International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211
Volume 10, Issue 8, August-2022, Impact Factor: 7.429, Available online at: www.ijaresm.com

Image Compression with Learning using Generative Adversarial Network

Mohit[1], Dr. Kumud Saxena[2], Mr. Yaduvir Singh[3]

[1,2,3]Noida Institute of Engineering and Technology, Greator Noida, U.P., India

---*****************---

ABSTRACT

Traditionally image compression techniques such as JPEG is not designed specifically for the data being compressed, and therefore do not achieve the best possible compression rates for images. In this project, we aim to model a deep neural network based compression architecture using a GAN model. A GAN consists of two networks called the generator and the discriminator. During training, the combination of discriminator and generator compete with each other trying to find minimum of their adversarial loss function. To win the competition the generator tries to generate image pattern that must look like a base image on other hand the discriminator finds a way to correctly discriminate image as real or generated. The discriminator's main work is to optimize generator's quality to fully spoof the discriminator, generally discriminator is used during training and remains untouched training. Here we are building a Generative adversarial network based image compression by learning them that helps us to generate the image at low bitrates. The framework contains a combination of an encoder, generator, multi-scale discriminator. The details in the image that are not required to store during the encoding are synthesized and achieves visual pleasing outcomes. The method we proposed is able to generate synthesized image using a semantic label map that leads to the storage cost reduction, synthesize of objects may include streets, trees, etc.

Keywords: GAN Loss, Encoder, Quantizer, semantic label maps, synthesization.

1 INTRODUCTION

Now a days image compression has become very popular using deep neural networks and in short we can say it as a deep compression system because it uses deep neural network structure for the compression and decompression of an image and passing an image through a neural unit is itself a attractive area of research. The deep compression systems (e.g. [1, 2, 5, 6, 8]) are competitive against the modern state-of-the-art codec in comparison with the quality they show after the encoding, the modern codec may include BPG [7], JPEG2000 [9], WebP [10], etc. Besides the better results the deep compression systems show, these systems are even better on focusing a particular area in an image that makes them domain specific easily and provides efficiently processed promising results [11]. However each deep compression system uses their own distortion metrics for example the matrix may be PSNR, MS-SSIM, etc. [12]. Nowadays faster image access is required for everyone with better image quality and to obtain faster image access low bit rates are necessary to decode the image faster and optimal bitrates are required to obtain the both, thereby we are working on model that compresses an image to a very low bitrate around 1/10 bits per pixel but for the training purpose we are using adversarial loses because the traditional lose/distortion metrics favor the preservation of pixel-wise entropy instead of preserving global structure and the texture layout of an image. To capture global structure and texture details new kind of distortion metrics must be used, a best used method is adversarial losses [13], the adversarial losses have shown very promising results towards this and helps in developing powerful generator having good performance in the generation of visually like-able results also with semantic maps. Here we are working to propose a generative adversarial network framework based on dedicated principles with advantageous behavior to compress FHD images and along with this trying to obtain an image with bitrate of 1/10 bits per pixel. We are using two modes of GAN, these two modes are conditioned GAN and non-conditioned GAN [14, 13] with their certain features, the generative compression kind of network may be used as a normal generative compression that includes the preservation of whole image structure and helps in generating different structures having different level of scales. And the second mode includes generative compression with selected contents generating from semantic label maps completely with the preservation of region that user wants to define/keep in an image having high detailed region, so called selective generative compression. Selective generative compression usually needs semantic label map or instance maps that are viable through existing instance networks e.g., MASK R-CNN [15], PSPNet [16], these networks store maps as a graphics vector. The generation of contents in an image using semantic maps leads generator to generate contents having low bitrate in comparison to other compression (BPG [7]) and more the generation of area through semantic maps leads to save more storage space thereby giving significant storage cost reduction. The proposed framework can reduce the bitrate atleast half of the orginal bitrate without degrading the quality of an image.

2 RELATED WORK

There are most popular deep neural networks[1, 2, 3, 4, 11, 17, 6] are auto encoders and other neural network e.g., recurrent neural networks [18, 19] are to date DNN structures. The aforementioned architectures perform their tasks by converting the input image in a stream of bits and after that the bit stream is used to perform lossless compression using entropy coding for example using arithmetic coding. There are other models that rely on context modeling to catch the distribution in an image of bit stream [2, 19, 4, 5, 8]. In an unsupervised manner generative adversarial networks [13] are strong technics to find the distribution of bits/pixel values/probability that are hard to find. The GANs are showing great capabilities towards generating more natural and clear images than other known prior used approaches e.g., [20, 21, 3, 23]. Other approaches are conditioned [13, 14] based generative adversarial networks these networks best approachable towards image to image translation [24, 25, 26, 27]. One other model defined by Santurkar *et al.* [28] in which a generative model is made to learn image thumbnails and uses that model as a decoder for compression based on image thumbnails.

