Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol.12 Issue 12, DECEMBER 2023

Image Compression and Retrieval: A Comprehensive Survey

Bhavith P B Rao
Department of Artificial Intelligence and Machine Learning
Dayananda Sagar College of Engineering
Bangalore, India

Samarth Agali Prasad
Department of Artificial Intelligence and Machine Learning
Dayananda Sagar College of Engineering
Bangalore, India

Chandrahas B
Department of Artificial Intelligence and Machine Learning
Dayananda Sagar College of Engineering
Bangalore, India

Prof. Yashaswini B M
Department of Artificial Intelligence and Machine Learning
Dayananda Sagar College of Engineering
Bangalore, India

P Vineel Akash
Department of Artificial Intelligence and Machine Learning
Dayananda Sagar College of Engineering
Bangalore, India

***Abstract*—This abstract provides an overview of image compression techniques using deep learning methods. It discusses the use of autoencoders, variational autoencoders, convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs) for image compression. The concept of vector quantization and its application in image compression is introduced. Image interpolation and the development of Super-Resolution techniques to enhance image resolution are discussed. This paper contributes to a greater understanding of these topics, providing insights into advancements and potential directions for future research in image compression, deep learning architectures, and image enhancement techniques.**

***Keywords*—**deep learning, image compression, vector quantization, encoder-decoder architecture, super resolution.

I. INTRODUCTION

Image compression and retrieval are crucial tasks in various domains, ranging from multimedia communication and storage to content-based image analysis. The rapid growth in digital imagery and the amount of media consumed by the world on a daily basis. Thus there is a need for minimizing storage requirements, reducing transmission bandwidth, and ensuring faster image rendering.

Image compression is a critical technique employed to reduce the size of image files while ensuring acceptable visual quality. Its primary objective is to facilitate efficient transmission and storage of images, as mentioned in [13]. The field of image compression encompasses two main techniques: lossless compression and lossy compression. Lossless compression methods employ algorithms that identify and exploit patterns and redundancies within the image data, resulting in a compressed file that can be fully reconstructed without any loss of information. On the other hand, lossy compression methods achieve higher compression ratios by selectively discarding less significant data, such as imperceptible details or redundant colour information. This trade-off between file size reduction and preservation of visual quality allows for effective image compression in various applications, including digital imaging, multimedia systems, and web-based image transmission.

In applications where image quality preservation is crucial, lossless compression is essential, whereas lossy compression is commonly used for web image transmission. Compressive sensing, an emerging field, emphasizes efficient signal acquisition and storage through reconstruction from limited data samples [14-16]. Within the broader context of compressive sensing, we will specifically explore vector quantization (VQ) as a subtopic in our survey paper. VQ offers a means of reducing data dimensionality by grouping similar data points, thus facilitating efficient representation and compression of images.

Deep learning has revolutionized numerous disciplines by enabling the efficient representation and generation of complex data. One such area is image processing, where a large number of Deep learning techniques have been employed to perform tasks such as image compression. Deep learning techniques have brought about a revolutionary transformation in image compression by utilizing their capacity to learn hierarchical representations directly from unprocessed image data. These models excel at capturing and utilizing the spatial correlations and intricate patterns found in images. Compressed image methods based on deep learning have exhibited exceptional performance in terms of compression effectiveness and visual fidelity.

Encoder-decoder architectures have emerged as a significant paradigm in this field, addressing the difficulties posed by sequential and variable-sized inputs. The encoder compresses

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol.12 Issue 12, DECEMBER 2023

the input into a lower-dimensional latent representation, while the decoder reconstructs the output data from the latent space. These architectures' applications consist of image processing, natural language processing, and speech recognition.As stated previously, these architectures have the potential to revolutionize a variety of disciplines.
Image interpolation is a technique used in digital image processing to resize or distort an image from one-pixel grid to another. It is required when changing the total number of pixels, whereas remapping can occur when correcting for lens distortion or rotating an image. However, modern computer graphics models lack the ability to render elements independently of the pixel grid. Thus, Super-Resolution was developed so as to improve the resolution of optical systems. Although it has been a problem for many years, finding a solution is still a challenging task. This is due to the high subjectivity of a High-Resolution (HR) image generated from a low Resolution (LR) image. The high subjectivity could be due to multiple variations in lighting, camera angles, brightness, noise, contrast and other variables.

