American Academic Scientific Research Journal for Engineering, Technology, and Sciences

ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

http://asrjetsjournal.org/

A Deep Wavelet AutoEncoder Scheme for Image Compression

Houda Chakib[a*], Najlae Idrissi[b]

[a,b]Data4Earth Laboratory, Faculty of Sciences and Technics, Sultan Moulay Slimane, B.P 523, Beni Mellal 23000, Morroco

[a]Email: houda.chakib@yahoo.fr, [b]Email: lidrissi.najlae@gmail.com

Abstract

For many years and since its appearance, Digital Wavelet Transform DWT has been used with great success in a wide range of applications especially in image compression and signal de-noising. Combined with several and various approaches, this powerful mathematical tool has shown its strength to compress images with high compression ratio and good visual quality. This paper attempts to demonstrate that it is needless to follow the classical three stages process of compression: pixels transformation, quantization and binary coding when compressing images using the baseline method. Indeed, in this work, we propose a new scheme of image compression system based on an unsupervised convolutional neural network AutoEncoder (CAE) that will reconstruct the approximate sub-band issue from image decomposition by the wavelet transform DWT. In order To evaluate the model's performance we use Kodak dataset containing a set of 24 images never compressed with a lossy algorithm technique and applied the approach on every one of them. We compared our achieved results with those obtained using standard compression method. We draw this comparison in terms of four performance parameters: Structural Similarity Index Metrix SSIM, Peak Signal to Noise Ratio PSNR, Mean Square Error MSE and Compression Ratio CR. The proposed scheme offers significate improvement in distortion metrics over the traditional image compression method when evaluated for perceptual quality moreover it produces better visual quality images with clearer details and textures which demonstrates its effectiveness and its robustness.

Keywords: Wavelet Transform DWT; Unsupervised Neural Network; AutoEncoder; Approximate Image; RGB Image; Image Compression.

----

* Corresponding author.

1. Introduction

Day after day the number of data created and transmitted keep growing due, in the first hand to the huge increase of means of communications and computer applications like Internet of Things IoT, and in the second hand to the digitalization of every sector in human's life such medicine, multimedia domains where it is an evident need to reduce the size as data at hand can be extremely large. Consequently digital image compression has received significant attention from researchers in order to create effective techniques for compression algorithms that led to achieve a simplified model using less memory than the original one. Moreover, this last years over than 70% of internet traffic is streaming of digital media [1], it's why it has been challenging for classic compression algorithms to adapt to the rising demand and develop complete model compression pipeline that combines many approaches in the purpose to take advantages and benefits from every one of them. Among these approaches, we have Machine Learning algorithms ML especially Deep Neuronal Networks DNN.

In recent years, DNN have recently received lots of attention, been applied to different applications and achieved great accuracy improvements in many tasks mainly in image processing.

They are part of ML that works by using deeper convolutional layers. They are especially well suited to identification applications such as face recognition [2], text translation [3], voice recognition [4, 5] and advanced driver assistance systems including lane classification and traffic sign recognition [6]…. According to the architecture of a convolutional neuronal network, the image passes through many layers until the output one. During each layer learning, the DNN attempts to detect different features by several filters, which are applied to each training image at different resolutions, and the output of each convolved image is used as the input to the next layer. The filters can start as very simple features, such as brightness and edges, and increase in complexity to features that uniquely define the object as the layers progress [7].

In this work, we are interested on one of most popular algorithms for deep learning namely convolutional autoencoder CAE. In such architecture, the network reproduces its inputs at the output layer by learning efficient coding of unlabelled data using an unsupervised learning technique [8]. Moreover, the CAE contains an internal "bottleneck" layer, containing fewer nodes than input or output layers, which forces the network to develop a compact representation of the input data by ignoring insignificant features thus forces the learned representations to assume useful properties [8]. Autoencoders are applied to many problems, from classification [9], facial recognition [10], feature detection [11], anomaly detection to acquiring the meaning of words [12,13]. Autoencoders are also generative models which means that they can randomly generate new data that is similar to the data in its input [14].

The main core of the following paper is organized around four sections: section2 summarizes some related works in this field. Section3 describes in details the proposed approach including the architecture of the autoencoder and the steps followed to achieve this work. Section 4 reports experimental results and compares the obtained values of metrics with those of baseline methods. Finally, section5 concludes this paper with discuss the remaining challenges and possible directions for future works.