3 BACKGROUND

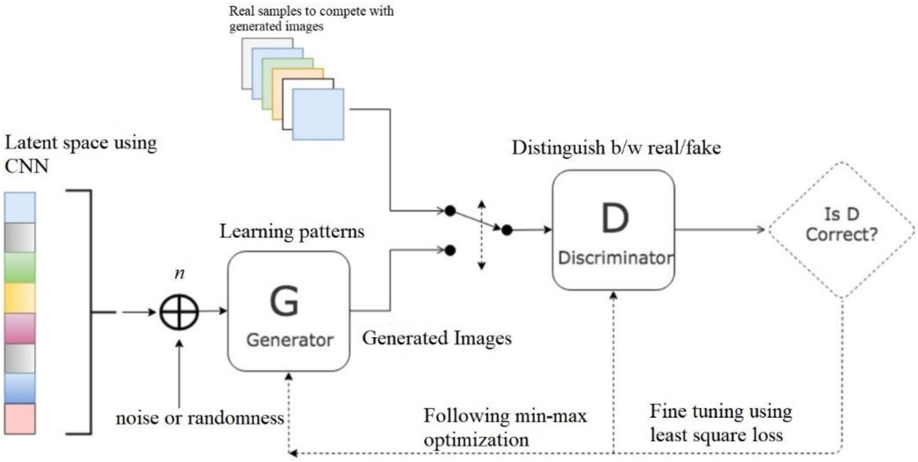

Fig. 1. GAN Arch.

GANs, for a given dataset X, the generative adversarial network is capable of finding unknown distribution p_x using a generator having training to map samples (denoted as symbol (r)) to distribution p_x and the generator is denoted as $\left(G(r)\right)$, the samples r are drown from a prior, know and fixed distribution p_r. Here in generative adversarial network both the discriminator and the generator are required to be trained simultaneously to find min-max point of the objective of min-max GAN/least GAN [30]

$$L_{GAN} \coloneqq max_D E\left[f\left(D(x)\right)\right] + E\left[g\left(D\left(G(r)\right)\right)\right] \quad (1)$$

Here generator and discriminator are denoted as G, D respectively and both are deep neural networks, and the functions g, f are considered from Nowozin et al. [29] tells that g, f during the solution of equation(1) must be taken care for their suitable values because their suitable values helps in minimization of f-divergence between p_x and $G(z)$. Here we are going to use least square generative network from [23], having $g(y) = y^2$and $f(y) = (y-1)^2$ following the properties of Pearson chi-square divergence. [30]

To execute the the task of compression let's take an image $x \in X$ accepting the equations from [22, 8], using them to make a model called encoder learnable denoted as E and also a model to decode is a decoder also called a generator G, and to make output more finite a quantizer is also made denoted as q. Preceding with the encoder, applying it on an image to map the image into a latent space by means of E that gives us latent space of image denoted as w. Now feeding it to the quantizer to generate a quantized representation $\hat{w} = q(w)$, and it can be used as a bitstream. To generate/reconstruct an image from the $\hat{w}$ the D(decoder) is applied to $G(\hat{w})$, producing $\hat{x}$. Now the decisive decision for bits required to encode $\hat{w}$ to convert it into bitstream is possible by means of entropy denoted as $h(\hat{w})$, that can be modeled through [8, 22]. The high quality bitrate versus distortion is usually a trade off between generated image quality and bit rate, the trade off is given as

$$E[d(x, \hat{x})] + \beta H(\hat{w}) \quad (2)$$

Here $E[d(x,\hat{x})]$ term denotes the similarity measure between x and $\hat{x}$ or one can say a loss to the reconstructor. As during trainig weight updatation is required through β of the equation(2) and to get optimal bitrates differentiable and suitable variable quantisation levels are used, denoted via $L = q_1, q_2, q_3$ .... Also to create bitstream of $\hat{w}$ differentiable entropy is maximized only upto dimensions of quantised feature map and also by quantisation levels,

$$dim\big(q(w)\big)log_2 L \geq H\big(q(w)\big) \quad (3)$$

It can be seen as a joining of compression based on learning and condition dependent generative network.

