



---

# **Image Compression Based on Deep Learning: A Review**

**Hajar Maseeh Yasin<sup>1\*</sup> and Adnan Mohsin Abdulazeez<sup>1</sup>**

*<sup>1</sup>Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq.*

## **Authors' contributions**

*This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.*

## **Article Information**

DOI: 10.9734/AJRCOS/2021/v8i1130193

### Editor(s):

(1) Dr. Manish Mahajan, CGC College of Engineering, India.

### Reviewers:

(1) Nur UYLAŞ SATI, Muğla Sıtkı Koçman University, Turkey.

(2) Emanuele Torti, University of Pavia, Italy.

Complete Peer review History: <http://www.sdiarticle4.com/review-history/67968>

**Review Article**

**Received 20 February 2021**

**Accepted 28 April 2021**

**Published 01 May 2021**

---

## **ABSTRACT**

Image compression is an essential technology for encoding and improving various forms of images in the digital era. The inventors have extended the principle of deep learning to the different states of neural networks as one of the most exciting machine learning methods to show that it is the most versatile way to analyze, classify, and compress images. Many neural networks are required for image compressions, such as deep neural networks, artificial neural networks, recurrent neural networks, and convolution neural networks. Therefore, this review paper discussed how to apply the rule of deep learning to various neural networks to obtain better compression in the image with high accuracy and minimize loss and superior visibility of the image. Therefore, deep learning and its application to different types of images in a justified manner with distinct analysis to obtain these things need deep learning.

**Keywords:** *Image compression; artificial neural networks; deep learning; recurrent neural networks; convolutional neural networks.*

## 1. INTRODUCTION

Image compression is a technique in which an image can be compressed by various methods [1]. Lossy and lossless are two types of image compression techniques [2], where there is a potential to remove any valuable data in the lossy for the original image. However, there is no risk of missing the original image data in Lossless [3]. Early methods compress images primarily by specifically using entropy coding to remove data redundant within the image matrix, like arithmetic coding [4], Huffman coding [5], and Golomb code [6]. In the late 1960s, the transforming coding of image compression, namely Fourier transform [7] and Hadamard transform, was proposed by encoding spatial frequencies [8]. The Discrete Cosine Transform [9] was proposed in 1974 by Ahmed et al. on [10] as an Image coding that can stack image energy and make compression in the frequency domain more efficient in the low-frequency domain. The quantization and prediction techniques are also introduced for reducing data redundant through entropy encoding and transforming techniques to minimize spatial and visual redundancy in the image. BPG [11], JPEG 2000 [12], JPEG [13], WebP [14], etc, are known as file compression (format) standards. JPEG is the most widely used or accepted in several techniques for loss image compression [15].

The ANNs are not fresh in the image coding field. Between the 1980s and 1990s, neural network research on image coding was a topic of much research [16]. Thanks to the availability of large quantities of data, the innovation of efficient processing systems, and sophisticated algorithms, it is now possible to train intense models of more than 1000 layers. Deep learning for image coding is valuable and has been continuously developed since 2015 [17,18].

Deep learning is a sub-field within machine learning that depends on learning several representations corresponding to a hierarchical structure of features, factors, or concepts [19]. Deep learning emerged from the research field of artificial intelligence neural networks. Pre-fed neural networks equipped with multiple hidden layers (called DNNs) are an excellent example of deep structure models. Backpropagation (BP) became popular in the 1980s [20] and it is a well-known algorithm for learning these networks' parameters. However, backpropagation alone did not work well with learning networks with many

hidden layers, and the difficulty increased as the depth of these networks increased [21]. Deep learning approaches may be instrumental in developing unsupervised learning processes. After all, we do not know what the name of a the paper structured contains three sections thing is about occasionally. In computer vision, a recurrent neural network (RNN) focused on deep learning is likely to become a prevalent network model and achieve more significant popularity in applied research with the advancement [22]. This an enticing aspect of reliable chemical processes.

Neural networks, especially CNNs, have recently achieved considerable progress in many areas, such as image comprehension, processing, and compression. CNN typically has one or two convolution layers [23]. Some parameters will be well trained in these layers depending on large image samples classified in an end-to-end technique for particular tasks. The qualified CNN is well adapted to solve grouping, identification, and prediction tasks with highly efficient adaptability on test results. The CNN foretelling signals' efficiency has exceeded that of the rule-based dataset features [24,25]. CNN can also be understood as a features extractor for converting the image into a compression representation function space, which helps manipulate images. It was also known as a potential alternative for compressive tasks based on these stellar features of CNN [26].

The past few years have seen good experience in many subjects, especially image processing and computer vision, of deep learning technology. Deep learning in video compression remains, however, in its infant stage. This article examines the democratic work on deep learning for image and video codec that has progressed actively for a few years. Therefore, this paper explores intelligent image compression algorithms based on the deep learning used in the last four years. The paper is organized as follows. Section 2 briefly present the image compression idea with its types. Section 3 reviews the deep learning with the most popular types of networks such as convolutional neural network and recurrent neural network that used in image compression. Section 4 reviews the prior studies in image compression using deep learning. The discussion on summarizing findings from previous studies will be present in Section 5. Finally, the paper is ended with a conclusion in Section 6.

## 2. IMAGE COMPRESSION

Image compression is a method commonly used for minimizing the size of the image during image recording and processing. With growing image quality and scale, compression has become critical in everyday life. Compression plays a significant role in storing a vast number of images digitally with the increased use of cloud computing [27,28].

The compression of images is helpful because the data used to view an image decreased and reduced the space available to store the image. The quality of the images' processing will improve [29]. Different color channels define the strength of a pixel position in the image, which includes a value translated to an appearance by assigning the value in the color map to a level. Grey-scale is the most common color map with all colors set from black to white. In addition to gray images, true-color images consist of a red, green, and blue vector for each pixel location—this type of image is used as a three-dimensional image [30,31].

Three distinct forms of redundancy exist in the graphical image (unequal sensitivity of the human eye to different visual information). However, most image compression models adopt Fig. 1. In the first aspect, the original image is denoted by  $I(m,n)$  and the compressed one by  $I'(m,n)$ —the source decoder used to delete redundancies in the image content. A channel encoder is used as an overhead to tackle noise by introducing parity bits. The channel may either be a contact network or a storage device. The job of channel decoder is to revoke the operation that done by the channel encoder.

