International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Customized Approach to Compress the Images with Deep Learning Model for Embedded System

[1]Abhiram Potlapalli, [2]Dr. Seetharam Khetavath,

Submitted: 05/02/2024 **Revised**: 13/03/2024 **Accepted**: 19/03/2024

Abstract: The automated systems like embedded systems will work autonomously, take decisions and actions independently. And they will use small resources like compressed memory unit and processing. But if we want intelligent system then we should integrate Machine Learning (ML) then the agent will take decisions and actions its own be fully autonomous. But Theses embedded systems used in scientific community required high computational speed Despite this when we are working with images, the low capacity of embedded systems greatly hinders this integration, so the possibility of being able to integrate them into a wide range of micro-controllers can be a great advantage. So this paper implements a Customized intelligent system to take image and compress the size of it, and send it to embedded system to take intelligent decisions. The proposed system is competitive if compared to other commercial systems with optimal results.

Keywords*: Customized CNN, Compressing Images, Embedded systems, deep learning, image processing.*

1. Introduction

The integration of digital technologies, such as the Internet of Things (IoT), Big Data, and artificial intelligence (AI), has significantly transformed manufacturing industries and automation processes in recent years. This evolution has led to a more interconnected and optimized value chain, allowing for real-time adjustments and improvements. Key concepts like IoT enable the deployment of sensors throughout manufacturing plants, providing a constant stream of data that can be leveraged for critical actions.

However, the challenge arises in automating decision-making processes to ensure timely and accurate responses to the information gathered. While experts in the field possess the knowledge to define the necessary conditions for optimal actions, relying solely on human decision-making can introduce delays. This necessitates the implementation of intelligent agents equipped with AI capabilities to make decisions based on real-time data autonomously.

One specific application of AI in manufacturing involves image compression. Traditional hand-crafted algorithms for image compression have limitations, as they need a deeper understanding of the content they compress. This limitation has led to the exploration of deep neural networks, which have shown promise in achieving higher compression rates due to their ability to learn and comprehend the content of images.

[1]*Research Scholar, Department of ECE, Chaitanya Deemed to be University, Hanamkonda, Warangal, Telangana, India*
[2]*Professor, Department of ECE, Chaitanya Deemed to be University, Hanamkonda, Warangal, Telangana, India*
Author Email : abhiram.potlapalli@gmail.com

Previous research efforts, such as those utilizing Gaussian models and pattern recognition methods, have focused on image compression but struggled to prevent loss. On the other hand, recent studies, including works by researchers mentioned [3], [9], [10], [13], [22], have implemented deep learning models with various filters and layers, achieving improved results. However, these systems still face challenges in achieving content-based compression.

A notable advancement in this area is the implementation of vision transformer models with attention layers, as demonstrated in the research cited [11]. These models offer a more optimal approach to image compression by leveraging attention mechanisms to focus on relevant content, addressing the limitations of previous methods. As manufacturing industries continue to embrace digitalization, the intersection of AI and image compression technologies is poised to play a crucial role in enhancing efficiency and decision-making processes across the value chain.

Researchers have explored various strategies to improve compression efficiency and overcome the limitations of lossy compression. From autoregressive context models to checkerboard context models, generative compressed approaches, and convolution neural networks (CNNs), the focus has been on achieving compression without sacrificing prediction accuracy and perceptual quality. Integrating attention mechanisms, recurrent neural networks, and transformer-based models like Vision Transformer and Swin Transformer has brought about novel solutions, demonstrating advancements in capturing spatial dependencies and global-related feature learning.

Contribution

Proposed optimized GAN model with stabilized loss to add and compress the images.

Our model can use vary less number of parameters and it provides loss less compression.

Our model is robust we tested the model with various data sets and consistently performing.

Related Work

The quest for efficient image compression techniques has spurred diverse approaches from various researchers, each aiming to strike a balance between reducing image size and resolution while minimizing information loss. Traditional methods often relied on machine learning (ML) algorithms such as principal component analysis, clustering, and supervised dimensionality reduction models. However, many of these approaches resulted in compressed images with noticeable loss.

