Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776810

HIGH-FIDELITY GENERATIVE IMAGE COMPRESSION USING GAN'S AND KNN APPROACH

Gadde Suresh [1], K.S.Thivya [2] & M.Anand [3]
Research scholar [1], Professor [2&3]
Department of Electronics and Communication Engineering,
Dr. MGR. Educational and Research Institute , Chennai 95.

Abstract
We conducted a study on combining Generative Adversarial Networks and learnt about the compression to create an advanced generative lossy compression system by utilizing KNN and GAN's approach. Our study focused on examining factors such as normalization layers, generator and discriminator architectures, training strategies, and perceptual losses. Our system is capable of producing visually pleasing reconstructions that are similar to the original input, can operate at a wide range of bitrates, and can handle high-resolution images. We tested our system using various perceptual metrics and a user study, which showed that our approach was better than existing approach, even when using more than 2 x bitrate. In summary, our study bridged the gap between rate-distortion-perception theory and practice.

1 Introduction

The number of images being produced by cameras is increasing rapidly. To save these images efficiently, lossy compression algorithms are used to reduce their size while still retaining their important visual features. Many different algorithms have been proposed over the years [50,43,54,58], including using video compression algorithms for single image compression. Recently, there has been a lot of interest in deep learning-based lossy compression [45,5,31,58], which involves training a neural network to balance compression rates and visual quality. This has resulted in new and improved state-of-the-art methods for image compression.

Although many image compression methods have been proposed, they all tend to cause noticeable image degradation as the compression factor increases. Classical algorithms start to exhibit artifacts like blocking or banding, while learning based approaches reveal issues with the distortion metric used to train the networks. Despite the development of various perceptual metrics, the weakness of each metric can be exploited by the learning algorithm, adding to problems such as checkerboard artifacts or blurry reconstructions. For instance, relying on MS-SSIM [53,52,58] can result in poor text reconstructions, while using MSE can lead to blurry reconstructions.

Agustsson et al [3] showed that Generative Adversarial Networks (GANs) could be used to reduce compression artifacts in extremely low bitrate compression (<0.08 bpp). While their approach produced reconstructions that looked convincing to the human eye, they tended to lose some of the finer details and deviate from the original input.

Blau and Michaeli [9] discovered a trade-off between three factors: rate (how compressed an image is), distortion (how different the compressed image is from the original), and perceptual quality (how visually pleasing the compressed image is). They defined distortion as a similarity metric comparing pairs of images and perceptual quality as the distance between the image distribution and the distribution of the reconstructions produced by the decoder, measured as a distribution divergence. They found that improving perceptual quality will always lead to worse distortion and vice versa, so a balance must be struck between the two. To do this, distortion can be traded for better perceptual quality by minimizing the difference between the input and reconstructed distributions, which can be done using Generative Adversarial Networks (GANs). While the theory is well-established, they only tested this trade-off on small datasets.

In this paper a new approach to compress high-resolution images while maintaining their visual quality is proposed. The proposed methodology is compared with previous existing methodology and it is found that the results obtained by proposed methodology is visually preferred even when the previous approaches have used higher bitrates. Various quantitative metrics have been discussed to evaluate the performance of the suggested approach and to show that the results are consistent with a theory called rate-distortion-perception. Our research demonstrates that although no metric could accurately predict the exact ranking of the user study, metrics such as FID and KID can be helpful in guiding the exploration process. In other words, these metrics can provide useful insights that can aid in making informed decisions during the exploration phase. Ensuring a comprehensive perceptual evaluation involves utilizing a diverse range of metrics that cover various aspects, such as no-reference metrics, pair-wise similarities, distributional similarities, and deep feature-based metrics derived from different network architectures. By using this ensemble of metrics, a more robust and comprehensive evaluation of perceptual qualities can be achieved. Our analysis involves a thorough examination of the suggested architecture and its constituent parts, which include normalization layers, generator and discriminator architectures, training methods, and the loss function. We assess these components based on both perceptual metrics and stability to gain a comprehensive understanding of their effectiveness.

2 Related Work

JPEG is the most commonly used algorithm for lossy compression [50]. There have been multiple attempts to develop alternatives to JPEG, such as WebP [54] and JPEG2000 [43], which involve the use of manually crafted algorithms. BPG [7] achieves high Peak Signal-to-Noise Ratio (PSNR) at different bitrates by utilizing the HEVC video codec [42]. In contrast, neural compression techniques aim to optimize Shannon's rate-distortion trade-off directly [14].

In the beginning, RNNs were utilized in the initial works [45, 47], whereas subsequent works were based on auto-encoders [5, 44, 1]. To achieve a reduced bitrate, various approaches have been employed to enhance the modeling of the probability density of auto-encoder latents, which, in turn, leads to more efficient arithmetic coding. These methods include hierarchical priors, auto-regression with different context shapes, or a combination of both [6, 31, 28, 39, 32, 26, 33]. State-of-the-art models, such as the one proposed by Minnen et al. [32], now surpass BPG in terms of PSNR.

