $\hat{x}$. The relative output is computed as:

$$\begin{aligned} \overline{C(\boldsymbol{x})} &= C(\boldsymbol{x}) - \mathbb{E}(C(\hat{\boldsymbol{x}})) \\ \overline{C(\hat{\boldsymbol{x}})} &= C(\hat{\boldsymbol{x}}) - \mathbb{E}(C(\boldsymbol{x})) \end{aligned} \tag{4}$$

Then the Rate-distortion loss in Eq. (1) is incorporated into the generator loss. The final generator losses and discrminator losses are listed as

$$\begin{aligned} \mathcal{L}_g =& (\mathbb{E}[(\overline{C(\hat{\boldsymbol{x}})} - 1)^2] + \mathbb{E}[(\overline{C(\boldsymbol{x})} + 1)^2]) * \beta + \\ & \lambda \times \mathcal{D}(\boldsymbol{x}, \hat{\boldsymbol{x}}) + \mathcal{R}(\hat{\boldsymbol{y}}) + \mathcal{R}(\hat{\boldsymbol{z}}) \\ \mathcal{L}_d =& \mathbb{E}[(\overline{C(\boldsymbol{x})} - 1)^2] + \mathbb{E}[(\overline{C(\hat{\boldsymbol{x}})} + 1)^2] \end{aligned} \tag{5}$$

where β defines the weight of generator, and is set as 0.015 in our experiments. The distortion term is a combination of MSE and LPIPS, as

$$\mathcal{D}(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \text{MSE} + k_p * \text{LPIPS} \tag{6}$$

where k_p controls the weight of LPIPS and k_p is set as 10^3 in our experiments. λ controls the rate and will be given in Section 4.

4. Implementation Details

Training Details For training, we used a 2K resolution high-quality *DIV2K* dataset [35] with 800 images to avoid over-fitting to JPEG compressed artifacts and distributions. We cropped them into samples with the size of $192\times192\times3$, instead of 256×256 to reduce the training time. The number of channels N in Figure 1 is set as 192.

Our submitted method *ANTxNN_SSIM* is optimized by only MS-SSIM quality metric [36], i.e. $\mathcal{D}(\boldsymbol{x}, \hat{\boldsymbol{x}}) = 1 - \text{MS-SSIM}(\boldsymbol{x}, \hat{\boldsymbol{x}})$. When optimized by MS-SSIM, λ belongs to the set $\{1.5, 2, 6, 18, 24\}$ to get five models. They were trained using the Adam [37] algorithm with a batch size of 8. The learning rate was kept at a fixed value of 1×10^{-4} for $500k$ iterations, and was reduced to 1×10^{-5} for the last $80k$ iterations. Then the bit allocation is solved as a knapsack problem to push the total bits to the targeted bpp.

Our submitted method *ANTxNN* is optimized by a combined distortion metrics, including LPIPS, mean square error (MSE) and GAN loss. λ belongs to the set $\{10^{-3} * (0.45, 0.5, 0.75, 1, 2, 4, 6, 8)\}$ to get eight models. Both discriminator and generator were trained using the Adam [37] algorithm with a batch size of 8. All the parameters are firstly initialized by a model trained for $500k$ iterations by MSE only. Then, the learning rate was kept at a fixed value of 1×10^{-4} for $300k$ iterations, and was reduced to 1×10^{-5} for the last $80k$ iterations. Learning rate decay can train the models to the convergence.

Textual Region Enhancement Due to the special attention of human visual system on textual regions, we used a simple OCR detection algorithm to get the enhanced mask and merge sub-masks into a whole region of interest (ROI). Then we crop the ROI from reconstructed images and encode the residual image using VVC [39] encoder, as a ROI-based enhancement.

Quantized rate computation We apply layer-wise fake quantization after each convolution in a hyper decoder and autoregressive model with a fixed encoder and decoder. And, we also apply fake quantization after the softmax operator to the weights of GMM models, although softmax is not totally fixed-point calculation, as

$$\begin{aligned} \mu_i^{(k)} &= \text{round}(\mu_i^{(k)}/q_\mu) * q_\mu \\ \sigma_i^{(k)} &= \text{round}(\sigma_i^{(k)}/q_\sigma) * q_\sigma \\ w_i^{(k)} &= \text{floor}(w_i^{(k)}/q_w) * q_w, k = 0, 1 \\ w_i^{(2)} &= 1.0 - w_i^{(0)} - w_i^{(1)} \end{aligned} \tag{7}$$

In our experiments, q_μ is set as 0.2, q_σ is set as 0.1 and q_w is set as 0.0001. Quantized rate computation leads to less than 3% rate increment. But, quantized rate computation can avoid the floating-point calculation errors and improve the computation stability.