Some researchers explored ANN in pursuing non-lossy compression. For instance, in [1], an ANN was employed, utilizing signal-to-noise ratio as a metric for compression. However, the authors highlighted limitations related to feature extraction and other considerations.

Another innovative approach, as described in [2], involved a generative compressed approach with a decoder designed to capture both positive and negative relationships in image compression. This method allowed for the effective reconstruction of images.

With advancements in deep learning, the researchers used CNN and RNN models for image compression. Ballé, J., Laparra, V., in [3], proposed a CNN model to compress samples end-to-end, introducing a generalized loss function based on the likelihood of a generative model. This method provided a systematic way of compressing images while maintaining their quality. In [5], an ANN model comprising a hyperpriority was proposed. This model effectively captured spatial dependencies in the latent representation of images, utilizing the popular MS-SSIM index for evaluation. The results illustrated superior rate-distortion performance compared to other ANN-based methods, showcasing the potential of this approach in achieving high-quality compression.

A different avenue was explored in [4], where a generative adversarial network (GAN) and a multi-scale discriminator were implemented to learn compression parameters. This approach identified important and unimportant regions in an image, enabling compression in the unimportant regions. The content-based approach reduced storage costs and demonstrated a nuanced understanding of image content.

He, D., Zheng, Y., et al in [7], introduced an autoregressive context model and a checkerboard context model that reorganizes decoder order, reducing decoding time and fitting into devices without sacrificing prediction accuracy. In [8], the authors suggested that the performance of such models could be enhanced by leveraging spatial-channel dependencies in latent space and optimizing context adaptivity.

The works in [9, 10] focused on visually pleasing reconstructions that are perceptually similar to input images across a range of bitrates. These approaches bridge the gap between rate-distortion-perception theory and practice, evaluating their methods quantitatively with various perceptual metrics. Rippel and Bourdev [11] implemented systematic transformers based on Vision Transformer (ViT) and Swin Transformer, emphasizing attention layers to meet expectations in image compression.

Toderici et al. [13] explored recurrent neural networks (RNNs) and long short-term memory (LSTM) models with additive reconstruction architectures, introducing a scaled-additive framework that demonstrated significant improvements. Some studies, like [12] and [15], concentrated on simple intelligent models prioritizing image size over content-based compression. On the other hand, Gaussian distribution methods [16] and Gaussian mixture models [18] were implemented for compression but did not provide lossless results.

Low bit-rate-based deep learning models were proposed in [19, 21], with a CNN model for semantic-based image compression presented in [20]. Medical image compression, focusing on lossless compression, was addressed by Nagoor in [22] using a CNN model. In [23], pattern recognition models were employed for lossless image compression.

The transition from old machine learning models to advanced vision transformer models reflects the evolution in image compression research. Many researchers have worked on addressing the challenge of content-based image compression without sacrificing image content. The field continues to evolve, emphasizing novel techniques prioritizing compression efficiency and preserving essential image information.

These diverse approaches highlight the dynamic landscape of image compression research, showcasing a shift towards deep learning techniques and a quest for methods that go beyond traditional lossy compression, aiming to preserve as much information as possible while efficiently reducing image size.

Methodology

We have implemented an optimized GAN [6] model for image compression. The model takes images as input and outputs compressed versions while ensuring the preservation of data through the addition of noise from the generator to the discriminator. Figure 1 serves as an illustrative representation of the proposed model architecture.

This innovative approach uses the GAN framework to achieve image compression without data loss. The generator introduces controlled noise during the compression process, and the discriminator plays a crucial role in evaluating the compressed images. This dynamic interplay between the generator and discriminator, as showcased in our proposed model, highlights the potential for novel and effective solutions in image compression.