GANs, introduced by Goodfellow et al. [17], have facilitated significant advancements in the field of unconditional and conditional image generation. The latest GAN models can generate high-resolution photo-realistic images [10, 21, 38], marking a remarkable progress in the field.The progress made in this area can be largely attributed to several key factors, including the expansion of training data and model size [10], the introduction of novel network architectures [21], and the development of new normalization techniques that help stabilize the training process [36]. These factors have collectively contributed to the significant advancements in the field. In addition to unconditional and conditional image generation, adversarial losses have also resulted in significant progress in various image enhancement tasks. These include tasks such as compression artifact removal [15], image de-noising [11], and image super-resolution [25]. Moreover, adversarial losses have also been utilized in the past to enhance the visual quality of neural compression systems [39, 40, 48, 3, 9]. For instance, [39] integrates an adversarial loss as a component in their full-resolution compression system. However, they do not thoroughly assess the advantages of this loss in terms of the quality of their reconstructions.

Although [40] offers a preliminary implementation of a low-resolution compression system with a GAN discriminator as the decoder, the primary focus of [48, 9] is to integrate a GAN loss into the rate-distortion objective in a logically coherent manner. To be specific, [48] suggests enhancing the rate-distortion objective by incorporating a distribution constraint to ensure that the reconstructions' distribution corresponds to the input distribution at all rates. On the other hand, [9] introduces and examines a triple trade-off between rate, distortion, and distribution matching. Lastly, [3] demonstrates that utilizing GAN-based compression systems at significantly low bitrates can result in bitrate savings of 2 times greater than those obtained from state-of-the-art engineered and learned compression algorithms.

3 Method

3.1 Background

Conditional GANs:

In the context of machine learning, Conditional Generative Adversarial Networks (GANs) refer to a method that enables the learning of a generative model of a conditional distribution p(X|S), where a given data point X is associated with additional information or context S, such as class labels or semantic maps. The joint distribution p (X,S) between the data point and context is unknown, and the conditional GANs help in estimating it. The approach has been used in various applications to generate images and other data that are conditioned on specific contexts. In Conditional GANs, two opposing networks are trained to learn a generative model of a conditional distribution p (X|S). The generator G, which is dependent on the information s, transforms samples y from a fixed known distribution pY into p(X|S). On the other hand, the discriminator D receives (x, s) input and evaluates the probability of it being a sample from p (X|S) rather than from G's output. The aim is to train the generator G to generate samples that can deceive the discriminator D into classifying them as real data coming from the distribution p(X|S). To achieve this, a "non-saturating" loss can be optimized, with s being kept constant during the process[17,58].

$$V_{DG} = E_{X\sim P_{data(x)}} [\log D (X)] + E_{Z - P_{data(z)}} [\log (1- D (G (Z))] \qquad (1)$$

G- generator, X- sample from real data, Z- sample from generator, D- Discriminator, $P_{data(x)}$- distribution of real data, $P_{data(z)}$- distribution of generator data, D(x) – Discriminator network and G(x)- generator network

KNN (K-Nearest Neighbors) classifier:

The KNN classifier is a machine learning algorithm utilized for regression and classification analysis. Its method involves identifying the k nearest data points to a given input data point from the training set, and classifying the input based on the most frequent class among its k-nearest neighbours. Typically, k is set to an odd number to prevent ties, and the similarity between data points is measured using a distance metric, such as Euclidean or Manhattan distance. KNN is a straightforward and efficient algorithm that can handle both binary and multi-class classification problems. To mathematically represent the k-NN algorithm, the following steps are followed. Firstly, the distance between the input data point and each data point in the training dataset is computed, which can be done using Euclidean distance, Manhattan distance, or other distance metrics. Secondly, the k data points that are the closest to the input data point are selected based on the computed distances. Thirdly, the number of data points in each class is counted among the k nearest neighbors. Finally, the majority class among the k nearest neighbors is assigned as the label for the input data point. The k-NN classifier can be expressed mathematically with the following equation:

$$y = \text{mode}(\{yi \mid xi \in Nk(x)\}) \qquad (2)$$

Here, y denotes the anticipated class label for the input data point x. The i-th data point in the training dataset is assigned a class label of yi, and Nk(x) signifies the set of k data points that are nearest to the input data point x. The mode function identifies the most frequently occurring class label among the k nearest neighbors.