II. IMAGE COMPRESSION USING DEEP LEARNING TECHNIQUES

Deep Leaning(DL) techniques have revolutionized image compression and reconstruction by utilizing their ability to learn complex representations directly from raw image data. DL-based compression models have capabilities in capturing intricate image features and preserving visual quality during compression.

A. Autoencoders

Autoencoders (AE) are neural networks used for unsupervised learning tasks, including image compression. By learning to reconstruct an input image from a compressed representation, AE-based compression techniques can effectively reduce image size while preserving important visual features. AE are widely used for image compressions because they allow reduction in the dimensionality of the input image. An AE consists of three parts: an input layer, a bottleneck layer where the latent space is represented, and an output layer. AE-based image compression models were developed as early as 2016[1].

B. Variational Autoencoders

Variational Autoencoders (VAE) add a probabilistic element to the compression process. This allows to generate high-quality images from compressed encodings. VAE produce better results than simple AE but also use up more computation power, for still image compression.
In [3], the authors introduced a VAE architecture using nonlinear transform and uniform quantizer for image compression. In [4], the authors introduced a VAE-based architecture meant for image compression on high-resolution images. To enhance the training process, the authors incorporated a non-local attention module (NLAM). However, this resulted in a significant increase in the overall complexity of the model.

C. Convolutional Neural Networks

Convolutional Neural Networks (CNN) have shown great potential in image compression due to its feature extraction characteristics. CNN can learn compact representations of images and reduce file size while maintaining perceptual quality.
In [5]– [7], the authors used CNNs to compress images. These compression models scored better than JPEG and JPEG-2000 in structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR).

D. Recurrent Neural Networks

Recurrent Neural Networks (RNN) have been applied to image compression by leveraging sequential modeling. An RNN-based image compression architecture combines convolutional layers for initial feature extraction, GDN and IGDN layers for feature normalization and reconstruction, RNN modules for capturing temporal dependencies, and binarized convolutional layers for efficient compression. RNN-based compression models capture temporal dependencies in images and can achieve efficient compression by encoding and reconstructing images in a sequential manner. In [8], the authors introduced an RNN-based compression architecture that utilized a stop code tolerant (SCT) approach for training the model. Their model was evaluated on the Kodak and ImageNet datasets. In another study [9], an RNN-based compression method was proposed specifically for still images. The training dataset in this study consisted of Kodak images. The performance of this model surpassed that of JPEG, JPEG-2000, and WebP. The authors in [10], illustrated the use of RNN with entropy encoding to compress images.

E. Generative Adversarial Networks

Generative Adversarial Networks (GAN) are used in image compression by incorporating a generator and a discriminator network. GAN-based compression techniques generate compressed representations that can be decoded to high-quality images while achieving high compression ratios. The authors in [11], illustrated the use of GAN for the compression and classification of semantic data. In [12], the authors demonstrated the use of unified binary GAN (BGAN+) for image compression and image retrieval. The model achieved better than JPEG and JPEG-2000. The visual quality of the reconstructed image was significantly improved compared to JPEG and JPEG-2000 at low bit-rates.

F. Performance Evaluation

Image compression plays a crucial role in reducing the size of digital images while maintaining acceptable visual quality. To evaluate the effectiveness of various compression techniques, performance metrics like Minimum bpp, PSNR, and MS-SSIM are used.
Minimum bpp (bits per pixel): Minimum bpp refers to the minimum number of bits required to represent a single pixel in a compressed image. It measures the average compression ratio achieved by a particular algorithm. Lower minimum bpp values indicate more efficient compression, as fewer bits are needed to encode each pixel.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol.12 Issue 12, DECEMBER 2023

PSNR (Peak Signal-to-Noise Ratio): PSNR is a widely used metric to evaluate the quality of compressed images. It measures the ratio between the maximum possible power of a signal (in this case, an image) and the power of the noise introduced by compression. PSNR is expressed in decibels (dB) and higher PSNR values indicate better quality.

MS-SSIM (Multi-Scale Structural Similarity Index): MS-SSIM is an advanced metric that considers both structural and perceptual similarity between the original and compressed images. Unlike PSNR, MS-SSIM takes into account human visual perception. MS-SSIM outputs a value between 0 and 1, where higher values indicate better similarity.