5. Results

In this section, experimental results and ablation studies are present. First, we have shown the results of our submitted methods on CLIC validation dataset in Table. 2 and Table 3. *ANTxNN_SSIM* ranks the 2-nd in terms of MS-SSIM. *ANTxNN* ranks the 6-th, 2-nd and 3-rd in terms of FID at the rate of 0.075bpp, 0.15bpp and 0.30bpp, separately.

Table 2: Results of our submitted methods *ANTxNN*.

Model	PSNR	MS-SSIM	FID	Rate (bpp)
ANTxNN	33.726	0.97748	144.509	0.30
ANTxNN	31.315	0.96098	158.102	0.15
ANTxNN	27.465	0.91785	192.697	0.075

Table 3: Results of our submitted methods *ANTxNN_SSIM*.

Model	PSNR	MS-SSIM	FID	Rate (bpp)
ANTxNN_SSIM	33.726	0.97748	144.509	0.30
ANTxNN_SSIM	31.315	0.96098	158.102	0.15
ANTxNN_SSIM	27.465	0.91785	192.697	0.075

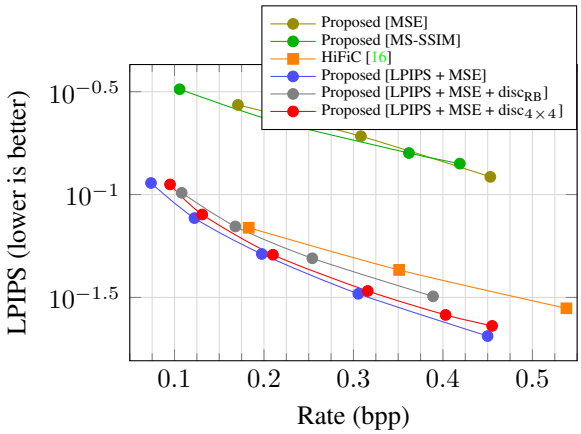

Figure 2: Comparison of R-D curves using LPIPS quality metric on the Kodak [38] dataset.

Besides, ablation study on rate-LPIPS rate curves is shown in Fig. 2. We can observe that the models optimized by MSE or MS-SSIM can not achieve good LPIPS values. By incorporating LPIPS into the loss function, the perceptual quality of reconstructed images can be significantly improved. We also compare our results with recent perceptual-optimized compression approach [16][1]. It can be seen our method achieves lower LPIPS performance compared to [16]. Models optimized by LPIPS+MSE achieved the lowest LPIPS, and adding GAN slightly increases the value of LPIPS.

[1] https://hific.github.io

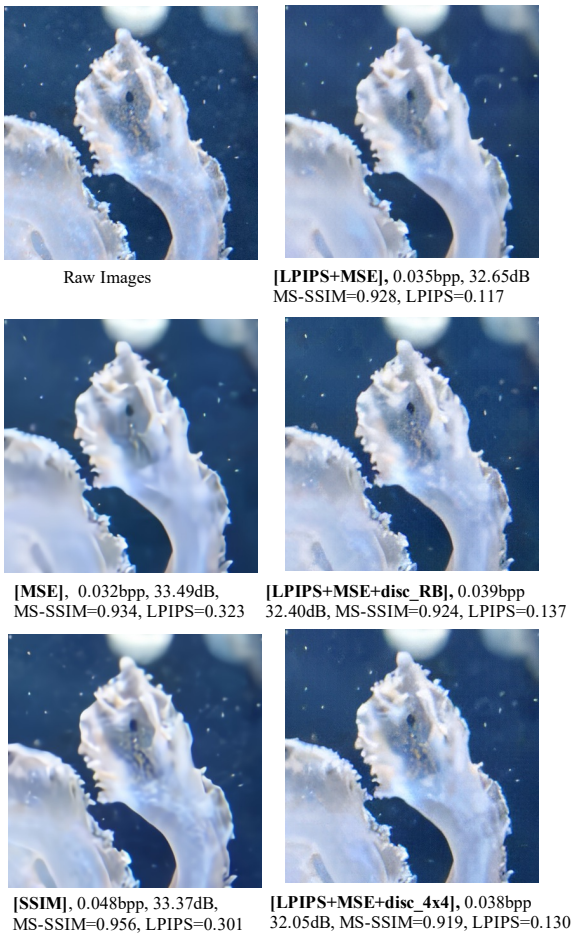

Figure 3: Visualization of different artifacts.