Neural Image Compression

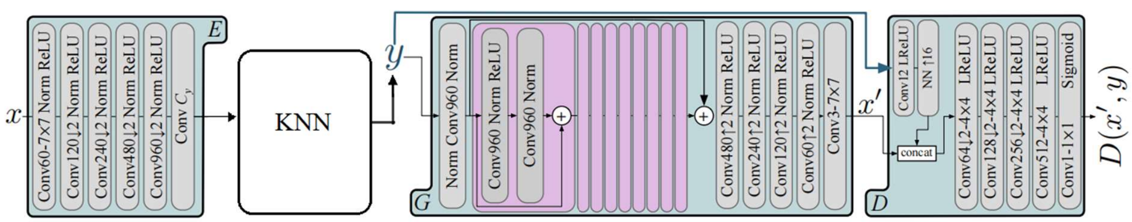
The concept of learned lossy compression relies on Shannon's rate-distortion theory, which considers the trade-off between the amount of information (rate) and the level of distortion that can be tolerated during compression. Typically, this problem is approached by utilizing an auto-encoder architecture that involves two main components: an encoder (E) and a decoder (G). In learned lossy compression, Shannon's rate-distortion theory [14,58] is used as a basis. An auto-encoder is used, which consists of an encoder E and a decoder G. In this process, to encode an image x, it is quantized into a latent representation y = E(x). The decoder G is used to reconstruct the image x' using y. The lossy compression results in a certain level of distortion, which can be measured using various metrics, such as MSE (mean squared error). The process of storing the quantized latent y is done through the introduction of a probability model P for y. With the help of an entropy coding algorithm, such as arithmetic coding, we can store y in a lossless manner. The bitrate required to store y can be calculated as r(y) = -log(P(y)), where the entropy coder may have a minor overhead of bits. By parameterizing the encoder E, decoder G, and probability model P as convolutional neural networks (CNNs), we can train them simultaneously to minimize the trade-off between rate and distortion, where a parameter ζ controls the trade-off:

$$V_{EG} = E_{X\sim Px}\,[\zeta\, r(y) + d\,(x, x')] \quad (3)$$

3.2 Formulation and Optimization

Our approach to neural image compression involves integrating a conditional GAN into the formulation, which involves combining equations 1 and 2 to train the networks E, G, K, and D. We utilize y = E(x) and s = y, while also incorporating the "perceptual distortion" dP = LPIPS based on the findings of a previous study that employed a VGG-based loss. In accordance with the formalism of another study, we group dP with MSE to form our overall distortion d = kMMSE+kP dP, where kM and kP are hyperparameters. By adjusting the hyperparameters ζ, ج, and to fine-tune the trade-off between the terms, we can achieve optimal results.

$$L_{EGK} = E_{x\sim px}\,[\,\zeta\, r(y) + d\,(x,x') - ج\log D\,(x',y)] \quad (4)$$



Our architecture is illustrated in Figure 1. ConvC refers to a convolution operation with C channels, using 3x3 filters, unless stated otherwise. Strided down or up convolutions are denoted by ↓2 and ↑2 respectively. ChannelNorm, as explained in the text, is used for normalization. LReLU with α=0.2 represents the leaky ReLU activation function [56,58]. Upsampling using nearest neighbor interpolation by a factor of 16 is referred to as NN↑16. Finally, Q quantization is employed [58].

Constrained Rate

While training a neural compression model with respect to the loss function mentioned in Equation 2, the only conflicting factor with the rate term r(y) is the solitary term d(x, x'). By manipulating only ζ, it is possible to regulate the ultimate (mean) bitrate of the model. Nevertheless, in our context, the MSE, dP, and -log(D(x')) are in conflict with the rate. When ζ is constant, models with varying kM, kP, and ج would have different bitrates, leading to difficulty in comparisons. To mitigate this issue, we introduce a hyper-parameter, "rate target" (rt), and modify Eq. 3 by replacing ζ with an adaptive term, ζ'. ζ' depends on two hyper-parameters, ζ(a) and ζ(b), such that ζ' equals ζ(a) if r(y) is greater than rt and ζ' equals ζ(b) otherwise. We set ζ(a) to be much greater than ζ(b) to train models with an average bitrate close to rt.

3.3 Architecture

Figure 1 depicts our architecture, comprising of the encoder E, generator G, discriminator D and KNN classifier block. We adopt the straight-through estimator, as in [44,58], for rounding y before inputting it into G. While our E, G, and D are based on [51,58,3], we introduce several distinctive modifications in the discriminator and normalization layers, which will be described in detail in the upcoming sections. While both [51,58, 3] utilize a multi-scale patch-discriminator D, we implement a single-scale patch-discriminator D and replace InstanceNorm [49,58] with SpectralNorm [36,58]. In contrast to [3], we condition D on y by concatenating an upscaled version of it with the image, as illustrated in Figure 1 [58]. This approach is inspired by the use of a conditional GAN formulation associated with KNN, where D can access the conditioning information (in our case, y as described in Section 3.2).

4 Experiments

Our study compares the performance of the proposed method, "High-Fidelity Generative Image Compression Using GAN's and KNN," which itself branches in to three approaches: Hi-fi-GAN's+KNN High, Hi-fi-GAN's+KNN Medium, and Hi-fi-GAN's+KNN Low with an approach that solely uses GAN, called "High-Fidelity Generative Image Compression," across three variations: Hi-fi-low, Hi-fi-medium, and Hi-fi-high [58]. To evaluate the effectiveness of these approaches, we use metrics such as MSE, SSIM, Histogram, and Compression Size.