### 2.1 Lossy Compression

It is focused primarily on file size after compression, where the size of the image is significantly reduced, but the image quality will diminish relative to the original image. Lossy compression enables fidelity for a specific transmission and storage. It also decreases the number of bits required for a picture with information loss [33]. As shown in Fig. 2 (a), the lossy compression requires three stages: acquisition, quantization, and entropy coding.

### 2.2 Lossless Compression

The lossless compression does not substantially reduce the file's size, but the image quality will be maintained to the highest degree possible. This form of reduction is helpful for applications for analysis and image evaluation. This compression's fundamental purpose is to render the image with as few bits as possible and improve the transfer rate, lower storage requirements. [34]. Although loss-free compression results in a very high rate of bit reduction because images are not chosen for use, the loss-free compression is sufficient for large applications like medical, deep sensing, security, and research applications. Lossless compression refers to different formats, such as videos, pdf files, text documents, etc. [35].

A traditional diagram of lossless compression will represent, as shown in Fig. 2 (b). In the first Stage, the mathematical transform is implemented to eliminate the redundancy between pixels. In the second Stage, redundant coding is carried out by entropy coding technique.

## 3. DEEP LEARNING

Deep learning (also called deep-structured learning or hierarchical learning) has emerged since 2006 as a new field within machine learning research [37]. Research techniques for deep learning have evolved over the past years and have affected a wide range of work on processing signals and information in traditional and modern forms. This technique is within the broad areas that include the basic concepts of machine learning and artificial intelligence [38].

Deep learning is involved in machine learning theory. It has differentiated theoretical models such as deep neural or multi-layered networks of a dynamic network of simple nonlinear basic units [39]. A potential advantage of such thick networks is to process data at many levels of abstraction with varieties of representations. All the representations have been obtained by learning an efficient hierarchy of symbols from distributed and massive data [40]. Deep learning effectively removes the requirement of handcrafted representations, which can be very useful when dealing with data, such as acoustic and visual signals, while performing the same [41,42]. Fig. 3 is an example of fully deep neural networks containing more than two hidden layers.

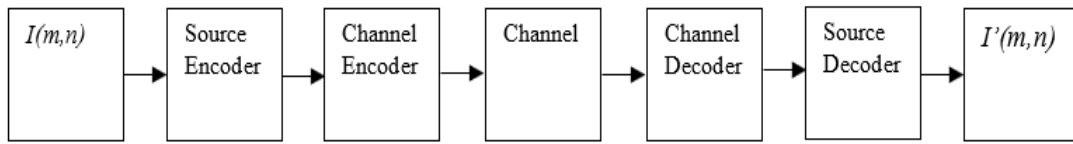


Fig. 1. Image compression model [32]

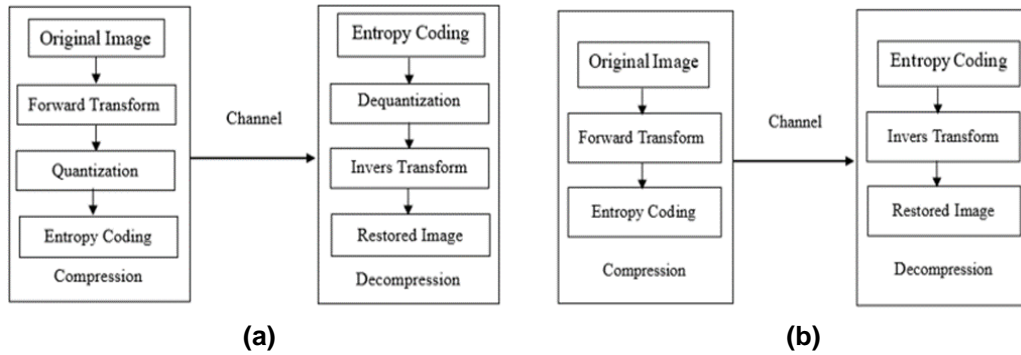


Fig. 2. (a) Lossy compression block diagram (b) Lossless block diagram [36]

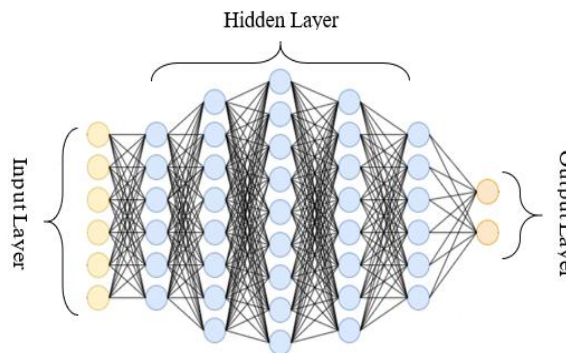


Fig. 3. Fully deep neural network architecture [43]

Deep learning has gained much interest lately, and it has been widely implemented in many machine vision problems. Deep learning approaches are now being used for low-level processing tasks [44]. Dong et al. [45] showed that an end-to-end manner could equip a convolutional neural network (CNN). They realized their system for JPEG compressive image restoration in the same year. A procedure to minimize the weights and speeds of skaters has been performed. In other words, they use pictures that have degraded differently. Kim et al. [46] suggest an intense network learn residual. Understanding the network structure based on sparse coding contributes to even more stable and effective machine learning. Driven by studies into sparse coding, Wang et al. [47] use the sparsity-based dual-domain approach to build a neural network that imitates the form of light

codes. Deep learning is an essential tool for low-level machine vision issues [25].

### 3.1 Recurrent Neural Network (RNN)

The recurrent neural network is one of the types of neural networks which is store past-related information. RNN units have multiple connections to themselves and convert knowledge out of the past. RNN uses old data from the models to shape current performance [48]. In applications such as machine translation, face recognition, and image compression, recurrent neural network architectures have achieved state-of-the-art outcomes [49].

The recurrent neural network structure ranges from completely interconnected to partly connected nets, and many layer feedforward

networks include particular input and output layers. As shown in Fig. 4 (b) [50], the completely (fully) connected networks do not have individual input layers of nodes, and each node connects to all other nodes as input. The node with itself feedback is possible [51].