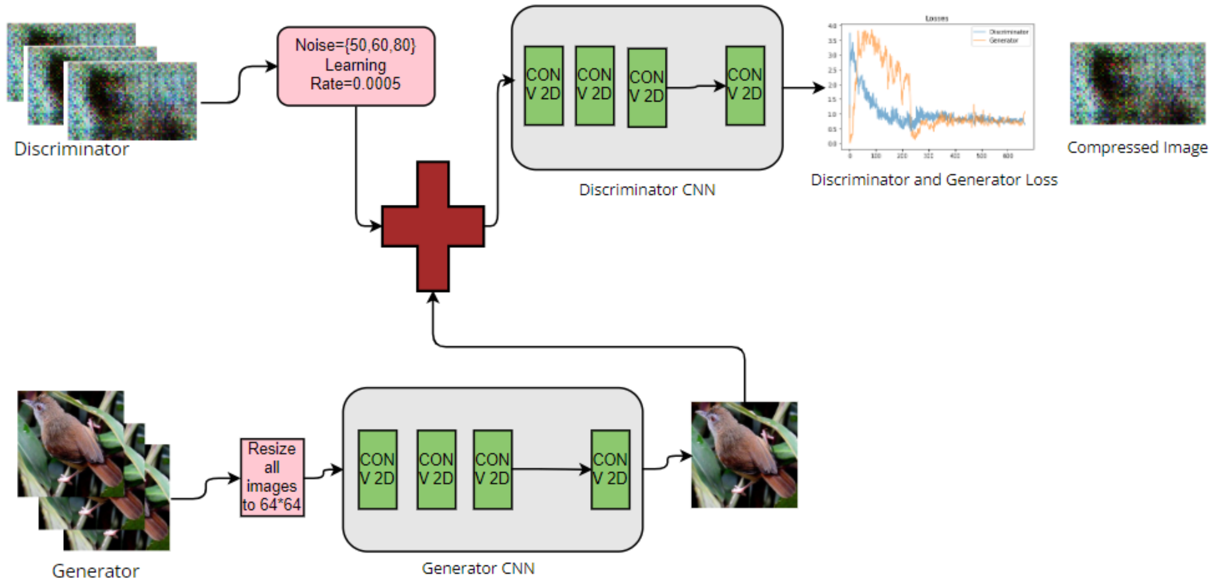

Figure 1 Propose G-D model

Data analysis and preprocessing

The dataset originates from Kaggle and comprises over 200 bird images. As part of the preprocessing steps, these images were resized to dimensions of 64x64, as illustrated in Figure 2. Resizing is a common practice in deep learning to standardize input dimensions and reduce computational complexity. However, it is essential to note that resizing may result in some loss of information, and the dataset's diversity and quality may influence the model's effectiveness.

Figure 2 samples used for compression after resizing to 64*64

I

mplementation

We developed a deep learning model, delineated in Table 1 that has been meticulously crafted to address the task of compressing images without compromising their inherent content. The initial preprocessing steps involve converting images to the RGB format and resizing them to dimensions of 64x64x3. The generator's architectural framework is noteworthy, featuring six convolutional layers accompanied by max pooling and batch normalization, succeeded by four densely connected layers. The intermediate CONV-2D layers employ the rectified linear unit (ReLU) activation function, while the final layer deploys the softmax activation function, yielding an output dimension of 3x1. Simultaneously, the discriminator, equipped with six CONV-2D layers, processes the output from the generator, ultimately producing images of a standardized 64x64 size. The convolutional operation f(x,y) (1) involves multiplying filter elements with their corresponding counterparts in the input image, followed by a summation process that yields scalar values at each spatial coordinate (x, y). Using softmax activation in the generator hints at a categorical output, potentially associated with specific image classes or channels. To optimize model efficacy, crucial considerations include implementing judicious loss functions, a well-balanced training strategy, explicit evaluation metrics, hyperparameter refinement, and potential integration of data augmentation techniques. The model's ultimate success hinges on its capacity to generalize to unseen data, underscoring the significance of robust training and meticulous evaluation practices in achieving optimal performance.

$$f(x, y) = \sum (I * W)(x, y) \qquad (1)$$

The pooling layers reduced the spatial dimensions of the feature maps through down sampling techniques like max pooling (equation 2). The pooling operation (Pool) occurred at spatial coordinates (x, y) to generate the resulting pooled feature maps. With equation (3) and (4) calculates the predicted values and finds the loss.