2. Literature Review

Digital image compression is a topical research area in the field of image processing due to its large number of applications such as aerial surveillance, recognition, medicine, multimedia… and every day the number of images treated is growing rapidly which indicates that there is an evident and urgent need to develop more and more efficiency image compression schemes. Moreover, this large amount of images need to be secured and easily transmitted. For all those reasons and others, this domain has received significant attention from researchers whose major focus is to develop different compression schemes that provide good visual quality with fewer bits to represent digital images in order to reduce the memory required for their storage.

For many years, image compression techniques based on wavelet transform have been considered as a powerful tool that compress images at higher compression ratio thanks to their particularity to be a time-frequency representation [15]. Every stage of image compression process mainly those based on wavelet transforms has received an important attention from scientists and many works have emerged during the last decade proving that this algorithms are a valuable tool for image processing [16, 17, 18, 19, 20]. JPEG2000 (Join Photographic Experts Group 2000) standard is the most famous compression algorithm based wavelet transform, developed to replace its predecessor JPEG discrete cosine transform (DCT)-based method [21, 22, 23, 24, 25, 26]. For many years and thanks to several works emerged later, this algorithm have proved its efficiency as a powerful compression tool that produce the best quality or performance consequentially it is widely used in various domains such Internet, digital photography, remote sensing, mobile, medical imagery, digital libraries/archives, and E-commerce [27, 28, 29, 30, 31].

The last years and due to the Artificial Intelligence AI advances especially Machine learning ML, different image compression techniques combining multi-resolution aspect of wavelets to parallel processing of data and training process of neuronal networks emerged and the list is so huge to be listed [32, 33, 34, 35, 36]. In fact, neural network implementation in every stage of an image compression system has received an important attention from scientists and many works have been developed during last decade supporting combination between the two algorithms [37, 38, 39, 40]. Furthermore and since the rise of Deep Learning, networks based deep layers have widely been involved in image compression process especially unsupervised called autoencoder. In the beginning, Mark A. Kramer [8] developed a Nonlinear Principal Component Analysis technique NLPCA to identify and remove correlations among problem variables in the purpose of dimensionality reduction. Later, this technique has been considered as the pioneer architecture of the autoencoders. Thus, began the ere of compressive autoencoder and many works have emerged with promising results surpassing existing lossy image compression algorithms [41, 42, 43, 44, 45, 46, 47].

In the field of image compression based wavelet transform and autoencoder algorithms, Y. Chuxi and colleagues [48] proposed a deep image compression approach in wavelet transform domain based on high frequency sub-band prediction by low-frequency sub-band. In the same time, all sub-bands feed different autoencoders which encode them with a conditional probability model for entropy coding moreover, the entire training process is unsupervised, and the auto-encoders and the conditional probability model are trained jointly. The propose model outputs results that outperform JPEG, JPEG2000, BPG, and some mainstream neural network-based image compression. H. MA and colleagues [49] proposed iWave++ a versatile end-to-end optimized image compression scheme based on wavelet transform in which a trained autoencoder converts images into coefficients without any information loss after, the coefficients are optionally quantized and encoded into bits. Their experimental results obtained using the Kodak dataset demonstrate that lossy iWave++ leads to 17.34 percent bits saving over BPG and lossless iWave++ achieves comparable or better performance than FLIF. T. Williams and colleagues [50] developed a scheme for image classification that converted data to the wavelet domain, consequently important features learning occurred over differing low to high frequencies and by processing the fused features mapping led to advance in the detection and classification accuracy. In spite of their proposed methods having limitations in their present structure, shown results demonstrate that their wavelet-based ensemble network would perform at a greater accuracy and comparable to greater computational cost than traditional deep neural network methods. A. Paul and colleagues [51] proposed a denoising method composed by dual branch deep neural network based architecture working on wavelet-transformed bands to remove multiple frequently encountered noise patterns from hyperspectral images. Experimental results demonstrate the superior performance of the proposed network compared to other state-of-the-art denoising methods with PSNR 36.74, SSIM 0.97 and overall accuracy 94.03 %. Q. Feng and colleagues [52] developed a model based on Haar wavelet transformation that used as feature extraction for data compression and a pseudoinverse learning algorithm based autoencoders to identify and recognize images. Experimental results show that the proposed model, compared with filters, CNN, and pseudoinverse learning autoencoders, takes less training time, at the same time it acquires comparative recognition accuracy. Q. Zho and colleagues [53] proposed a wavelet loss function to better generate and reconstruct images. In fact, wavelet transform is applied to the reconstructed image loss function of the auto-encoder, and the frequency characteristics of the decomposed image are used to constrain it. The authors conducted their comparative experiments on two larger-size image datasets (FaceSrub, COIL20) and a small-size image dataset (Fashion_MNIST), and proved the effectiveness of the wavelet loss function. At the same time, they proposed a new image quality index called wavelet high-frequency signal-to-noise ratio (WHF-SNR), which can better measure the quality of the reconstructed image of the auto-encoder. H. Luo and colleagues [54] presented in their paper a novel scheme to learn high-level representative features and conduct classification for hyperspectral image (HSI) data in an automatic fashion. Experimental results on two real HSI data sets demonstrate that the proposed strategy improves classification performance in comparison with other state-of-the-art handcrafted feature extractors and their combinations.