Table 1: Summary of performance evaluation

Research	Model	Minimum bpp	PSNR (dB)	MS-SSIM
[3]	VAE	0.15	30.76	0.955
[4]	VAE	0.2	30	0.7768
[5]	CNN	0.0726	23.93	0.8118
[6]	CNN	0.519	33.62	0.981
[7]	CNN	0.2	31	0.7878
[9]	RNN	0.5	33.59	0.9877
[11]	GAN	0.0983	28.54	0.973

III. VECTOR QUANTIZATION

Vector quantization (VQ) is an extensively employed data compression method that efficiently stores and transmits data by processing k-dimensional vectors (k-pixel blocks) rather than individual scalars [17-19]. In image compression, VQ maps input vectors to an output space, representing the original image faithfully while reducing the number of distinct vectors used for compression. VQ accomplishes significant compression ratios while maintaining acceptable image quality by effectively capturing image correlations and redundancies. Its ability to manage k-dimensional vectors and exploit image correlations has made VQ the preferred method for efficient image compression, thereby enabling applications that require optimal data storage and transmission efficiency.

A. Lossy compression schemes for image coding

Lossy compression techniques exploit the human eye's tolerance for image distortion, achieving lower bit rates by sacrificing data [17, 18]. Quantization during encoding eliminates redundant information, resulting in smaller bit sizes [17, 18]. The trade-off between image quality and bitrate is dependent on the application and extent of degradation desired [19]. Techniques such as vector quantization and hybrid coding optimize compression efficacy [20] and provide a balance between compression rates and image quality. These methods play a vital role in numerous applications that require efficient data transmission and storage.

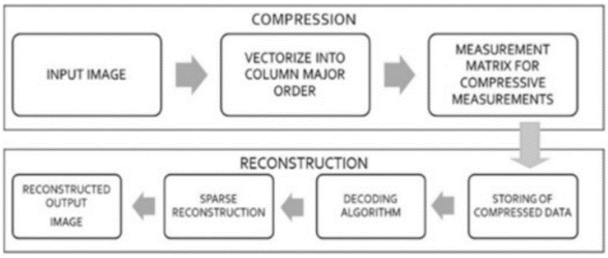


Fig.1: Flowchart for Lossy compression image coding

B. Vector quantisation in image compression

Vector quantization (VQ) is a coding method derived from Shannon's rate distortion theory, which implies that coding vectors of information yields superior performance to coding scalars [21]. Numerous applications, such as image coding, have involved extensive research. In VQ, an image is divided into tiles, and each tile is regarded as a k-dimensional vector U. The encoder selects the codeword y that minimises the distortion measure d(u, y) for each tile from the codebook. The index j corresponding to the chosen codeword is transmitted over the channel. If the channel is error-free, the decoder retrieves codeword y from the received index and outputs it as the reconstructed image tile if the channel is error-free. Mathematically, VQ encoding is a mapping from a k-dimensional vector space to a finite set of symbols, J

$$VQ: u = (u_1, u_2, \dots, u_k) \rightarrow j \quad (1)$$

where k = nm, j E J, and J has size J = Y. The rate, R, of the quantization is

$$R = \log_2 Y \quad (2)$$

where R is bits per input vector. The compression rate is R/k bits per pixel. Typically, Y is chosen to be a power of 2, so R is an integer. Consequently, VQ encoding generates codes of R bits in length with every R-bit code corresponding to some y E Y.

C. Practical limitations of basic vector quantisation

Vector quantization (VQ) for image coding is hindered by computational complexity and memory requirements, which restricts its application to small dimensions and low bitrates. Efforts to reduce complexity through techniques such as cluster merging and dimension reduction result in marginally inferior codebooks, whereas the generation of optimal codebooks requires a significant amount of computation time. In VQ-based image coding, additional research is required to strike a balance between codebook quality and computational efficacy. [22]

D. Experimental Results and Comparison

In our research, we compared VQ-based compression to conventional methods such as JPEG and PNG, concentrating on compression efficiency, image quality, bandwidth reduction, and preservation of fidelity. VQ attained high compression ratios by exploiting image correlations efficiently, particularly with small block sizes. Subjective and objective evaluations confirmed that VQ maintained acceptable image quality at lower bitrates, frequently outperforming conventional methods. Analysis of bandwidth reduction demonstrated VQ's efficiency

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol.12 Issue 12, DECEMBER 2023

in optimizing network utilization, making it advantageous for constrained networks. Moreover, VQ effectively reconstructed images with minimal distortion, preserving essential image components and visual characteristics. Our findings demonstrate that VQ is a promising alternative to conventional compression methods.