$$p(x, y) = Max\_Pool(f)(x, y) \qquad (2)$$

$$y = f(wx + b) \qquad (3)$$

$$loss(y_{true}, y_{pred}) = \frac{1}{m} - \sum_{c=1}^{C} (y_c log(p_c)) \qquad (4)$$

Table 1 layer wise parameters of generator and discriminator models

Generator Model	Discriminator Model
Input image 64*64*3	Input 1*8192
Conv 2D 32*32*32	Conv 2D 4*4*512
Batch Normalization	Conv 2D 8*8*512
Conv 2D 16*16*64	Batch Normalization
Batch Normalization	Conv 2D 16*16*256
Conv 2D 16*16*64	Batch Normalization
Conv 2D 8*8*128	Conv 2D 32*32*128
Conv 2D 8*8*256	Conv 2D 32*32*128
Batch Normalization	Conv 2D 64*64*64
Conv 2D 4*4*512	Conv 2D 64*64*64
Flatten(1*8192)	Batch Normalization
Dense 1	Output 64*64*1

In GAN Loss of Discriminator is $L_D = Error(D_x, 1) + Error(D(G_x, 0))$ and loss function for Generator is $L_G = Error(D(D_x), 1)$ from this with cross entropy loss function we optimized the both the losses with respect toN_{sp}.

$$\frac{p_{data}(x)}{D_x} = N_{sp} + \frac{p_g(x)}{1-D_x} \qquad (1)$$

$$\frac{p_{data}(x)}{D_x} = \frac{N_{sp}-N_{sp}\,D(x)+p_g(x)}{1-D_x} \qquad (2)$$

$$D_x = \frac{(1-D_x)p_{data}(x)}{N_{sp}-N_{sp}D_x+p_g(x)} \qquad (3)$$

$$V(G, D) = E_{x-p_{data}}[log\, D_x] + E_{x-p_g}[log(1 - D_x] \qquad (4)$$

$$V(G, D) = E_{x-p_{data}}\left[log\left(\frac{(1-D_x p_{data}(x)}{N_{sp}-N_{sp}D_x+p_g(x)}\right)\right] + E_{x-p_g}\left[log\left(1 - \frac{1-D_x p_{data}(x)}{N_{sp}-N_{sp}D_x+p_g(x)}\right)\right] \qquad (5)$$

N_{sp} Represents the noised data in the provided framework, constrained within the range $0 < N_{sp} \le 1$. Equation (5) is the derived GAN optimizer. In this context, x denotes the real data, G_x represents synthetic data generated by the model, D_x signifies the discriminator evaluation for real data, and D(G_x) reflects the generator's evaluation for synthetic data. The schematic depiction in Figure 1 elucidates the functioning of the GAN in the context of image compression.

The generated and discriminated losses were initially set at 0.0004 and 0.004, respectively. After iteratively experimenting with various values, adjustments were made to optimize these parameters. The noise level, denoted as 50, 60, 80, and 100, was also systematically tuned. Notably, when the noise level was set to 100, the model could provide accurate values within a relatively short duration.

Different batch sizes of 8, 16, and 32 were employed during training. It was observed that for batch sizes of 16 and 32, the model tended to become overfitted. Conversely, with a batch size of 8, the model demonstrated its best performance. This suggests that careful consideration and fine-tuning of hyperparameters, such as noise level and batch size, are crucial for achieving optimal results in GAN-based image compression.

Result analysis

This model has undergone comprehensive training with a diverse set of hyperparameters. The initial learning rates for the generator and discriminator were set at 0.004. The noise levels were systematically adjusted, including values 50, 60, and 100. Additionally, various batch sizes were employed during training, and meticulous selection was made to identify the hyperparameter configurations that yielded optimal performance.