Original Image	Original size	Hi-fi-Gan’s+KNN High	Hi-fi-Gan’s+KNN Medium	Hi-fi-Gan’s+KNN Low
 Image 1 Resolution: 2048*1366	4.3 MB	1.9 MB	1.8MB	1.9 MB
		Hi-fi-high	Hi-fi-medium	Hi-fi-low
		2.8 MB	2.0 MB	2.5 MB
 Image 2 Resolution:2016*1562	3.5 MB	Hi-fi-Gan’s+KNN High	Hi-fi-Gan’s+KNN Medium	Hi-fi-Gan’s+KNN Low
		1.5 MB	1.6 MB	1.6 MB
		Hi-fi-high	Hi-fi-medium	Hi-fi-low
		1.9 MB	2.0 MB	2.1 MB

 Image 3 Resolution:2040*1356	4.5 MB	Hi-fi-Gan’s+KNN High	Hi-fi-Gan’s+KNN Medium	Hi-fi-Gan’s+KNN Low
		2.7 MB	2.7 MB	2.8 MB
		Hi-fi-high	Hi-fi-medium	Hi-fi-low
		3.4 MB	3.7 MB	3.7 MB
 Image 4 Resolution:2040*1356	3.9 MB	Hi-fi-Gan’s+KNN High	Hi-fi-Gan’s+KNN Medium	Hi-fi-Gan’s+KNN Low
		1.4 MB	1.4 MB	1.4 MB
		Hi-fi-high	Hi-fi-medium	Hi-fi-low
		2.3 MB	2.6 MB	2.7 MB
 Image 5 Resolution:2040 *1536	3.4 MB	Hi-fi-Gan’s+KNN High	Hi-fi-Gan’s+KNN Medium	Hi-fi-Gan’s+KNN Low
		623.9 KB	617.5 KB	638.0 KB
		Hi-fi-high	Hi-fi-medium	Hi-fi-low
		1.0 MB	1.2 MB	1.3 MB
 Image 6	4.1 MB	Hi-fi-Gan’s+KNN High	Hi-fi-Gan’s+KNN Medium	Hi-fi-Gan’s+KNN Low
		1.7 MB	1.8 MB	1.8 MB
		Hi-fi-high	Hi-fi-medium	Hi-fi-low

Resolution:2040*1200		2.6 MB	2.9 MB	2.9MB
 Image 7 Resolution: 2040*1152	3.7 MB	Hi-fi-Gan's+KNN High	Hi-fi-Gan's+KNN Medium	Hi-fi-Gan's+KNN Low
		1.4 MB	1.4 MB	1.4 MB
		Hi-fi-high	Hi-fi-medium	Hi-fi-low
		2.5 MB	2.4 MB	2.4 MB

Table 1. Compressed size is showcased for Hi-fi-Gan's+KNN high, medium and low approach Vs Hi-fi-high, medium and low approach

COMPRESSED IMAGE VISUAL CLARITY		
Hi-fi-Gan's+KNN High	Hi-fi-Gan's+KNN Medium	Hi-fi-Gan's+KNN Low
 Image 1		
 Image 2		
 Image 3		

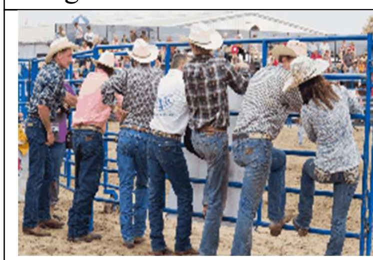
 Image 4		
 Image 5		
 Image 6		
 Image 7		

Table 2: Compressed Image visual quality is showcased for Hi-fi-Gan's+KNN high, medium and low approach.

Histogram plot for Original image and Hi-fi-Gan's+KNN high, medium and low approach is shown below.

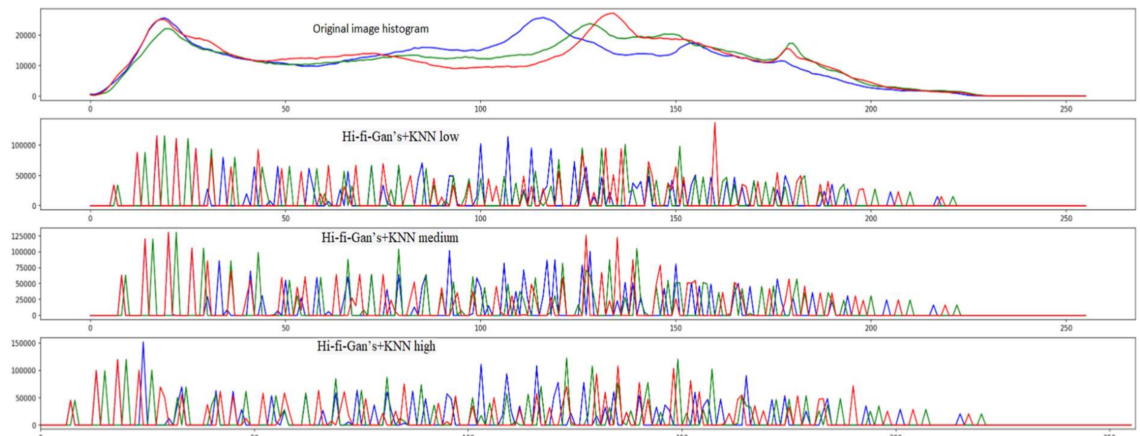


Figure 2 : Histogram plot for image 1.