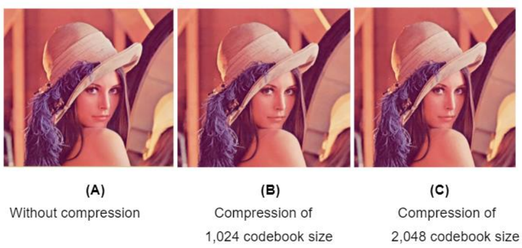
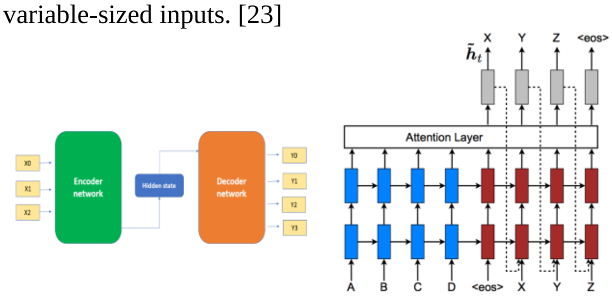


Fig.2: (A) "Lena" without compression. (B) VQ-compressed "Lena" image, codebook size of 1,024, compression time of 6 minutes, 2.6 bits/pixel. (C) VQ-compressed "Lena" image, codebook size of 2,048, compression time of 10 minutes and 2.9 bits per pixel.

From the above figures, it is evident that the quality of the images increases with codebook size. Increasing the codebook size increases, the computations per image. Hence, a flexible choice of the trade-off between compression ratio and fidelity must be made in accordance with the application requirements and resource constraints.

IV. ENCODER – DECODER ARCHITECTURE

The encoder-decoder architecture is a powerful model that has found success in various domains, including both natural language processing (NLP) and image processing. While attention-based (transformer) encoder-decoder models have become the de-facto standard in NLP, similar advancements have also been made in the field of image processing. The encoder-decoder architecture has exhibited notable progression throughout its history, accommodating the expanding complexities of deep learning objectives. During the initial phases, the framework predominantly consisted of fully connected dense layers, convolutional layers, recurrent networks, and their respective variations, as mentioned in [30]. The initial encoder-decoder models exhibited proficiency in specific tasks, however, they encountered difficulties in processing complex data structures, such as sequential or variable-sized inputs. [23]



(a) (b)

Fig.3: (a) Encoder-decoder architecture (b) Attention based encoder-decoder architecture

However, RNN-based encoder-decoder models introduced the concept of sequence-to-sequence learning, allowing the processing of sequential data, such as language translation. CNN-based encoder-decoder architectures became instrumental in image processing tasks, leveraging their ability to capture spatial features and perform tasks like image segmentation and generation. [24][25][31]

A general encoder-decoder architecture consists of two key components: an encoder network and a decoder network. The encoder network takes in input data, such as an image or a sequence of text, and processes it to generate a compressed representation called the latent space or hidden state. This latent space representation captures the essential information and patterns present in the input data in a more compact and abstract form. The latent space serves as a compressed and informative representation that retains the most important features of the input while discarding unnecessary details. The decoder network takes this latent space representation as input and aims to reconstruct the original data or generate a desired output. By utilizing the latent space representation, the decoder network can effectively reconstruct the data, preserving its essential content and characteristics.

The significance of the latent space lies in its ability to capture the underlying structure and essence of the input data, enabling efficient storage, transmission, and manipulation. It allows for various tasks such as image compression, text generation, or other generative tasks, offering a powerful framework for information representation and generation in the field of deep learning. [26]

Furthermore, the development of attention mechanisms, such as the transformer model, brought further advancements to encoder-decoder architectures. The attention mechanism allows the decoder to selectively focus on different parts of the input data during the reconstruction or generation process. By assigning varying degrees of importance to different elements of the input, such as specific regions or features in an image, the attention mechanism enables the decoder to generate more accurate and contextually relevant outputs. [27] Transformer based encoder-decoder models have become the standard in natural language processing (NLP), achieving state-of-the-art results in tasks like machine translation and text generation.