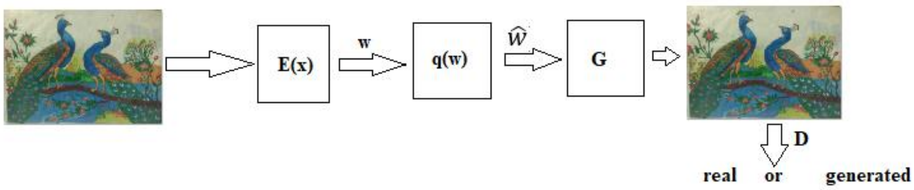

Fig. 2. Model Arch.

From the above fig. 2 a brief structure of the workflow can be assumed, here the output is further added up with a certain degree of noise to make it spoof proof and to build a better generator, the noise is basically accepted from p_r (fixed prior probability) and producing a hidden vector denoted as r. Now the the generator G attempts to construct an image from the latent vector (r) by means of $G(r)$, the constructed image denoted as $x^{'}$ having consistency with the p_x. The objective of min-max Gan loss to find the min max points is expressed as

$$min_{e,G} L_{Gan} + \lambda E[d(x,x^{'})] + \beta H(\hat{w}) \text{ [30]} \quad (4)$$

The term lambda in equation (4) stablizes the distortion against entropy term and generative adversarial loss. The important feature of equation (4) due to the elements of original image in the form of $\hat{w}$ that makes the equation saving the some details about base image. Here the addition of entropy and the distortion term generates an important effect on the model's learning capability because of the bitrate controller β, if we set bitrate controller to zero then the quantized representation of the base image would contain nearly the same bitrate or some additional learned bits and the automatically learned dimensions that will nearly reconstruct the image to the original image losslessly, due to this, the distortion term will have no better imapct on preserving/synthesizing image contents and also GAN loss effect would become uneffective. The second case arise for controlling the entropy to original distribution of $\hat{w}$ helps in making the distribution of quantized bitstream of x determinstic and the loss term behaves like regularizer through this the purpose of adding some randomness r is fullfilled, randomness/noise is independent of x helps in building a standard generative network.

Though by applying constraints on $H(\hat{w})$, the encoder and the generator will not be able to surpass the loss factor, under this condition only remaining way for G, E is to maintain the objective of equation (1) and the distortion factor of equation (4), balancing these two factors make generative network adaptable to generate nearly original contents and along with preserving original contents. By the means of above training with distortion factor, leads to the balanced generative network and helps to remove mode collapse. Proceeding with semantic maps further helps in reducing bitrates because more the uses of semantic maps increases chance on synthesizing more conrents. The sementic maps are only fed to the discriminator during period of training. Here the semantic label maps behave as a additional information to the generative network model.

DISCUSSION AND CONCLUSION

We building a generative adverssarial network with learning image capability to distinguish different areas in an image and this learned image pattern helps in preserving user desired details with low bitrates by means of introducing quantization and entropy module. Convincingly the generative compression method provides results in finer details. In practice the discriminator moves towards minimization for the $(D(x)-1)^2 + \Big(D\big(G(r)\big)\Big)^2$ for both generated and real images. Through the least loss function the generator move towards minimization of for images that were real by means of difference between expected value and predicted. Usually this task involves balancing 1/0 class labels for real or fake images respectively, helping in minimization of least square loss. The advantage of least loss is that it follows the rule of more the error, more the penalty, thereby producing large value correction instead of vanishing gradient.
The future work may include the better implementation of the current work using better workflow, methods and also having the a different lower bitrate color format for each pixel for advance in image compression. Another future advancement might be the opportunity to perform preservable region or synthesizable region detection for better encoding

and decoding, further research may involve spatial bits prediction or spatial coordinates measurement for different region to preserve for better accuracy, quality and performance, etc.