3. Work Methodology

3.1. Scheme of paper idea

The aim of this present work is in first hand, to develop an efficient image compression system which combines the features of both wavelet transform and convolutional autoencoder and in second hand to design an effective loss compression scheme that is suitable to high-dimensional data. In fact, in one side we want take advantages from the wavelet decomposition as a multiresolution time-frequency representation that uses both time and the frequency of the signal to analyze. In the other side, we exploit the nonlinear dimensionality reduction aspect of unsupervised deep neural network autoencoder. Thus, we propose a new approach as a mixture of the two techniques illustrated by the scheme in figure 1, where we summarized all the steps followed:

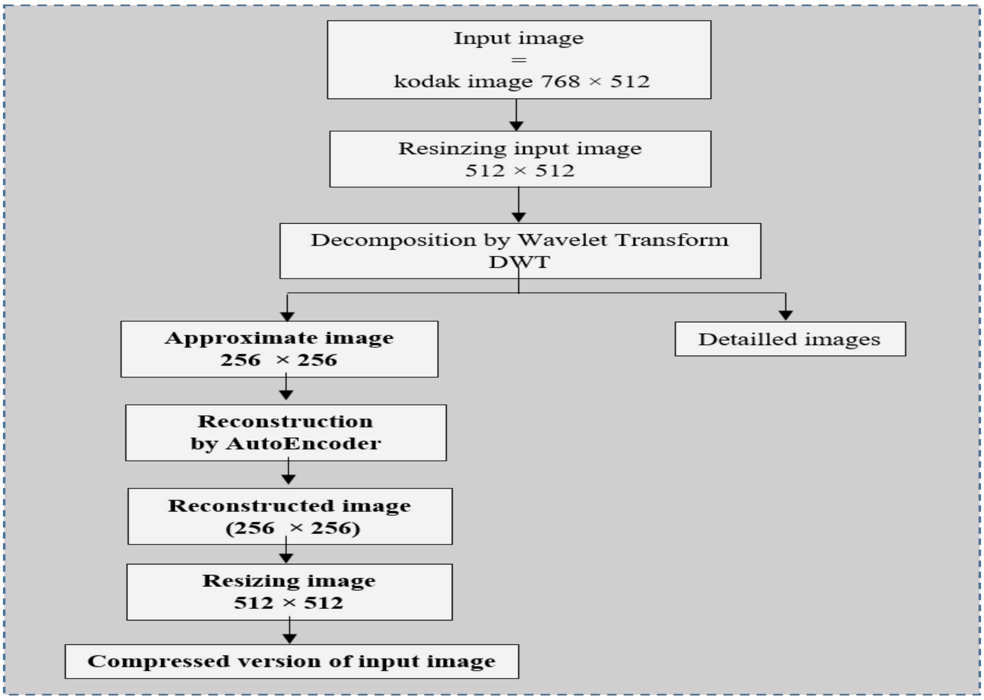

Figure 1: Steps illustrating the proposed scheme.