In image processing, attention based encoder-decoder models have led to improved reconstruction quality, better preservation of fine details, and the ability to handle complex visual scenes more effectively. The attention mechanism has revolutionized the encoder-decoder architecture, enabling it to capture and utilize the most relevant information from the input data, resulting in enhanced performance and more precise image processing capabilities. [32]

In recent years, encoder-decoder architectures, in combination with super-resolution models, have emerged as a powerful solution to overcome the limitations of traditional image compression and reconstruction techniques. By incorporating an encoder network to compress the image into a compact latent representation and a decoder network for reconstruction,

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol.12 Issue 12, DECEMBER 2023

encoder-decoder architectures offer the potential for improved image quality and reduced storage requirements. [28]
One notable advantage of encoder-decoder architectures is their ability to leverage the power of deep learning to capture intricate image features during compression. By learning the underlying patterns and structures of the image data, the encoder network can create a compressed representation that retains essential information while discarding redundant details. This enables more efficient storage and transmission of images without significant loss of quality. [28][29]
By leveraging encoder-decoder architectures along with super-resolution models, researchers and practitioners have been able to overcome the limitations of traditional image compression techniques. These advanced approaches provide a promising avenue for achieving high-quality image reconstructions with reduced storage requirements, offering significant advantages in applications where bandwidth optimization, storage efficiency, and preservation of image quality are crucial considerations.

V. SUPER RESOLUTION

The concept of image Super-Resolution focuses on recovering a high-resolution (HR) image from a low-resolution (LR) image input. The LR image is obtained by applying a degradation function to the HR image, including scaling factor, blur type, and noise. The SR process aims to predict the inverse of the degradation function and estimate the HR image corresponding to the input LR image.

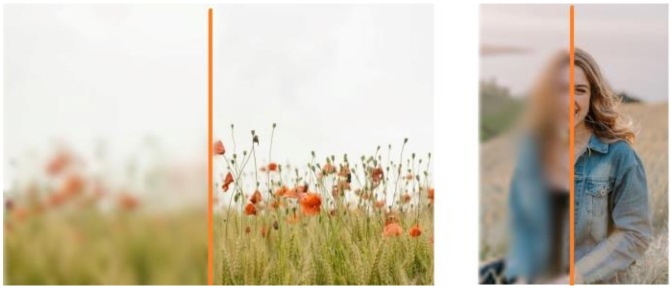


Fig.4: Examples of Super-Resolution

There are several different metrics that can be used to determine the quality of an image. One such metric is the Peak Signal to Noise Ratio (PSNR). PSNR is determined by the individual pixel intensity values of the generated HR image and the reference LR image and thus might not be able to identify when the overall image is not visually similar to the reference image. However, it is still used to compare new techniques to existing images [33-34]. To determine the structural quality of an image, the Structural Similarity Index Metric (SSIM) [35] was developed to measure the structural similarity between images by comparing the contrast and structural details within the reference image.
Apart from the above-mentioned quantitative methods, qualitative methods which take into account the subjectivity of Super-Resolution such as Opinion Scoring were also developed. Opinion Scoring asks quality texters to grade the quality of images based on certain criteria. However, due to the use of humans, this could lead to several flaws, such as inconsistencies and human errors. Due to this, perceptual quality was developed. Perceptual quality is a measure of the perceptual impact of unwanted signals and how they detract from the enjoyment or interpretation of an image or video. The most common perceptual quality metrics are based on modeling the human visual system (HVS) using aspects such as contrast and orientation sensitivity and frequency among others.
A variety of deep learning methods were developed over the years to solve the SR problem; wherein the models discussed are trained using both low and high-resolution images (LR-HR pairs). The models can be classified based on the upsampling method [36], network, learning algorithm and model frameworks.
Many models using supervised Super-Resolution are commonly used for performing Super-Resolution tasks. Models such as SRResnet and SRGAN [37] are commonly used. However, new innovations in the field of deep learning has led to the development of attention based [38], feedback based [39] and wavelet based approaches to Super-Resolution [40].
The main issue with supervised methods is that the LR images are generated using degradation methods. One approach is using weakly supervised methods where in the unpaired LR and HR images are used. Although these methods are used it does not have associations between the LR and HR images. Another method is Zero Shot Super-Resolution (ZSSR) [37], which augments a single image and uses the augmentations for training. Another method commonly used is the use of deep image priors. [41]
Some major domains in which Super-Resolution is used are face image Super-Resolution, real word image Super-Resolution, remote sensing and satellite imaging, and video Super-Resolution. Alongside the above-mentioned domains fields such as medical imaging, scientific exploration, and law enforcement stand to benefit from the field of super imaging.
Some potential complications arise from a lack of computing power due to the models' complexity, lack of data, and subjectivity of Super-Resolution as there are a large number of factors that whilst not directly contrasting in nature are difficult to incorporate in Super-Resolution models such as contrast, brightness and structural integrity of the images.