REFERENCES

[1] Lucas T., Wenzhe S., Andrew C., Ferenc H.: Lossy image compression with compressive autoencoders. In International Conference on Learning Representations (ICLR), 2017.
[2] Johannes B., Valero L., and Eero P Simoncelli: End-to-end optimized image compression. In International Conference on Learning Representations (ICLR), 2017.
[3] Eirikur A., Fabian M., Michael T., Lukas C., Radu T., Luca B., and Luc van Gool: Soft-to-hard vector quantization for end-to-end learning compressible representations. In Advances in Neural Information Processing Systems, pp. 1141–1151, 2017.
[4] Mu L., Wangmeng Z., Shuhang G., Debin Z., David Z.: Learning convolutional networks for content-weighted image compression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3214–3223, 2018.
[5] Oren R., Lubomir B.: Real-time adaptive image compression. In Proceedings of the International Conference on Machine Learning, Proceedings of Machine Learning Research, pp. 2922–2930, 2017.
[6] Johannes B., David M., Saurabh S., Sung J. H., Nick J.: Variational image compression with a scale hyperprior. In International Conference on Learning Representations (ICLR), 2018.
[7] Leonardo G., Lorenzo S., Marco B., Alberto D. B.: Deep generative adversarial compression artifact removal. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4826–4835, 2017.
[8] Fabian M., Eirikur A., Michael T., Radu T., Luc Van Gool: Conditional probability models for deep image compression. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
[9] David S. T., Michael W. M.: JPEG 2000: Image Compression Fundamentals, Standards and Practice. Kluwer Academic Publishers, Norwell, MA, USA, 2001.
[10] Web. WebP Image format. https://developers. google.com/speed/webp/.
[11] Robert T., Fabian M., Eirikur A., Michael T., Radu T., Luc Van Gool: Towards image understanding from deep compression without decoding. In International Conference on Learning Representations (ICLR), 2018.
[12] Z. Wang, Eero. P. S., A. C. Bovik: Multiscale structural similarity for image quality assessment. In Asilomar Conference on Signals, Systems Computers, pp. 1398–1402, 2003.
[13] Ian G., Jean Pouget-Abadie, Mehdi M., Bing X., David Warde-Farley, Sherjil O., Aaron C., Yoshua B.: Generative adversarial nets. In Advances in Neural Information Processing Systems, pp. 2672–2680, 2014.
[14] Mehdi M., Simon O.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014.
[15] Kaiming H., Georgia G., Piotr D., Ross G.: Mask r-cnn. In IEEE International Conference on Computer Vision (ICCV), pp 2980–2988, 2017.
[16] Hengshuang Z., Jianping S., Xiaojuan Q., Xiaogang W., Jiaya J.: Pyramid scene parsing network. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
[17] David M., Johannes B., George D. T.: Joint autoregressive and hierarchical priors for learned image compression. In Advances in Neural Information Processing Systems, pp. 10771–10780, 2018.
[18] George T., Sean M O'Malley, Sung J. H., Damien V., David M., Shumeet B., Michele C., Rahul S.: Variable rate image compression with recurrent neural networks. arXiv preprint arXiv:1511.06085, 2015.
[19] George T., Damien V., Nick J., Sung J. H., David M., Joel S., Michele C.. Full resolution image compression with recurrent neural networks. arXiv:1608.05148, 2016.
[20] Tim S., Ian G., Wojciech Z., Vicki C., Alec R., Xi C.: Improved techniques for training gans. In Advances in Neural Information Processing Systems, pp. 2234–2242, 2016.
[21] Martin A., L´eon Bottou: Towards principled methods for training generative adversarial networks. arXiv:1701.04862, 2017.
[22] Martin A., Soumith C., L´eon B.: Wasserstein generative adversarial networks. In Proceedings of the International Conference on Machine Learning, pp. 214–223, 2017.
[23] Xudong M., Qing L., Haoran X., Raymond Y. L., Zhen W., Stephen P. S.: Least squares generative adversarial networks. In IEEE International Conference on Computer Vision (ICCV), pp. 2813– 2821, 2017.
[24] Phillip I., Jun-Yan Z., Tinghui Z., Alexei A E.: Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134, 2017.
[25] Ting-Chun W., Ming-Yu L., Jun-Yan Z., Andrew T., Jan K., Bryan C.: High-resolution image synthesis and semantic manipulation with conditional gans. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
[26] Jun-Yan Z., Taesung P., Phillip I., Alexei A E.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2223–2232, 2017.

[27] Ming-Yu L., Thomas B., Jan K.: Unsupervised image-to-image translation networks. In Advances in Neural Information Processing Systems, pp. 700–708, 2017
[28] Shibani S., David B., Nir S.: Generative compression. Picture Coding Symposium (PCS), 2018.
[29] Sebastian N., Botond C., Ryota T.: f-GAN: Training generative neural samplers using variational divergence minimization. In Advances in Neural Information Processing Systems, pp. 271–279, 2016.
[30] Eirikur A., Michael T., Fabian M., Radu T., Luc V. G.: Generative Adversarial Networks for Extreme Learned Image Compression. pp. 221-224, 2019.