Fig. 3 visualize the compressed artifacts by different methods using the same network, i.e optimizing by MSE, MS-SSIM, LPIPS+MSE, LPIPS+MSE+GAN. We can observe adding discriminator can improve the perceptual quality significantly than optimizing LPIPS, MSE or MS-SSIM only. Since 4×4 discriminator is slightly better than discriminator with residual block, so our submitted method *ANTxNN* used 4×4 discriminator.

6. Conclusion

In this paper, we have described our methods *ANTxNN* and *ANTxNN_SSIM*, which are submitted to challenge on learned image compression (CLIC) 2021. RaLSGANs are incorporated into rate-distortion optimization stably to improve perceptual quality of reconstructed images than other GAN losses. We also discuss two types of discriminator networks. Besides, visualization comparisons have demonstrated RaLSGANs-optimized compression model can generate higher subjective quality than optimizing LPIPS, PSNR, or SSIM only.

References

[1] G. K Wallace, *"The JPEG still picture compression standard"*, IEEE Trans. on Consumer Electronics, vol. 38, no. 1, pp. 43-59, Feb. 1991. 1

[2] Majid Rabbani, Rajan Joshi, *"An overview of the JPEG2000 still image compression standard"*, ELSEVIER Signal Processing: Image Communication, vol. 17, no, 1, pp. 3-48, Jan. 2002. 1

[3] G. J. Sullivan, J. Ohm, W. 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, Dec. 2012. 1

[4] G. J. Sullivan and J. R. Ohm, *"Versatile video coding Towards the next generation of video compression"*, Picture Coding Symposium, Jun. 2018. 1

[5] J. Ballé, P. A. Chou, D. Minnen, S. Singh, N. Johnston, E. Agustsson, S. J. Hwang, G. Toderici, *"Nonlinear Transform Coding"*, IEEE Journal of Selected Topics in Signal Processing, arXiv preprint, arXiv. 2007.03034. 1

[6] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, N. Johnston, *"Variational Image Compression with a Scale Hyperprior"*, Intl. Conf. on Learning Representations (ICLR), pp. 1-23, 2018. 1

[7] D. Minnen, J. Ballé, G. Toderici, *"Joint Autoregressive and Hierarchical Priors for Learned Image Compression"*, arXiv.1809.02736. 1

[8] Z. Cheng, H. Sun, M. Takeuchi, J. Katto, *"Learned Image Compression with Discretized Gaussian Mixture Likelihoods and Attention Modules"*, IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR) 2020. 1, 2

[9] J. Lee, S. Cho, M. Kim, *"An End-to-End Joint Learning Scheme of Image Compression and Quality Enhancement with Improved Entropy Minimization"*, arXiv preprint, arXiv. 1912.12817. 1, 2

[10] D. Minnen, S. Singh, *"Channel-wise Autoregressive Entropy Models for Learned Image Compression"*, IEEE International Conference on Image Processing (ICIP) 2020. 1

[11] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, O. Wang, *"The Unreasonable Effectiveness of Deep Features as a Perceptual Metric"*, IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR) 2020. 1

[12] Workshop and Challenge on Learned Image Compression(CLIC), CVPR Workshop, http://www.compression.cc/challenge/ 1

[13] Y. Kim, S. Cho, J. Lee, S-Y Jeong, J. S. Choi, J. Do, *"Towards the Perceptual Quality Enhancement of Low Bit-rate Compressed Images"*, CVPR Workshop CLIC, 2020. 1, 2

[14] J. Lee, D. Kim, Y. Kim, H. Kwon, J. Kim and T. Lee, *"A Training Method for Image Compression Networks to Improve Perceptual Quality of Reconstructions"*, CVPR Workshop CLIC, 2020. 1, 2

[15] H. Akutsu, A. Suzuki, Z. Zhong, and K. Aizawa, *"Ultra Low Bitrate Learned Image Compression by Selective Detail Decoding"*, CVPR Workshop CLIC, 2020. 1, 2

[16] F. Mentzer, G. Toderici, M. Tschanne, E. Agustsson, *"High-Fidelity Generative Image Compression"*, NeurIPS 2020 1, 4