Figure 3 and 4 Observations reveal that higher noise levels, specifically at 100 and above, result in less clear generator samples. The generator and discriminator losses are recorded at 0.5879 and 1.059, respectively. However, as the noise is gradually reduced, a notable improvement in sample clarity is observed, with the generator and discriminator losses overlapping.

Further insights from Figures 5 and 6 underscore the importance of precisely adjusting noise levels and batch sizes to achieve the best fit. It is evident that when these parameters are appropriately tuned, the model exhibits optimal performance. Conversely, improper adjustments lead to overfitting, highlighting the sensitivity of the model's performance to the chosen hyperparameters. This iterative exploration and fine-tuning process underscore the importance of a systematic approach to hyperparameter optimization in GAN-based models for image compression.

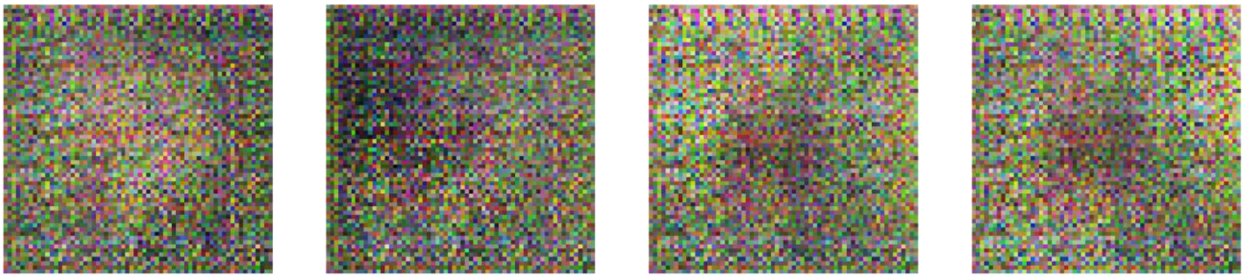

Figure 3 samples with high noise given by discriminator, when discriminator loss is 0.58790 and generator loss is 1.059

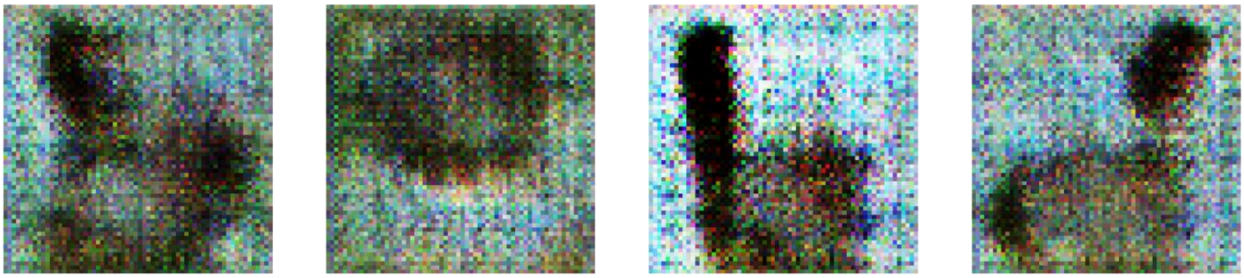

Figure 4 samples with low noise given by discriminator, when discriminator loss is 0.517 and generator loss is 1.361