To achieve our work we followed the following steps according to the previous scheme:

- An input Kodak image with initial size 768 × 512 is resized to 512 × 512.
- The study is divided into two parallel parts: the first one the digital image is decomposed by a DWT into sub-band in the horizontal and vertical directions in one level. From this emerge three detailed sub-images: horizontal high-pass sub-image, vertical high-pass sub-image and diagonal high-pass sub-image and one approximate low-pass sub-image as indicated in figure 2:

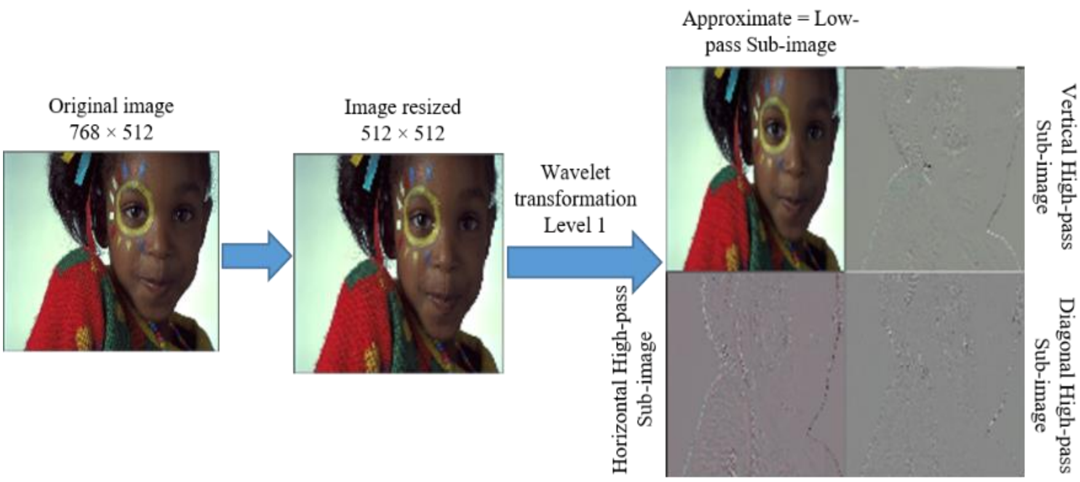

Figure 2: Wavelet image decomposition in Level1.

- In this stage of the research, our interest focuses in the approximate image. The idea is to put this image as input layer for a CAE to reproduce it at the output layer. In other words, the goal is to reconstruct the approximate image by forcing the CAE to determine a compressed version of the image in its input by considering only meaningful features and with the lowest amount of loss.
- To evaluate the compression system performance, we consider the following metrics:

- CR (Compression Ratio) is defined as [16, 17]:

$$CR = \frac{Size\ of\ original\ image}{Size\ of\ compressed\ image} \tag{1}$$

- MSE (Mean Square Error): calculates the mean square error between each pixels for the two images to compare. It is defined by the following equation [16, 17]:

$$MSE = \frac{1}{C \times M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(I_{ij} - J_{ji}\right)^2 \tag{2}$$

Where C is the number of channel whereas $M \times N$ is the size of each image.

- SSIM (Structural Similarity Index Metric) measures similarities within pixels of two images. This measurement involves luminance differences, contrast differences and structural variations and defined as [57, 58, 59]:

$$SSIM(x, y) = \frac{\left(2\mu_x\mu_y + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \tag{3}$$

With : μ_x Pixel sample mean of x;

μ_y Pixel sample mean of y;

σ_x^2 Variance of x;

σ_y^2 Variance of y;

σ_{xy} Covariance of x and y;

$c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$: Two variables to stabilize the division with weak denominator; L the dynamic range of the pixel-values

$k_1 = 0.01\ and\ k_2 = 0.03$

This structural metric ranges from minus one (opposite contrast) to zero (completely different) to one (completely identical) [57, 60].