V. CONCLUSION

DL techniques have brought significant advancements to image compression and reconstruction tasks by leveraging their ability to learn intricate image representations. These techniques outperform traditional methods by providing superior compression efficiency, reduced artifacts, and improved visual quality in reconstructed images. VQ is an effective coding technique with the potential for significant bit-rate reduction and entropy enhancement. It permits fixed-length code representations and can be designed to be relatively error-tolerant. However, it necessitates cautious codebook design and can present computational difficulties and image artifacts. The encoder-decoder architecture has proven to be a powerful framework in deep learning, finding applications in diverse fields such as natural language

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol.12 Issue 12, DECEMBER 2023

processing and image processing. With its ability to compress, reconstruct and translate data, this architecture offers improved efficiency and quality. The integration of attention mechanisms and transformer models has further enhanced its capabilities. Continued research and advancements in encoder-decoder architectures promise to address complex data processing challenges and drive innovation across domains. Super-Resolution poses a unique problem by offering a broad variety of possible solutions due to its subjectivity, with a broad number of classifications of the solutions based on a variety of factors. Due to the uniqueness of the problem, it has a large variety of applications in a large number of fields. Advancements in the future of computing power, big data, and new advancements in the field of deep learning will help address the problems posed by Super-Resolution.

REFERENCES

[1] J. Balle, V. Laparra and E. P. Simoncelli, “End-to-end optimized image ́ compression,” arXiv preprint, 2016. arXiv:1611.01704.
[2] Learning-Driven Lossy Image Compression; A Comprehensive Survey arXiv:2201.09240 Sonain Jamil, Md. Jalil Piran, Muhib Ur Rahman
[3] L. Zhou, C. Cai, Y. Gao, S. Su and J. Wu, ”Variational Autoencoder for Low Bit-rate Image Compression,” in Proc. of The IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2018.
[4] T. Chen, H. Liu, Z. Ma, Q. Shen, X. Cao and Y. Wang, “End-to-End Learnt Image Compression via Non-Local Attention Optimization and Improved Context Modeling,” IEEE Transactions on Image Processing, vol. 30, pp. 3179–3191, Feb. 2021.
[5] S. K. Raman, A. Ramesh, V. Naganoor, S. Dash, G. Kumaravelu and H. Lee, “CompressNet: Generative Compression at Extremely Low Bitrates,” in Proc. of The IEEE Winter Conf. on Applications of Computer Vision, pp. 2325–2333, 2020.
[6] Z. Cheng, H. Sun, M. Takeuchi and J. Katto, “Learned image compression with discretized gaussian mixture likelihoods and attention modules,” in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, pp. 7939–7948, 2020.
[7] J. Lee, S. Cho and M. Kim, “An end-to-end joint learning scheme of image compression and quality enhancement with improved entropy minimization,” arXiv preprint arXiv:1912.12817, 2020.
[8] K. Islam, L. M. Dang, S. Lee and H. Moon, “Image Compression With Recurrent Neural Network and Generalized Divisive Normalization,” in ... 14 Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 1875-1879, 2021.
[9] G. Toderici, S. M. O’Malley, S. J. Hwang, D. Vincent, D. Minnen, S. Baluja, M. Covell and R. Sukthankar, “Variable Rate Image Compression with Recurrent Neural Networks,” CoRR, vol. abs/1511.06085, 2016
[10]G. Toderici, D. Vincent, N. Johnston, S. Jin Hwang, D. Minnen, J. Shor and M. Covell, “Full Resolution Image Compression with Recurrent Neural Networks,” in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 5306–5314, 2017