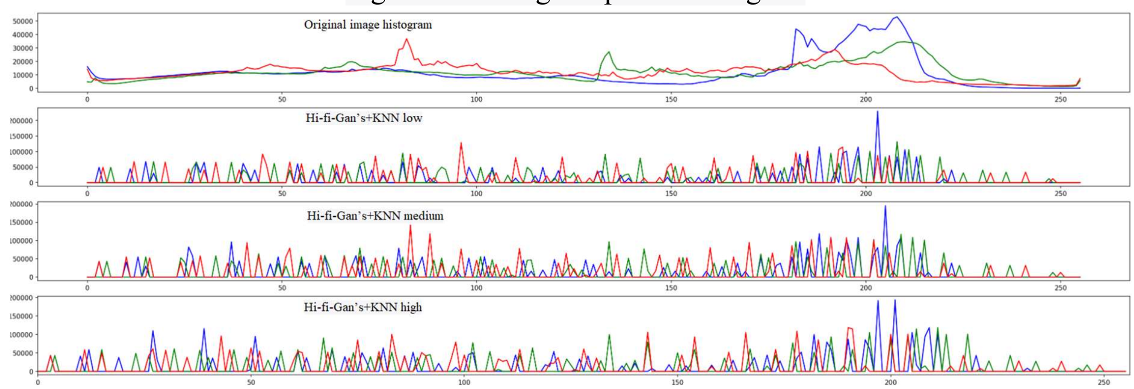


Figure 3 : Histogram plot for image 2.

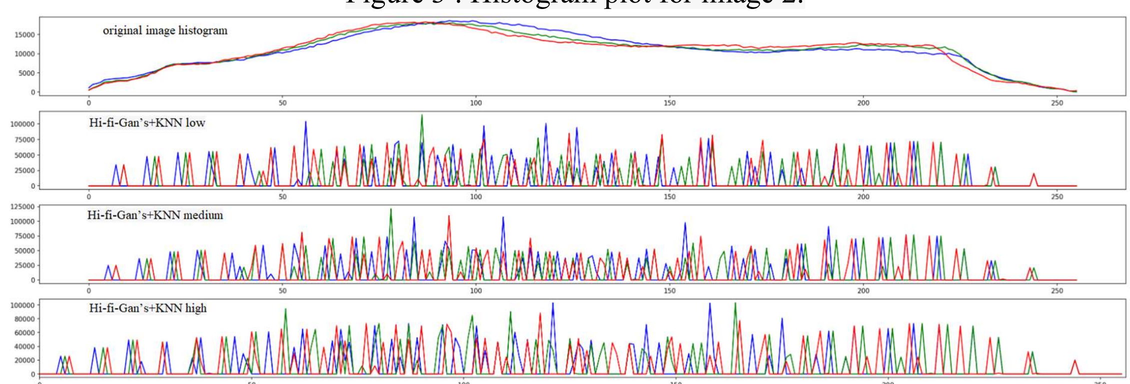


Figure 3 : Histogram plot for image 3.

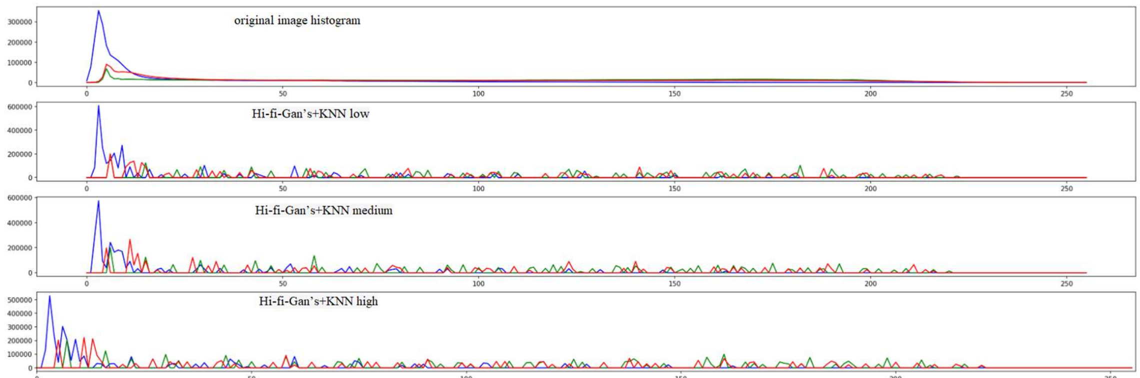


Figure 4 : Histogram plot for image 4.