- PSNR (Peak Signal to Noise Ratio) measures, in decibels, the perceptual quality of the image compressed on the ter of distortion. It is commonly used to quantify reconstruction quality for images and video subject to lossy compression and defined as [55, 56, 57]:

$$PSNR = 10log_{10}\left(\frac{MAX^2}{MSE}\right) \quad (4)$$

Where MAX is the maximum valid value for a pixel

- As indicated in the scheme relating to the steps followed to achieve this work, in some cases we need to resize the images used. To accomplish this stage we opted for the interpolation algorithms based on the use of interpolation kernel that calculates the value of a pixel using a weighted average of neighboring pixel values. These algorithms are efficient in increasing image resolution as well as in decreasing it [61]. We tried several interpolation algorithms finally, we choose to work with the bi-cubic algorithm because of the greater number of known pixel values considered while estimating the desired value which makes it the slowest algorithm in the terms of processing time whereas it produces noticeably sharper image with best quality. In effect, this algorithm takes into account the 16 pixels surrounding the considered pixel and use an interpolation kernel 4×4 to calculate the value of the desired pixel. Figure 3 demonstrates the two process of increasing or reducing the image:

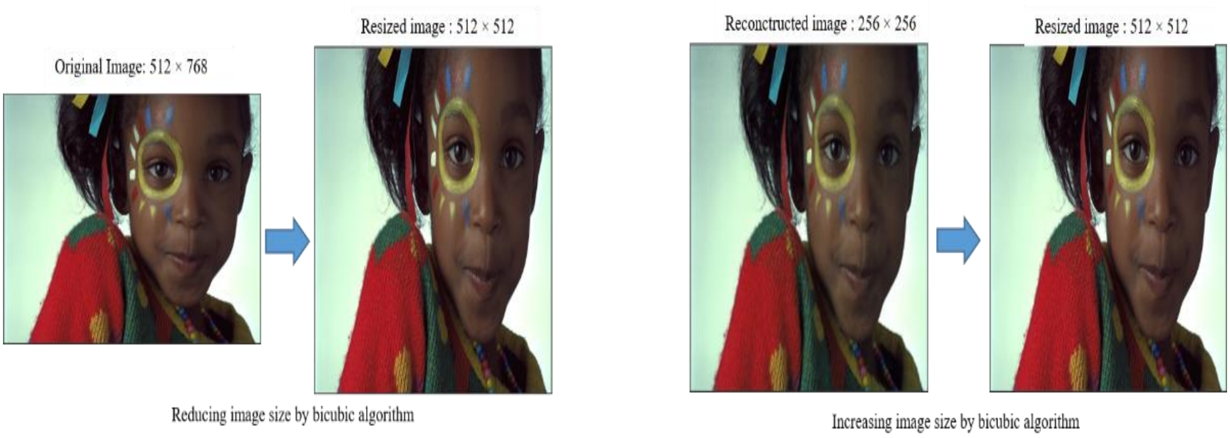

Figure 3: Resizing the original image by bi-cubic algorithm.

3.2. Proposed AutoEncoder Architecture

As mentioned before, the main idea of this work is to reconstruct the approximate image issue from decomposing the original image by wavelet Bior1.1 using an unsupervised deep neural network namely autoencoder defined with two dense layers: an encoder, which compresses the images into a latent vector, and a decoder that reconstructs the original image from the latent space. The raison of using an autoencoder architecture to reconstruct images is because during the training process the CAE network learns different features at different depth and as we go deeper and deeper it learns higher features which allows it to be more rich in features therefore we propose for this work a CAE with the architecture shown in figure 4.