[11] R. Torfason, F. Mentzer, E. Augustsson, M. Tschannen, R. Timofte and L. V. Gool, “Towards Image Understanding from Deep Compression Without Decoding,” in Proc. of Int. Conf. on Learning Representations, 2018.
[12] J. Song, T. He, L. Gao, X. Xu, A. Hanjalic and T. H. Shen, “Unified Binary Generative Adversarial Network for Image Retrieval and Compression,” International Journal of Computer Vision, vol. 12.
[13] Image compression using vector quantization. Pamela Cosman,Karen L. Oehler,E.A. Riskin,Robert M Gray
[14] Compressed sensing. David L. Donoho
[15] Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information.Emmanuel J. Candès,Justin Romberg,Terence Tao
[16] Near-optimal signal recovery from random projections: Universal encoding strategies. Emmanuel J. Candes,Terence Tao
[17] A mathematical theory of communication. C. E. SHANNON
[18] Vector quantization and signal compression. Gersho A,Gray R.M
[19] Introduction to data compression. Sayood, K
[20] The JPEG still picture compression standard. Gregory K. Wallace
[21] Vector quantization. Robert M. Gray
[22] Image Coding Using Differential Vector Quantization. James E. Fowler, Jr.,Matthew R. Carbonara,Stanley C. Ahalt
[23] Understanding How Encoder-Decoder Architectures Attend Kyle Aitken, Vinay V Ramasesh, Yuan Cao Google Research, Niru Maheswaranathan
[24] A survey on deep learning approaches for medical images and a systematic look up into real-time object detection AmritaKaur, YadwinderSingh, NirvairNeeru, LakhwinderKaur, AshimaSingh
[25] Neural Machine Translation by jointly learning to align and translate Dzmitry Bahdanau, KyungHyun Cho, Yoshua Bengio
[26] Efficient structuring of the latent space for controllable data reconstruction and compression Elena Trunz, Michael Weinmann, Sebastian Merzbach, Reinhard Klein
[27] Attention is all you need Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin
[28] Image and Video Compression with Neural Networks: A Review Siwei Ma, Xinfeng Zhang, Chuanmin Jia, Zhenghui Zhao,Shiqi Wang, Shanshe Wang
[29] An End-to-End Deep Learning Image Compression Framework Based on Semantic Analysis Cheng Wang, Yifei Han, Weidong Wang
[30] Sequence to Sequence Learning with Neural Networks Ilya Sutskever, Oriol Vinyals, Quoc V. Le
[31] SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla
[32] Attention mechanisms in computer vision: A survey Meng-Hao Guo, Tian-Xing Xu, Jiang-Jiang Liu, Zheng-Ning Liu, Peng-Tao Jiang, Tai-Jiang Mu, Song-Hai Zhang, Ralph R. Martin, Ming-Ming Cheng, Shi-Min Hu
[33] Image quality metrics: PSNR vs. SSIM A. Horé, D. Ziou
[34] Stego image quality and the reliability of PSNR A Almohammad, G Ghinea
[35] Analyzing the role of visual structure in the recognition of natural image content with multi-scale SSIM D.M. Rouse, S.S. Hemami
[36] An investigation of Pre-Upsampling generative modelling and Generative adversarial networks in audio Super-Resolution James King, Ramon Vinas Torne, Alexander Campbell, Pietro Lio
[37] Photo-realistic single image Super-Resolution using a generative adversarial network C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi
[38] Second-order Attention Network for Single Image Super-Resolution Tao Dai, Jianrui Cai3 , Yongbing Zhang, Shu-Tao Xia, Lei Zhang
[39] Feedback Network for Image Super-Resolution Zhen Li, Jinglei Yang. Zheng Liu. Xiaomin Yang, Gwanggil Jeon, Wei Wu
[40] Wavelet-based residual attention network for image Super-Resolution Shengke Xue, Wenyuan Qiu, Fan Liu, Xinyu Jin
[41] Deep Image Prior Dmitry Ulyanov, Andrea Vedaldi, Victor Lempitsky