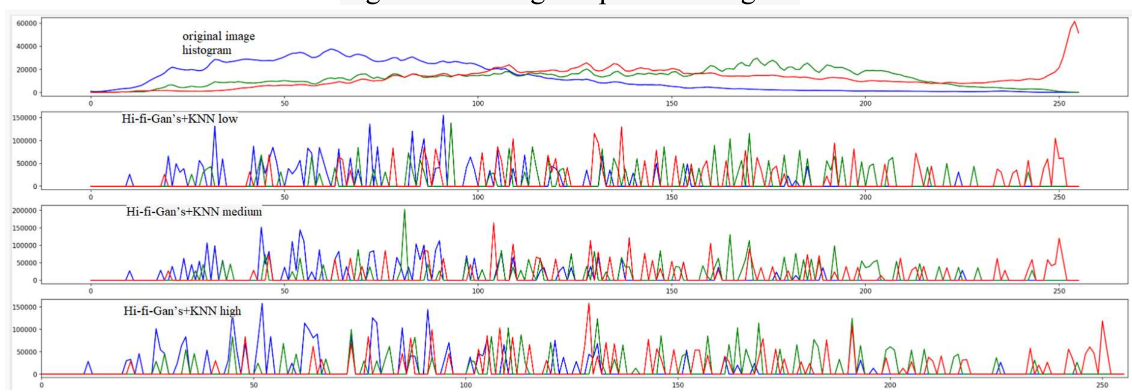


Figure 5 : Histogram plot for image 5.

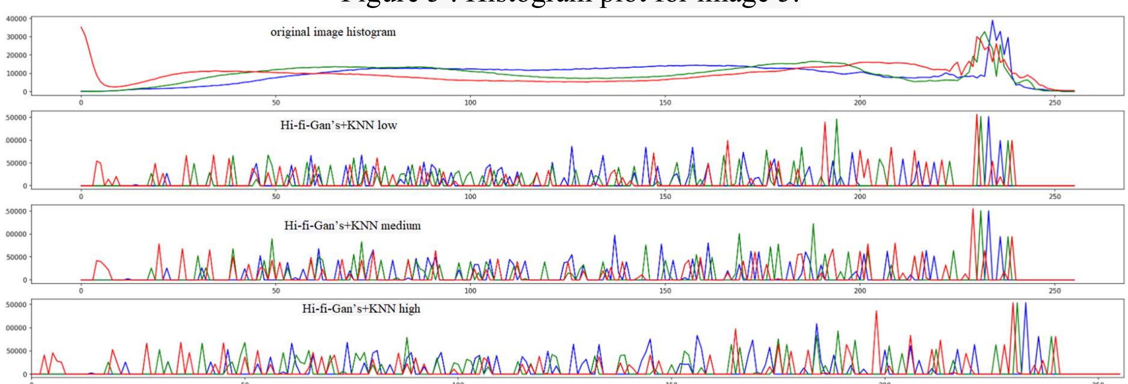


Figure 6 : Histogram plot for image 6.

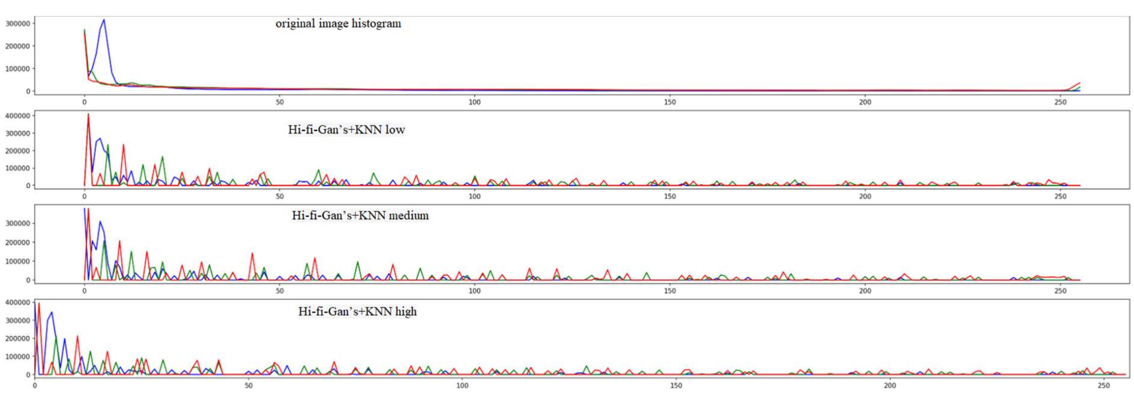


Figure 7 : Histogram plot for image 7.