As shown in Fig. 4 (a), the simple partial recurrent neural network is used to learn character strings. Since some nodes are a feedforward framework, other nodes' sequential context will be provided, and feedback from other nodes is received. Weights from the context units (C1 and C2) will be processed using backpropagation [52], like those for the input units. The context units receive time-delayed feedback from the second layer units, in the case of simple RNN. Inputs and their desired successor outputs consist of training data. The net can be trained to predict the next letter in a character string and validate a character string [53].

### 3.2 Convolutional Neural Network (CNN)

Convolution neural networks are primarily used for image compression and classification but have proven successful for a variety of tasks, such as speech recognition [54], pose estimation [55], and visual saliency detection [56]. There are numerous CNN structure variations, including fully convolution networks and deep supervised networks [57,58]. However, their underlying structure is similar, as it involves three layers:

convolution, pooling, and fully-connected layers [59].

The convolution layer learns feature representations of input images by using several convolution kernels. Here, a kernel is a matrix of weight values used for computing feature maps that provide feature information such as edges or curves. Furthermore, the pooling layer is usually placed between two convolution layers and aims to reduce feature maps' resolution. This method is practical as kernels in succeeding convolution layers can encode more features like pooling and convolution layers [60,61].

After many convolution and pooling layers, it is common to use one or more full-connected layers that link nodes in the past layer with every single node of the present layer. Lastly, as a final step, the network uses a SoftMax activation function to produce probabilities of classes from the input data [62,63]. Fig. 5 shows the architecture and the training process discussed in the above sections.

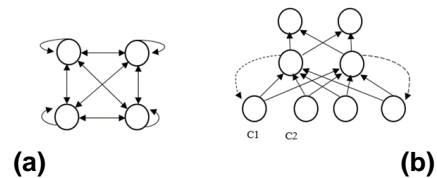


Fig. 4. (a) RNN Fully Connected (b) Simple RNN [50]

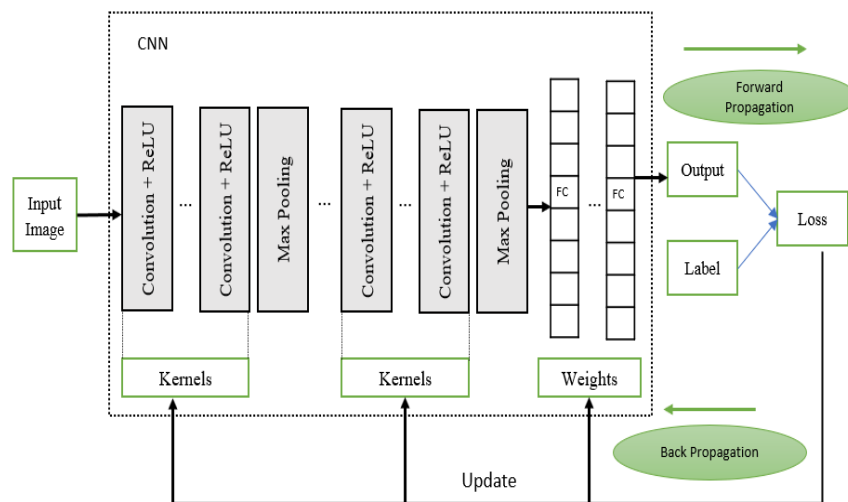


Fig. 5. CNN Architecture [64]

#### 4. LITERATURE SURVEY

The deep learning concept applies to artificial neural networks with many layers. In the last few decades, deep learning recognized as one of the most effective methods for managing massive data sets, and it has been a widespread method in the literature. The interest in more hidden layers surpassing classical algorithms has evolved recently in image encoding and compression. Therefore, in this section, we summarize previous work in using deep learning for image compression.

K. GU et al. [65]. Plane to develop an image quality index for quality assessment of blinds using deep learning. Extensive experiments are performed to measure their TID2013 system's feasibility compared to classical full-reference, state-of-the-art reduced-reference, and no-reference image consistency measurement methods.

K. Watkins et al. [66]. Introduce a neural network for gray image compression and focus on designing a fault-tolerant distribution system with stream error correction capability. Next, a deep neural network (DNN) suggests image compression using the Levenberg-Marquardt algorithm. They show experimentally that their DNN increases the restored images' consistency above that of the De DCT Zonal coding, DCT Threshold coding, Set Dividing up in Hierarchical Trees, and Gaussian Pyramid.

H. Deng et al. [67]. Use a genetic algorithm to choose the correct deep autoencoder to process images. The critical information can be extracted from the image via the optimized network. The standard image analysis shows that the proposed algorithm achieves a higher signal-to-noise ratio and retains performance even at high compression ratios.

F. Hussain et al. [68]. Explore the Deep Neural Network (DNN) ability to reveal the relationship between compression/decompression. The goal is to create a DNN capable of image compression with an algorithm that has better generalization property and requires less training time. ReLUs that more accurately map to biological neurons shortens the encoding/decoding time and improves its generalization ability. The introduction of ReLUs ensures an efficient gradient propagation, causes the proposed network to be sparse and is calculated effectively so that these networks are suited for real-time system compression.

G. Toderici et al. [69]. Proposed a general structure for variable-rate image compression and a novel structure focused on recurrent convolution and deconvolution Long Short-Term Memory networks. These structures solve the principal problems which have prevented autoencoders from handling current algorithms for image compression: Firstly, only once (not every image), regardless of the input image size and the required compression rate, must the networks be trained. Secondly, networks are increasingly sending more bits. Thirdly, the most effective image is rebuilding.

X. Sun et al. [70]. Proposed a joint deep-network-based image restoration algorithm to create a restoration system for multiple degradations. There are two phases of the convolutional neural network used in their proposal. Firstly, with two cascaded neural networks, a de-blocking subnet is created. A 20-layer profound network with skipping connections performs super-resolution—a novel deep network created by cascading these two phases.