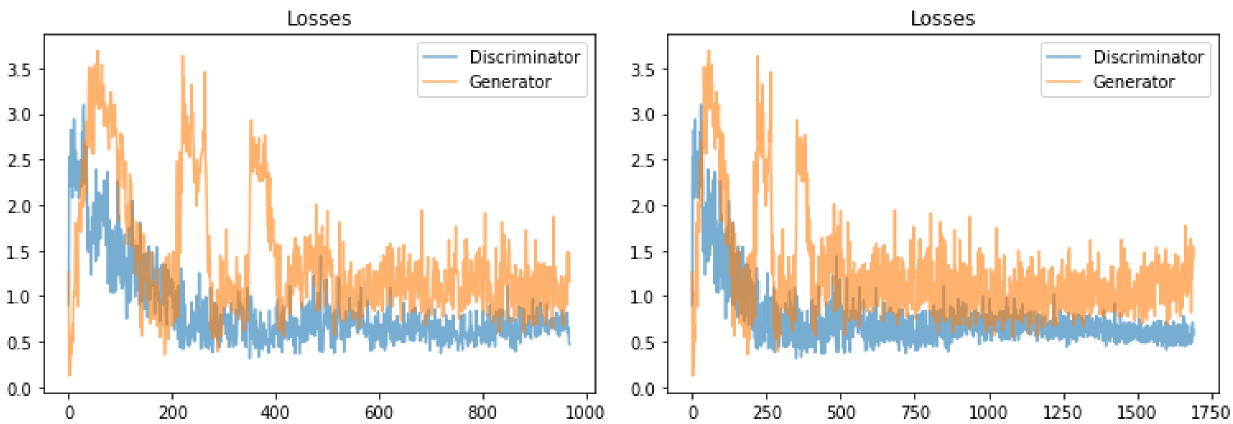

Figure 5 generator and discriminator losses epoch by epoch where batch size is 16

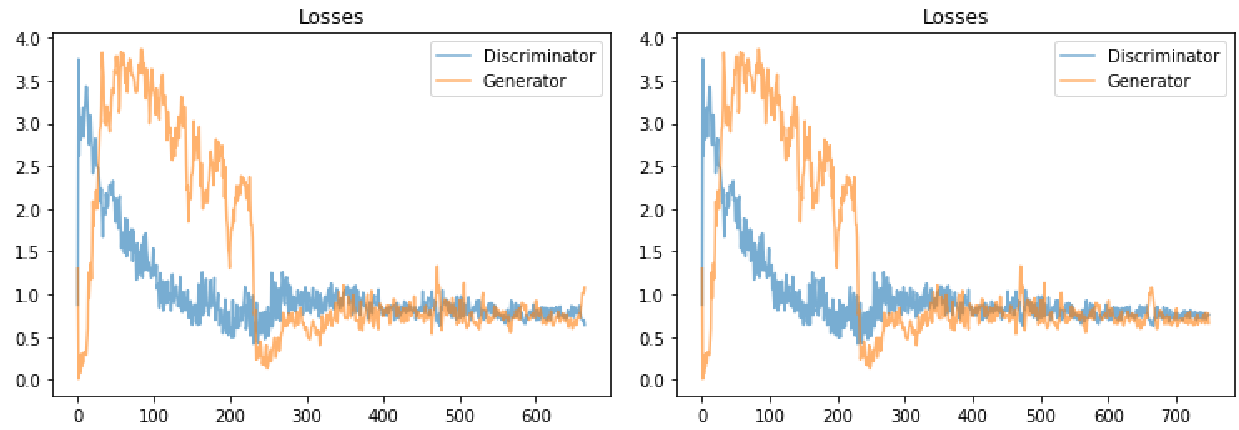

Figure 5 generator and discriminator losses epoch by epoch where batch size is 8

Conclusion

Implementing an optimized GAN for image compression using the Kaggle bird image dataset significantly advances preserving image content during compression. This approach involves training a generator and discriminator to reconstruct images from scratch, employing a competitive setup where one component minimizes the likelihood of images after compression, and the other minimizes the likelihood after compression. The Kaggle bird image dataset provides a diverse and practical testing ground, encompassing variations in bird species, backgrounds, and lighting conditions.

The essential observation is that the implemented model successfully reduces image size without perceptible loss of content. This achievement validates the efficacy of the GAN-based approach in content-preserving compression, a crucial consideration in image compression applications where maintaining visual fidelity is paramount. The work contributes to the evolving landscape of image compression techniques, showcasing the potential of optimized GANs beyond traditional applications.

References

[1] Lee, J., Cho, S., & Beack, S. K. (2018). Context-adaptive entropy model for end-to-end optimized image compression. arXiv preprint arXiv:1809.10452.

[2] Agustsson, E., Minnen, D., Toderici, G., & Mentzer, F. (2023). Multi-realism image compression with a conditional generator. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 22324-22333).