		Mean Square Error	Structural Similarity Index Measure
Image 1	Hi-fi-Gan's+KNN low	69.30	0.86
	Hi-fi-Gan's+KNN medium	41.16	0.89
	Hi-fi-Gan's+KNN high	28.10	0.91
Image 2	Hi-fi-Gan's+KNN low	72.34	0.83
	Hi-fi-Gan's+KNN medium	71.52	0.88
	Hi-fi-Gan's+KNN high	43.94	0.91
Image 3	Hi-fi-Gan's+KNN low	87.55	0.83
	Hi-fi-Gan's+KNN medium	49.73	0.90
	Hi-fi-Gan's+KNN high	31.39	0.93
		48.73	0.79

Image 4	Hi-fi-Gan’s+KNN low	35.74	0.84
	Hi-fi-Gan’s+KNN medium	31.56	0.86
	Hi-fi-Gan’s+KNN high		
Image 5	Hi-fi-Gan’s+KNN low	50.12	0.83
	Hi-fi-Gan’s+KNN medium	47.59	0.84
	Hi-fi-Gan’s+KNN high	46.05	0.85
Image 6	Hi-fi-Gan’s+KNN low	84.51	0.77
	Hi-fi-Gan’s+KNN medium	82.25	0.83
	Hi-fi-Gan’s+KNN high	59.46	0.87
Image 7	Hi-fi-Gan’s+KNN low	85.99	0.81
		57.93	0.85

	Hi-fi-Gan's+KNN medium Hi-fi-Gan's+KNN high	45.78	0.87

Table 3. Test result for proposed system for different images

It is evident from table 3, that Hi-fi-Gan's+KNN high approach produces higher structural similarity index measure and lesser Mean square error when compared to Hi-fi-Gan's+KNN low and medium approach.

Conclusion:

In this paper a novel compression technique is proposed High-Fidelity Generative Image Compression Using GAN's and KNN," which itself branches in to three approaches: Hi-fi-GAN's+KNN High, Hi-fi-GAN's+KNN Medium, and Hi-fi-GAN's+KNN Low. The architecture of the proposed system was discussed , the learning rate as well as performance of the system is better as well as compression size and the clarity of the compressed image is more when compared to the approach that solely uses GAN, called "High-Fidelity Generative Image Compression," across three variations: Hi-fi-low, Hi-fi-medium, and Hi-fi-high. The architecture was designed with an encoder E, generator G, discriminator D and KNN classifier block. The Gan works on two parameters (x,s). x represents data point or picture location and s represents extra information from absolute location about features. In our work x is fixed as 7*7 (49 pixels). This is done to avoid multiple size of samples that can be generated by the generator G. In Gan network Y represents the total number of samples. The Y is also fixed in order to avoid non saturating loss. It is fixed as 120,240, 480 and 960. By implementing our approach instead of training the whole image , 7*7 (49 pixel) in one single location of the sample is trained, thereby reducing the permutation and accuracy is increased. In the discriminator part rectified linear unit and Leaky ReLU is employed to improve the accuracy. Our proposed architecture outperforms the High-Fidelity Generative Image Compression technique with Gan's network.

References

[1] Eirikur Agustsson, Fabian Mentzer, Michael Tschannen, Lukas Cavigelli, Radu Timofte, Luca Benini, and Luc V Gool. Soft-to-hard vector quantization for end-to-end learning compressible representations. In *Advances in Neural Information Processing Systems*, pages 1141–1151, 2017.

[2] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[3] Eirikur Agustsson, Michael Tschannen, Fabian Mentzer, Radu Timofte, and Luc Van Gool. Generative adversarial networks for extreme learned image compression. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

[4] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

[5] Johannes Ballé, Valero Laparra, and Eero P Simoncelli. End-to-end optimized image compression. *arXiv preprint arXiv:1611.01704*, 2016.

[6] Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston. Variational image compression with a scale hyperprior. In *International Conference on Learning Representations (ICLR)*, 2018.

[7] Fabrice Bellard. BPG Image format. https://bellard.org/bpg/.

[8] Mikołaj Binkowski, Dougal J Sutherland, Michael Arbel, and Arthur Gretton. Demystifying mmd gans. In *International Conference on Learning Representations*, 2018.

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

[10] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fifidelity natural image synthesis. In *International Conference on Learning Representations*, 2019.

[11] Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang. Image blind denoising with generative adversarial network based noise modeling. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[12] Rémi Coulom. Whole-history rating: A bayesian rating system for players of time-varying strength. In *International Conference on Computers and Games*, pages 113–124. Springer, 2008.

[13] Stéphane Coulombe and Steven Pigeon. Low-complexity transcoding of jpeg images with near-optimal quality using a predictive quality factor and scaling parameters. *IEEE Transactions on Image Processing*, 19(3):712–721, 2009.

[14] Thomas M Cover and Joy A Thomas. Elements of information theory. John Wiley & Sons, 2012.

[15] Leonardo Galteri, Lorenzo Seidenari, Marco Bertini, and Alberto Del Bimbo. Deep generative adversarial compression artifact removal. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4826–4835, 2017.

[16] Mark E Glickman. A comprehensive guide to chess ratings. American Chess Journal, 3(1):59–102, 1995.

[17] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Advances in neural information processing systems, pages 2672–2680, 2014.

[18] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In Advances in Neural Information Processing Systems, pages 6626–6637, 2017.

[19] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International Conference on Machine Learning, pages 448–456, 2015.