F. Jiang et al. [71]. A new compression method is proposed based on a deep neural network. Two neural networks are cleanly assembled in a high-quality image compression end-to-end encoding environment at low bit rates. The first convolutionary neural networks, known as the Compact Convolutional Neural Network (ComCNN), learn from an image that preserves the structural data and is coded with an image codec (e.g., JPEG, JPEG2000, or BPG). The so-called Convolutionary Reconstruction Neural Network (RecCNN) is the second CNN to replicate the high-quality decoded image at the encoding point. They create a single end-to-end learning algorithm that simultaneously lets ComCNN and RecCNN communicate efficiently with two CNNs and allows the decoded image to be reconstituted exactly with RecCNN. Such a design also makes the theoretical compression framework compliant with existing image coding standards.

W. Shi et al. [72]. Use Deep learning to solve the two most critical issues of compressed sensing (CS), i.e., creating a sampling operator and developing a fast nonlinear reconstruction algorithm. Develop a deep network composed of three sub-networks: compressed sampling, initial reconstruction, and deep reconstruction, which has a high relationship with the conventional block compressed sensing smoothly predicted Landweber algorithm. The sampling operator

automatically learns to prevent intricate artificial designs. In the sense of sampling measurements, the reconstruction subnetwork will efficiently recover the original image.

G. Toderici et al. [73]. Introduce several neural networks for lossy compression methods. During implementation, each of these neural network architectures can provide varying compression rates but need to be trained once. Each architecture has a recurrent neural network, decoder, encoder, multinomial discretized, and entropy coding. The model is fundamentally new, which comes from the combination of Gated Recurrent Unit and Residual Net. They study the "One Shot" versus "Additive" reconstruction approach and launch a new scaled-additive architecture.

Z. Ching et al. [74]. A loss image compression scheme has been developed which uses convolution autoencoder (CAE) functions to achieve high-quality loss coding during training. Initially, they designed a novel active CAE structure to replace the conventional transformations and then trained it using a rate-distortion loss function. Authors use principal components analysis (PCA) to flip the maps produced by CAE, then qualify quantization and entropy coder to make codes.

J. Zeng et al. [75]. Proposed a general hybrid JPEG-phase research deep-learning system that incorporates field expertise under the rich stage analysis models. Their system proposed occupies two significant phases. The first phase is handmade, aligned with the convolution phase and the quantitative measurements and truncation phase of the wealthy models. The second phase is a deep neural network with many deep subnetworks, in which the model parameters are learned during preparation.

D. Minnen et al. [76]. Introduce an algorithm mixing deep neural networks with a tiled network adaptation to the quality-sensitive bit rate. They illustrate the importance of the prediction of the spatial context.

F. Mentzer et al. [77]. Reflect on the latter problem and introduces a new strategy to navigate an image compression auto-encoder rate-distortion trade-off. The key idea is to model the entropy of latent representation directly using a framework: a 3D-CNN that learns a conditional probability model of the autoencoder's latent distribution. During the training, the autoencoder

uses the context model to test its entropy and the context model is simultaneously modified to learn how the latent representation relies on the symbols.

E. Peixoto et al. [78]. Use a Convolutional Neural Networks to create a new prediction model in an image. They suggest an intra-prediction solution multi-mode using two CNN-based prediction modes and all intra-mode modes previously used in the highly efficient video encoding standard. They also consider any allocation strategy that only improves bitstream if there is a substantially reduced reconstruction error.

P. Akyazi et al. [79]. Introduced two end-to-end structures for image compression dependent on convolutional neural network (CNN). Two dimensions of wavelet division as a pre-processing step before training. It features extraction for the compression of wavelet coefficients that will use in the proposed networks. Training is complete, and multiple models that operate at various rates are created using a regularization in the loss function.

S. Li et al. [80]. Propose a two-stage sub-band coding system for coefficients besides analysis by Filter banks based on convolutional networks using the SPIHT-like algorithm and then primitive adaptive arithmetic coding (AAC). The SPIHT-like algorithm stretches the spatial orientation tree to manipulate inter-sub band reliance on sub-bands and directions of various sizes. For knowledge-theoretical calculations, reciprocal information is calculated to formulate these dependencies. Different primitives have been designed to encode the bitstream created by adapting its multiple lists and passes. Neural networks boost AAC's likelihood estimates where nonlinear estimation will be based on scales, paths, positions, and coefficient importance contexts.

B. Zheng et al. [81]. A pixel location map and quantification tables were proposed as inputs to an indirectly dual-domain convolution network (IDCN). The dual-domain correction unit was presented as a fundamental feature in the formulation of the IDCN. The implied dual-domain translation enables the IDCN to interact with color images in the discrete cosine transform (DCT) domain priors. A versatile IDCN version has also been built to deal with a variety of compression qualities.

P. Guo et al. [82]. Reports end-to-end image compression systems based on convolutions of neural networks (CNNs) for the retina optical coherence tomography (OCT) images with an image compression ratio of up to 80. Our compression scheme comprises three components: pre-processing of data, CNN compression, and CNN reconstruction. The pre-processing module was designed to reduce OCT noise and to break the area of interest. The proposed system has been trained and tested with pathological knowledge on ophthalmic OCT images. When the compression ratio hit 40 percent, reconstructed images could still be more than 99% identical in terms of MS-SSIM.

M. Li et al. [83]. Presents context-based convolutional networks (CCNs) to be precise and efficient. In particular, a 3D zigzag scan and 3D code division technique is used to describe the coding contexts needed for the parallel entropy decoding, all of which are used to place the binary masks on invariant translation CCN convolution filters. CCNs have also shown their ability for entropy modeling in lossless and loss picture compression. The first is to use a CCN to calculate the Bernoulli distribution of each code directly for entropy measurement. A discrete mix of Gaussian distributions with approximate parameters of three CCNs is the distribution of each code group. The CCNs entropy model and the analysis/synthesis transformations are designed for the efficiency of rate distortion.

T. Hoang et al. [84]. Propose a new layered image compression architecture with a matched semantical segmentation encoder-decoder. A semantic segmentation network is applied in both the encoder and decoder to the up-sampled image. However, the semantic zed section extracted from the image taken up is not exactly from the original image. A neural convolution network with a particular structure is applied to its original distribution on the nonlinear map of the segment removed to achieve this quality difference.