[3] Ballé, J., Laparra, V., & Simoncelli, E. P. (2016). End-to-end optimized image compression. arXiv preprint arXiv:1611.01704.

[4] Agustsson, E., Tschannen, M., Mentzer, F., Timofte, R., & Gool, L. V. (2019). Generative adversarial networks for extreme learned image compression. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 221-231).

[5] Ballé, J., Minnen, D., Singh, S., Hwang, S. J., & Johnston, N. (2018). Variational image compression with a scale hyperprior. arXiv preprint arXiv:1802.01436.

[6] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

[7] He, D., Zheng, Y., Sun, B., Wang, Y., & Qin, H. (2021). Checkerboard context model for efficient learned image compression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 14771-14780).

[8] Koyuncu, A. B., Gao, H., Boev, A., Gaikov, G., Alshina, E., & Steinbach, E. (2022, October). Contextformer: A transformer with spatio-channel attention for context modeling in learned image compression. In European Conference on Computer Vision (pp. 447-463). Cham: Springer Nature Switzerland.

[9] Mentzer, F., Toderici, G. D., Tschannen, M., & Agustsson, E. (2020). High-fidelity generative image compression. Advances in Neural Information Processing Systems, 33, 11913-11924.

[10] Yang, R., Van Gool, L., & Timofte, R. (2021). Perceptual learned video compression with recurrent conditional GAN. arXiv preprint arXiv:2109.03082, 1.

[11] Zou, R., Song, C., & Zhang, Z. (2022). The devil is in the details: Window-based attention for image compression. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 17492-17501).

[12] Rippel, O., & Bourdev, L. (2017, July). Real-time adaptive image compression. In International

Conference on Machine Learning (pp. 2922-2930). PMLR.

[13] Toderici, G., Vincent, D., Johnston, N., Jin Hwang, S., Minnen, D., Shor, J., & Covell, M. (2017). Full resolution image compression with recurrent neural networks. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (pp. 5306-5314).

[14] Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE transactions on image processing, 26(7), 3142-3155.

[15] Lin, F., Sun, H., Liu, J., & Katto, J. (2023, June). Multistage spatial context models for learned image compression. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.

[16] Luo, W., Li, Y., Urtasun, R., & Zemel, R. (2016). Understanding the effective receptive field in deep convolutional neural networks. *Advances in neural information processing systems*, *29*.

[17] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[18] Fu, H., Liang, F., Lin, J., Li, B., Akbari, M., Liang, J., ... & Han, J. (2021). Learned image compression with discretized gaussian-laplacian-logistic mixture model and concatenated residual modules. *arXiv preprint arXiv:2107.06463*.

[19] Santurkar, S., Budden, D., & Shavit, N. (2018, June). Generative compression. In 2018 Picture Coding Symposium (PCS) (pp. 258-262). IEEE.

[20] Mahalingaiah, K., Sharma, H., Kaplish, P., & Cheng, I. (2020). Semantic learning for image compression (SLIC). In Smart Multimedia: Second International Conference, ICSM 2019, San Diego, CA, USA, December 16–18, 2019, Revised Selected Papers 2 (pp. 57-66). Springer International Publishing.

[21] S. Hamis, T. Zaharia and O. Rousseau, "Image Compression at Very Low Bitrate Based on Deep Learned Super-Resolution," 2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT), Ancona, Italy, 2019, pp. 128-133, doi: 10.1109/ISCE.2019.8901038.

[22] O. H. Nagoor, J. Whittle, J. Deng, B. Mora and M. W. Jones, "Lossless Compression For Volumetric Medical Images Using Deep Neural Network With Local Sampling," 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 2020, pp. 2815-2819, doi: 10.1109/ICIP40778.2020.9191031.

[23] Weber, M., Renggli, C., Grabner, H., Zhang, C. (2021). Observer Dependent Lossy Image Compression. In: Akata, Z., Geiger, A., Sattler, T. (eds) Pattern Recognition. DAGM GCPR 2020. Lecture Notes in Computer Science(), vol 12544. Springer, Cham.