[20] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. In International Conference on Learning Representations (ICLR), 2017.
[21] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4401–4410, 2019.
[22] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. CoRR, abs/1412.6980, 2014.
[23] Kodak PhotoCD dataset. http://r0k.us/graphics/kodak/.
[24] Karol Kurach, Mario Lucic, Xiaohua Zhai, Marcin Michalski, and Sylvain Gelly. A large-scale study on regularization and normalization in gans. In International Conference on Machine Learning, pages 3581–3590, 2019.
[25] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image superresolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4681–4690, 2017.
[26] Jooyoung Lee, Seunghyun Cho, and Seung-Kwon Beack. Context-adaptive entropy model for end-to-end optimized image compression. arXiv preprint arXiv:1809.10452, 2018.
[27] Boyi Li, Felix Wu, Kilian Q Weinberger, and Serge Belongie. Positional normalization. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, Advances in Neural Information Processing Systems 32, pages 1622–1634. Curran Associates, Inc., 2019.
[28] Mu Li, Wangmeng Zuo, Shuhang Gu, Debin Zhao, and David Zhang. Learning convolutional networks for content-weighted image compression. arXiv preprint arXiv:1703.10553, 2017.
[29] Mario Lucic, Karol Kurach, Marcin Michalski, Sylvain Gelly, and Olivier Bousquet. Are gans created equal? a large-scale study. In Advances in Neural Information Processing Systems, pages 700–709, 2018.
[30] Detlev Marpe, Heiko Schwarz, and Thomas Wiegand. Context-based adaptive binary arithmetic coding in the h. 264/avc video compression standard. IEEE Transactions on circuits and systems for video technology, 13(7):620–636, 2003.
[31] Fabian Mentzer, Eirikur Agustsson, Michael Tschannen, Radu Timofte, and Luc Van Gool. Conditional probability models for deep image compression. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
[32] David Minnen, Johannes Ballé, and George D Toderici. Joint autoregressive and hierarchical priors for learned image compression. In Advances in Neural Information Processing Systems, pages 10771–10780, 2018.
[33] David Minnen and Saurabh Singh. Channel-wise autoregressive entropy models for learned image compression. arXiv preprint arXiv:2007.08739, 2020.
[34] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014.
[35] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. IEEE Signal Processing Letters, 20(3):209–212, 2012.

[36] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In International Conference on Learning Representations, 2018.

[37] Sahand Negahban, Sewoong Oh, and Devavrat Shah. Iterative ranking from pair-wise comparisons. In Advances in neural information processing systems, pages 2474–2482, 2012.

[38] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatiallyadaptive normalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2337–2346, 2019.

[39] Oren Rippel and Lubomir Bourdev. Real-time adaptive image compression. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2922–2930, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR.

[40] Shibani Santurkar, David Budden, and Nir Shavit. Generative compression. *arXiv preprint arXiv:1703.01467*, 2017.

[41] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[42] Gary J Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand. Overview of the high effificiency video coding (hevc) standard. *IEEE Transactions on circuits and systems for video technology*, 22(12):1649– 1668, 2012.

[43] David S. Taubman and Michael W. Marcellin. *JPEG 2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers, Norwell, MA, USA, 2001.

[44] Lucas Theis, Wenzhe Shi, Andrew Cunningham, and Ferenc Huszar. Lossy image compression with compressive autoencoders. In *International Conference on Learning Representations (ICLR)*, 2017.

[45] George Toderici, Sean M O'Malley, Sung Jin Hwang, Damien Vincent, David Minnen, Shumeet Baluja, Michele Covell, and Rahul Sukthankar. Variable rate image compression with recurrent neural networks. *arXiv preprint arXiv:1511.06085*, 2015.

[46] George Toderici, Lucas Theis, Nick Johnston, Eirikur Agustsson, Fabian Mentzer, Johannes Ballé, Wenzhe Shi, and Radu Timofte. *CLIC 2020: Challenge on Learned Image Compression*, 2020. http:// compression.cc.

[47] George Toderici, Damien Vincent, Nick Johnston, Sung Jin Hwang, David Minnen, Joel Shor, and Michele Covell. Full resolution image compression with recurrent neural networks. *arXiv preprint arXiv:1608.05148*, 2016.

[48] Michael Tschannen, Eirikur Agustsson, and Mario Lucic. Deep generative models for distributionpreserving lossy compression. In *Advances in Neural Information Processing Systems*, pages 5929–5940, 2018.

[49] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.

[50] Gregory K Wallace. The JPEG still picture compression standard. *IEEE transactions on consumer electronics*, 38(1):xviii–xxxiv, 1992.

[51] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. Highresolution image synthesis and semantic manipulation with conditional gans. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[52] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *Asilomar Conference on Signals, Systems Computers, 2003*, volume 2, pages 1398–1402 Vol.2, Nov 2003.

[53] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

[54] WebP Image format. https://developers.google.com/speed/webp/.

[55] Yuxin Wu and Kaiming He. Group normalization. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

[56] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015.

[57] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018.

[58] Fabian Mentzer, George Toderici,Michael Tschannen, Eirikur Agustsson. High-Fidelity Generative Image Compression. In 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada. arXiv:2006.09965v3 [eess.IV] 23 Oct 2020.