H. Liu et al. [85]. Propose a deep learning-based Picture Wise Just Noticeable Difference model for image coding. First, subedit the task of predicting positive decision outcomes Picture Wise Just Noticeable Difference as a classification problem and propose a framework that can fix the problem using just one binary

classifier. Second, create a deep learning-based classifier model that can estimate if a combination of colors is perceptually lossy or lossless. Thus, they suggested a shifting window-based search technique to forecast Picture Wise.

A. Krishnaraj et al. [86]. Developed and implemented a deep learning-based and discrete wavelet transform (DWT) model for image compression on the Internet of Underwater Things (IoUT). Convolutional Neural Network (CNN) is used at both encoding and decoding to incorporate high-quality images. They suggest that the machine learning supervised CNN is superior to current approaches such as super-resolution deep neural networks, JPEG, and JPEG2000 in compression efficiency and reconstructed image quality.

## 5. DISCUSSION

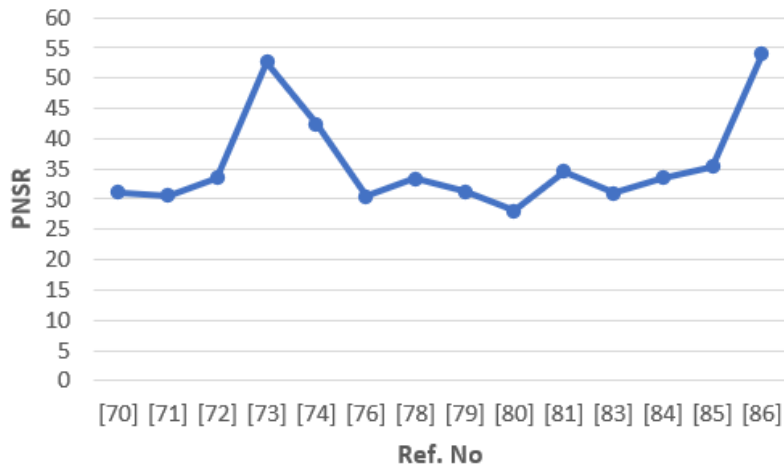
Besides the literature survey in the previous section, we conclude that one of the benefits of deep learning in image compression is folded over the last traditional methods. The excellent adaptability of neural networks' content is to form the model based on signal processing because the network elements are obtaining based on a lot of valuable data. In the latest coding standards, the models handcrafted based on the image is a prior science. In neural network models, the larger receptive field is used widely, which uses adjacent information and improves image compression efficiency by utilizing patterns from a distance. However, traditional encoders only use adjacent patterns and are hard to use in large samples. Finally, both texture and feature can be well represented using neural networks, improving the image compression process combined for human vision and computer vision analysis. While the current coding standards only seek to perform high pressure towards the task of human vision. We intend to research further the deep learning image compression in the representation and distribution of images with better quality and lower bit rates and the memory and computational efficiency in the accurate picture. Due to this issue, the deployment of deep learning has become extremely difficult. When using more complex neural networks, the performance of the networks is usually improved.



**Table 1. Summarization of literature survey**

Ref.	year	dataset	Methodology	PSNR	SSIM	result
[86]	2020	UWSN	Hybrid 20 CNN layer and Discrete Wavelet Transform (DWT)	53.961	-	79.7038% space saving
[85]	2020	MCL-JCI	Combine the CNN and JPEG coder for training and predicting Stage	35.41	0.82	get 92% for lossy/lossless features accuracy, 0.79 dB absolute prediction error of the model.
[84]	2020	ADE20K, Kodak	Hybrid a semantic segmentation network for an up-sampled image in both encoder and decoder. And the convolution neural networks to solve the semantic segment extractor.	33.57	0.977	get 35.31% BD-rate reduction over the HEVC-based (BPG) codec, 5% bitrate, and 24% encoding time-saving.
[83]	2020	Kodak and Tecnica	implemented Deep neural networks and entropy encoding.	31.01	0.978	The technique achieves a compression ratio equivalent to the new technique and is much faster.
[82]	2020	Kodak	Train the CNNs with an adversarial objective, objective function, patch discriminator, and MS-SSIM penalty.	-	0.985	gain 80% as compression ratio
[81]	2020	DIV2K, LIVE1	Hybrid deep neural networks with Discrete Cosine Transform	34.51	0.9220	the technique is much better than state-of-the-art techniques, and it has a promising application in particular.
[80]	2019	ImageNet	Hybrid SPIHT-like algorithm, arithmetic coding with deep neural networks	28.01	-	Archive 16.24%-bit rate as coding performance.
[79]	2019	CLIC2019	Hybrid the Haar wavelet technique with deep neural networks.	31.25	0.983	the methods outperform JPEG compression, reduce blurring artifacts and blocking, and save various details in the images' low bitrates.
[78]	2018	Kodak, CLIC	Use deep learning in multi-mode Intra prediction, with a block-based choice of the best mode.	33.4	0.92	reduce 28% bitrate compared to the baseline codec.
[77]	2018	ImageNet, Kodak, Urban100	combine the entropy coding with a deep neural network for training the auto-encoder.	-	0.982	the method outperforms the BPG, JPEG, and JPEG2000. Furthermore, it reduces the rate by 10 % for the context model.
[76]	2018	Kodak	Hybrid the deep neural networks with quality-sensitive bit rate adaptation.	30.418	-	the bit rate for both quantitative and subjective image quality will increase
[75]	2018	ImageNet	Hybrid the deep-learning model with threshold quantizers handcrafted	-	-	accuracy of this framework is 65.4% in the quantization and truncation phase

<b>Ref.</b>	<b>year</b>	<b>dataset</b>	<b>Methodology</b>	<b>PSNR</b>	<b>SSIM</b>	<b>result</b>
[74]	2018	Kodak	Hybrid PCA with convolution autoencoder	42.45	0.98	compared to JPEG2000, the method decrement 13.7% BD-rate
[73]	2017	32x32 Benchmark, Kodak	Hybrid Gated recurrent units with residual neural network	52.61	1.7998	Improve 4.3%–8.8% for AUC based on the perceptual metric.
[72]	2017	BSDS500, Set5, Set14	hybrid deep convolutional neural network with Landweber algorithm.	33.55	0.9081	Achieves significant performance improvements over several current state-of-the-art methods and runs in real-time.
[71]	2017	grayscale image	Combine JPEG, JPEG 2000 with compact convolutional neural network and reconstruction convolutional neural network	30.59	0.8895	saving a 5.22%-bit rate compare with JPEG 2000
[70]	2017	Set5, Set14, BSD100	deep-network structure with 28 CNN layers merged with cascaded neural network for de-blocking subnet	31.05	-	The method outperforms other methods separately in terms of subjective quality.
[69]	2016	32x32 Benchmark	Hybrid LSTM models with Recurrent Neural Network encoder and decoder	-	0.87	reduced storage size by 10% or more, and better quality than JPEG, JPEG 2000