[17] J. Ballé, Valero Laparra, Eero P. Simoncelli, *"End-to-End Optimized Image Compression"*, Intl. Conf. on Learning Representations (ICLR), pp. 1-27, April 24-26, 2017. 1

[18] Lucas Theis, Wenzhe Shi, Andrew Cunninghan and Ferenc Huszar, *"Lossy Image Compression with Compressive Autoencoders"*, Intl. Conf. on Learning Representations (ICLR), pp. 1-19, April 24-26, 2017. 1

[19] E. Agustsson, F. Mentzer, M. Tschannen, L. Cavigelli, R. Timofte, L. Benini, L. V. Gool, *"Soft-to-Hard Vector Quantization for End-to-End Learning Compressible Representations"*, Neural Information Processing Systems (NIPS) 2017, arXiv:1704.00648v2. 1

[20] E. Agustsson, L. Thesis, *"Universally Quantized Neural Compression"*, arXiv preprint, arXiv. 2006.09952, 2020. 1

[21] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. V. Gool, *"Generative Adversarial Networks for Extreme Learned Image Compression"*, arXiv:1804.02958. 2

[22] G. Toderici, S. M.O'Malley, S. J. Hwang, et al., *"Variable rate image compression with recurrent neural networks"*, arXiv: 1511.06085, 2015. 2

[23] M. Li, W. Zuo, S. Gu, D. Zhao, D. Zhang, *"Learning Convolutional Networks for Content-weighted Image Compression"*, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), June 17-22, 2018. 2

[24] Z. Cheng, H. Sun, M. Takeuchi, J. Katto, *"Deep Convolutional AutoEncoder-based Lossy Image Compression"*, Picture Coding Symposium, pp. 1-5, June 24-27, 2018. 2

[25] Z. Cheng, H. Sun, M. Takeuchi, J. Katto, *"Learning Image and Video Compression through Spatial-Temporal Energy Compaction"*, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), June 16-20, 2019. 2

[26] Z. Cheng, H. Sun, M. Takeuchi, J. Katto, *"Deep Residual Learning for Image Compression"*, CVPR Workshop, pp. 1-4, June 16-20, 2019. 2

[27] J. Lin, M. Akbari, H. Fu, Q. Zhang, S. Wang, J. Liang, D. Liu, F. Liang, G. Zhang, C. Tu, *"Learned Variable-Rate Multi-Frequency Image Compression using Modulated Generalized Octave Convolution"*, IEEE MMSP 2020. 2

[28] Z. Zhong, H. Akutsu, K. Aizawa, *"Channel-Level Variable Quantization Network for Deep Image Compression"*, Proceedings of International Joint Conference on Artificial Intelligence (IJCAI), 2020. 2

[29] Y. Choi, M. El-Khamy, and J. Lee, *"Variable Rate Deep Image Compression with a Conditional Autoencoder"*, Proceedings of the IEEE International Conference on Computer Vision, pp. 3146–3154, Oct. 2019. 2

[30] Ripple Oren, L. Bourdev, *"Real Time Adaptive Image Compression"*, Proc. of Machine Learning Research, Vol. 70, pp. 2922-2930, 2017. 2

[31] Yochai Blau and Tomer Michaeli. *"Rethinking lossy compression: The rate-distortion-perception tradeoff"*, arXiv preprint arXiv:1901.07821, 2019. 2

[32] Z. Guo, Z. Zhang, R. Feng, Z. Chen. *"Causal Contextual Prediction for Learned Image Compression"*, arXiv. 2011.09704. 2

[33] X. Mao, Q. Li, et al. *"Least Squares Generative Adversarial Networks"*, arXiv.1611.04076. 3

[34] Alexia Jolicoeur-Martineau, *"The relativistic discriminator: a key element missing from standard GAN"*, arXiv.1807.00734. 3

[35] E. Agustsson and Radu Timofte et al., *"NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study"*, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, July 2017. 3

[36] Z. Wang, E. P. Simoncelli and A. C. Bovik, *"Multiscale structural similarity for image quality assessment"*, The 36-th Asilomar Conference on Signals, Systems and Computers, Vol.2, pp. 1398-1402, Nov. 2013. 3

[37] D. P. Kingma and J. Ba, *"Adam: A method for stochastic optimization"*, arXiv:1412.6980, pp.1-15, Dec. 2014. 3

[38] Kodak Lossless True Color Image Suite, Download from http://r0k.us/graphics/kodak/ 4

[39] VVC Official Test Model VTM, https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM. 3