**Fig. 6. The PSNR of the compression image gain by the previous research**

This paper focuses on these papers that review applied deep learning on image compression to get best compressed with high quality and save more storage space. Table 1 shows the dataset, methodology, Peak signal-to-noise ratio (PSNR) value, structural index similarity (SSIM) value, and the result used by previous work (Precisely in the last four years). According to this table, we find that [53] is getting the best PSNR (53.961) and save up to 79% for storage space comparing with other studies, which is used the DWT with 20 convolution neural network layers. With the hybrid of traditional compression techniques with deep learning (CNN or RNN), we get good compression compared with image coding JPEG, JPEG 2000, PNG that researchers have proven. Fig. 6 shows the tendency result of the PSNR for deep learning approaches image compression by previous studies.

## 6. CONCLUSION

Currently, Deep Learning (DL) is a widespread research path. Using the CNN or Convolution Layer (CL), the pooling layer and other superficial structures let the network structure learn and extract the related features and use them. This functionality provides many conveniences for many experiments, removing the need for a highly complex modeling method. Deep learning is also used in image classification and object recognition, image segmentation, image compression, etc., and has been fantastic findings and development. In this review, the deep learning-based image compression methods have been examined and compared, particularly for new studies of deep-learning-based image compression techniques.

## COMPETING INTERESTS

Authors have declared that no competing interests exist.

## REFERENCES

1. Brifcani MA, Al-Bamerny JN. Image compression analysis using multistage vector quantization based on discrete wavelet transform," Proc. 2010 Int Conf Methods Model Comput Sci. ICM2CS. 2010;46–53.
2. Venugopal D, Mohan S, Raja S. An efficient block based lossless compression of medical images. Optik (Stuttg). 2016;127(2):754–758,.
3. Zebari DQ, Zeebaree AM, Abdulazeez H Haron, Hamed HNA. Improved threshold based and trainable fully automated segmentation for breast cancer boundary and pectoral muscle in mammogram images," IEEE Access, 2020;8:1–20. DOI: 10.1109/ACCESS.2020.3036072.
4. Han S, Mao H, Dally WJ. Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding," 4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc. 2016;1–14.
5. Ding JJ, Chen HH, Wei WY. Adaptive golomb code for joint geometrically distributed data and its application in image coding," IEEE Trans. Circuits Syst. Video Technol. 2013;23(4):661–670.
6. Poolakkachalil TK, Chandran S, Vijayalakshmi K. Analysis of application of arithmetic coding on dct and dct-dwt hybrid

- transforms of images for compression," 2017 Int. Conf. Networks Adv. Comput. Technol. Net ACT. 2017;288–293.
7. Andrews, Pratt W. Fourier transform coding of images. Proc Hawaii Int Conf Syst Sci. 1968;677–679,
  8. Pratt, Kane J. Hadamard transform image coding. Proc. IEEE. 1969;57:1.
  9. Abdulazeez M, Zeebaree DQ, Abdulqader DM. Wavelet applications in medical images: A Review," Transform. 2020;21: 17625–17276.
  10. Ahmed N, Natarajan T, Rao K K. Discrete cosine transform," IEEE Trans. Comput. 1974;100(1):90–93.
  11. Albalawi U, Mohanty SP, Kougianos E. A Hardware Architecture for Better Portable Graphics (BPG) Compression Encoder," Proc. - 2015 IEEE Int. Symp. Nanoelectron. Inf Syst iNIS. 2015;291–296.
  12. Li M, Zuo W, Gu S, Zhao D, Zhang D. Learning convolutional networks for content-weighted image compression," Proc. IEEE Comput. Soc Conf Comput Vis Pattern Recognit. 2018;3214–3223,.
  13. Zaineldin H, Elhosseini MA, Ali HA. Image compression algorithms in wireless multimedia sensor networks: A survey. Ain Shams Eng Journal. 2015;6(2):481–490.
  14. Saeed, J., & Abdulazeez, A. M. (2021). Facial Beauty Prediction and Analysis Based on Deep Convolutional Neural Network: A Review. *Journal of Soft Computing and Data Mining*, 2(1), 1-12..
  15. IEEE MR, Ahmed S, Alone. Image Compression using neural network. Int J Innov Sci Mod Eng. 2014;2(5):24–28.
  16. Dharwal R. Applications of Artificial Neural Networks: A review. Indian J Sci Technol. 2016;9(1):1–8,.
  17. Bargarai FAM, Abdulazeez AM, Tiryaki VM, Zeebaree DQ. Management of wireless communication systems using artificial intelligence-based software defined radio. Int J Interact Mob Technol. 2020;14(13):107–133. DOI: 10.3991/ijim.v14i13.14211.
  18. Misra J, Saha I. Artificial neural networks in hardware: A survey of two decades of progress. Neurocomputing. 2010;74(1–3):239–255.
  19. Kaşkavalci HC, Gören S. A deep learning based distributed smart surveillance architecture using edge and cloud computing," Proc. - 2019 Int. Conf. Deep Learn. Mach. Learn. Emerg. Appl. Deep. 2019;1–6.
  20. Abdullah, S. M. S., & Abdulazeez, A. M. (2021). Facial Expression Recognition Based on Deep Learning Convolution Neural Network: A Review. *Journal of Soft Computing and Data Mining*, 2(1), 53-65.
  21. Hussein P, Kandel CW, Bolan MB Wallace, Bagci U. Lung and pancreatic tumor characterization in the deep learning era: Novel supervised and unsupervised learning approaches. IEEE Trans Med Imaging. 2019;38(8):1777–1787,.
  22. Govinda E, Indira Dutt VBSS. Artificial neural networks in UWB image processing for early detection of breast cancer. Int J Adv Sci Technol. 2020;29(5):2717–2730.
  23. Schiopu Y Liu, Munteanu A. CNN-based prediction for lossless coding of photographic images," 2018 Pict. Coding Symp. PCS 2018 - Proc. 2018;3:16–20.
  24. Luo YW, Li Y, Han FJ, Huang SB. Grading image retrieval based on CNN deep features. Int. Conf. Adv. Commun. Technol. ICACT, vol. 2018-Febru, 2018; 148–152.
  25. Lipmann RP. Introduction to computing with Neural Nets," IEEE Assp Mag. 1987; 4(2):4–22,.
  26. Li Gu S, Van Gool L, Timofte R. Learning filter basis for convolutional neural network compression," Proc. IEEE Int. Conf. Comput. Vis. 2019;5622–5631.
  27. H Dathar, Abdulazeez AM. A Modified Convolutional Neural Networks Model for Medical Image Segmentation, learning. 2020;20:20.
  28. W Li, Pan Q, Liang S, Jiao JY. Research on fractal image compression hybrid algorithm based on convolutional neural network and gene expression programming. J Algorithms Comput. Technol. 2019;13.
  29. Abdulazeez M, Zeebaree DQ, Asaad D, Zebari GM, Mohammed I, Adeen N. The applications of discrete wavelet transform in image processing: A review. J Soft Comput. Data Min. 2020;2(1):31–43.
  30. Fu X, Zha ZJ, Wu F, Ding X, Paisley J. JPEG artifacts reduction via deep convolutional sparse coding. Proc. IEEE Int. Conf. Comput. Vis. 2019;2501–2510.
  31. Theis L, Shi, W. Lossy I. mage compression with compressive autoencoders. ICLR. 2017;1–19.

32. Tiwari G. A Comparative study on Image and Video Compression Techniques, IOSR J. VLSI Signal Process., 2013;3(3):69–73,.
33. Hu Y, Member S, Yang W, Ma Z, Member S. Learning End-to-End Lossy Image Compression: A Benchmark. *Electr. Eng. Syst. Sci.* 2020;2:1–15,.
34. Rusyn O, Lutsyk Y, Lysak A. Lukenyuk L. Pohreliuk. Lossless image compression in the remote sensing applications. *IEEE First Int. Conf. Data Stream Min. Process.* 2016;195–198,.
35. Kurihara S. Imaizumi, Shiota S. An encryption-then-compression system for lossless image. *IEICE Trans Inf Syst.* 2017;100(1):52–56.
36. Fred L, Kumar SN, Ajay Kumar H, Abisha W, Bat optimization-based vector quantization algorithm for medical image compression. Springer International Publishing. 2019;150.
37. Abdulqader M, Abdulazeez AM, Zeebaree DQ. Machine learning supervised algorithms of gene selection: A review,” *Technol. Reports Kansai Univ.* 2020;62(3): 233–244.
38. Lecun Y, Bengio Y, Hinton G. Deep learning,” *Nature.* 2015;521:7553:436–444,.
39. Zeebaree Q, Haron H, Abdulazeez AM, Zebari DA. Machine learning and Region Growing for Breast Cancer Segmentation,” *2019 Int. Conf. Adv. Sci. Eng. ICOASE.* 2019;88–93.
40. Bodyanskiy YP, Grimm S. Mashtalir, Vinarski V. Fast training of neural networks 2 architecture of two layers ann for finding principal,” *ICDM.* 2010;165–173.
41. Deng. A tutorial survey of architectures, algorithms, and applications for deep learning,” *APSIPA Trans. Signal Inf. Process.* 2014;3.
42. Abdulazeez M, Hajy DM, Zeebaree DQ, Zebari DA. Robust watermarking scheme based LWT and SVD using artificial bee colony optimization,” *Indones. J Electr Eng Comput Sci.* 2020;21(2):1218–1229. DOI: 10.11591/ijeecs.v21.i2.pp1218-1229.
43. Bahi, Batouche M. Deep learning for ligand-based virtual screening in Drug Discovery,” *Proc. - PAIS 2018 Int. Conf. Pattern Anal. Intell. Syst;* 2018.
44. Mahmood R, Abdulazeez AM, Orman Z. A new hand gesture recognition system using artificial neural network,” *2nd Int. Sci. Conf. ZAKHO.* 2017;1–14.
45. Dong C, Loy C, He K, Tang X. Image super-resolution using deep convolutional networks,” *IEEE Trans. Pattern Anal. Mach. Intell.* 2016;38(2):295–307.
46. Kim J, Jung Kwon L, Kyoung Mu L. Accurate image super-resolution using very deep convolutional networks. *Proc. IEEE Conf. Comput. Vis. pattern Recognit.* 2016;1646–1654. DOI: 10.11236/jph.60.8\_444.
47. Wang Z, Liu D, Chang S, Ling Q, Yang Y, Huang TS. D3: Deep dual-domain based fast restoration of jpeg-compressed images. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2016;2764–2772. DOI: 10.1109/CVPR.2016.302.
48. Mohammed, S. B., & Abdulazeez, A. M. (2021). Deep Convolution Neural Network for Facial Expression Recognition. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 18(4), 3578-3586.
49. Blanco M Delgado, Pegalajar MC. A real-coded genetic algorithm for training recurrent neural networks. *Neural Networks.* 2001;14(1):93–105. DOI: 10.1016/S0893-6080(00)00081-2.
50. Medsker L, Jain LC, Eds., *Recurrent Neural Network: Design and application.* London: CRC Press; 2001.
51. Doroshenko VA. Primenenie analogovykh vychislitel'nykh mashin dlia analiza élektroéntsefalogrammy. *Nerv Sist.* 1969;10:192–196.
52. Zeebaree Q, Abdulazeez AM, Mohammed O, Hassan S. Hiding Image by Using Contourlet Transform; 2020.
53. Jorgensen TD, Haynes BP, Norlund CCF. Pruning artificial neural networks using neural complexity measures. *Int. J. Neural Syst.* 2008;18(5):389–403. DOI: 10.1142/S012906570800166X.
54. Yu L, Zhang Y, Han H, Zhang L, Wu F. Robust median filtering forensics by CNN-based multiple residuals learning,” *IEEE Access.* 2019;7:120594–120602.
55. Jahwar, Mohsin Abdulazeez A. Meta-heuristic algorithms for k-means clustering: A review. *PalArch's J Archaeol. Egypt/Egyptology.* 2020;17(7):12002–12020,
56. Ji S, Xu W, Yang M, Yu K. 3D Convolutional neural networks for human

- action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* 2013;35(1):221–231,.
57. Zeebaree D, Abdulazeez AM, "Evaluating Data Mining Classification Methods Performance in Internet of Things Applications. *J Soft Comput. Data Min.* 2020;1(2):10–25.
  58. Zeebaree DQ, Abdulazeez AM, Zebari DA, Haron H, Hamed HNA. Multi-level fusion in ultrasound for cancer detection based on uniform LBP features. *Comput Mater Contin.* 2021;66(3):3363–3382.
  59. S. I. Saleem, A. M. Abdulazeez and Z. Orman, "A new segmentation framework for arabic handwritten text using machine learning techniques," *Computers, Materials & Continua*, 2021;68(2):2727–2754,.
  60. Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B. Recent advances in convolutional neural networks. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 2019;10629–10638,
  61. Qader Zeebaree D, Habibollah H, Mohsin Abdulazeez A. Gene selection and classification of microarray data using convolutional neural network. *Int. Conf. Adv. Sci. Eng. (ICOASE), IEEE.* 2018;145–150.
  62. Chen Y, Peng F, Kang X, Jane Wang Z. Depthwise separable convolutional neural network for image forensics," 2019 *IEEE Int. Conf. Vis. Commun. Image Process. VCIP.* 2019;1–4.
  63. Omar AM, Abdulazeez Sengur A, Al-Ali GS. Fused faster RCNNs for efficient detection of the license plates. *Indones. J. Electr. Eng. Comput. Sci.* 2020;19(2):974–982.
  64. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: An overview and application in radiology," *Insights Imaging.* 2018;9(4):611–629.
  65. Gu K, Zhai G, Yang X, Zhang W. Deep learning network for blind image quality assessment. *Int Conf Image Process.* 2014;511–515,.
  66. YZ. Watkins and M. R. Sayeh, "Image Data Compression and Noisy Channel Error Correction Using Deep Neural Network. *Procedia Comput. Sci.* 2016;95: 145–152,.
  67. Deng H, Liu H, Wang F, Wang Z, Wang Y. Image compression based on genetic algorithm and deep neural network," *Commun. Comput Inf Sci.* 2016;681:417–424.
  68. Hussain, Jeong J. Efficient deep neural network for digital image compression employing rectified linear neurons. *J. Sensors.* 2016;1–7,.
  69. Toderici et al. Variable rate image compression with recurrent neural networks," 4th *Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc.* 2016;1–12,
  70. Sun X, Li X, Zhuo L, Lam KM, Li J. A joint deep-network-based image restoration algorithm for multi- degradations," *IEEE Int. Conf. Multimed. Exp.* 2017;1–6.
  71. Jiang F, Tao W, Liu S, Ren J, Guo X, Zhao D. An end-to-end compression framework based on convolutional neural networks," *IEEE Trans. Circuits Syst Video Technol.* 2018;28(10):3007–3018.
  72. Shi W, Jiang F, Zhang S, Zhao D. Deep networks for compressed image sensing," *arXiv.* 2017;92:877–882.
  73. Toderici et al. Full resolution image compression with recurrent neural networks," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR.* 2017;5435–5443.
  74. Cheng Z, Sun H, Takeuchi M, Katto J. Deep convolutional auto encoder-based lossy image compression," 2018 *Pict. Coding Symp. PCS 2018;*253–257.
  75. Zeng S, Tan B Li, Huang J. Large-scale jpeg image steganalysis using hybrid deep-learning framework. *IEEE Trans. Inf. Forensics Secur.* 2018;13(5):1200–1214.
  76. Minnen D et al. Spatially adaptive image compression using a tiled deep network," *Proc. - Int. Conf. Image Process. ICIP, vol. 2017-Septe,* 2018;2796–2800.
  77. Mentzer F, Agustsson E, Tschannen M, Timofte R, Van Gool L. Conditional Probability Models for Deep Image Compression," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2018;4394–4402.
  78. Peixoto E, Hung EM, De Campos T. Multi-Mode Intra Prediction for Learning-Based Image Compression," *IEEE Xplore.* 2018;1296–1300.
  79. Akyazi, Ebrahimi T. A new end-to-end image compression system based on convolutional neural networks. 2019; 22.
  80. Li S, Zheng Z, Dai W, Xiong H. Lossy image compression with filter bank based convolutional networks. *Data Compression Conf Proc.* 2019;23–32.

81. Zheng B, Chen Y, Tian X, Zhou F, Liu X. Implicit dual-domain convolutional network for robust color image compression artifact reduction. *IEEE Trans. Circuits Syst. Video Technol.* 2020;30(11):3982–3994.
82. Guo Li D, Li X. Deep OCT image compression with convolutional neural networks. *Biomed Opt Express.* 2020;11(7):3543,.
83. M Li, K Ma J, You D, Zuo W. Efficient and Effective Context-Based Convolutional Entropy Modeling for Image Compression,” *IEEE Trans. Image Process.* 2020;29(1): 5900–5911.
84. Hoang TM, Zhou J, Fan Y. Image compression with encoder-decoder matched semantic segmentation. *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.* 2020;619–623.
85. H Liu et al. Deep Learning-Based Picture-Wise Just Noticeable Distortion Prediction Model for Image Compression,” *IEEE Trans. Image Process.* 2020;29:641–656.
86. Krishnaraj S, Mohamed Elhoseny N, Thenmozhi M, Mahmoud Selim M. Deep learning model for real-time image compression in Internet of Underwater Things (IoUT),” *J. Real-Time Image Process.* 2020;6(17):2097–2111.

---

© 2021 Yasin and Abdulazeez; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Peer-review history:*  
*The peer review history for this paper can be accessed here:*  
<http://www.sdiarticle4.com/review-history/67968>