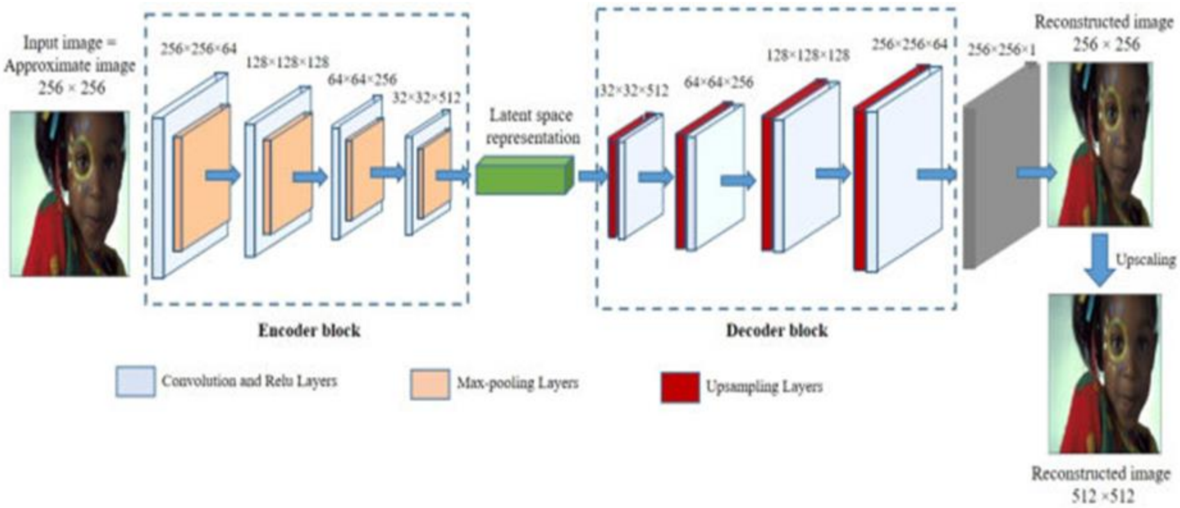

Figure4: Architecture of the autoencoder used to reconctruct the approximate image.

As indicated on this figure, the encoder part is structured around four convolutional blocks that have the following characteristics:

- The 1st convolutional block contains two layers working successively. The first one performs convolutional operations between the input data and 64 filters of 3×3 and generates 64 features maps activated by the nonlinear function ReLU (Rectified Linear Unit) defined as [62, 63]:

$$ReLU = max(x, 0) = \begin{cases} x, & if\ x \geq 0 \\ 0, & if\ x < 0 \end{cases} \qquad (5)$$

This activation function performs a nonlinear threshold operation, where any input value less than zero is set to zero. The second is a sampling layer obtained after performing maxpooling operation that acts as down-sampling function by dividing the input to pooling regions and computes the maximum value of each region. All these operations are carried out with the stride parameter set to 2.

- The 2nd, 3th and the 4th convolutional blocks have the same architecture as the previous one with the difference that the number of filters is increased to 128, 256 and 512 respectively. The output of the last convolutional block generates the latent features representation, which preserves only the most relevant aspects of the input image.

The second part of the CAE called decoder has four convolutional blocks aiming to reproduce the input image from the latent space. The layers have the following characteristics:

- The 1st to the 4th convolutional blocks represent the opposite side of the encoder and each block contains transpose convolutional layer, with the same number of filters and kernels as the encoder blocks in order to upsample the data by 2. Each block is followed by convolutional layer using ReLU activation function.

- An extra convolutional layer is added at the end of the CAE architecture adopted for this work. Its purpose is to output the reconstructed image representing a compressed version of the original image in the input of the network. This last layer contains one filter of size 3×3 and a stride of 1 with Sigmoid activation function defined as [62, 63]:

$$\boldsymbol{Sigm(x)} = \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e^{-x}}} \qquad (\mathbf{6})$$

We choose to train our CAE using Adam optimizer [64] and MSE (Mean Square Error) as Loss function. For each image in the dataset Kodak, The network training continues for 1000 epochs, which allows us to achieve an accuracy between 97% and 99% to reconstruct the approximate image. This accuracy is sufficient to get an image quality visually acceptable. Also, the network is trained with the Batch normalization algorithm to get normalized data as it is the best mean to regulate the network and to make it more strong [65].

3.3. Compression by the baseline algorithm

It is well known that wavelet based image representation has recently emerged as a powerful tool that in image processing and computer vision that investigates many fields such as compression, detection, recognition, image retrieval and colleagues Based on our hypothesis, baseline compression based wavelet consists, in the first set, in image decomposition in level one using the Bior1.1 wavelet. In fact, wavelet transform decomposes the digital image into sub-band in the horizontal and vertical directions. Low and high pass filters banks are applied to the image along rows and columns separately from this emerge three detailed sub-images: horizontal high-pass sub-image, vertical high-pass sub-image and diagonal high-pass sub-image and one approximate low-pass sub-image [17,19]. In the second step, quantization technique is applied to the spectral coefficients or the approximate image and finally a Huffman binary code is used to convert information under binary shape. The process of the compression by the baseline method is illustrated in figure 5.

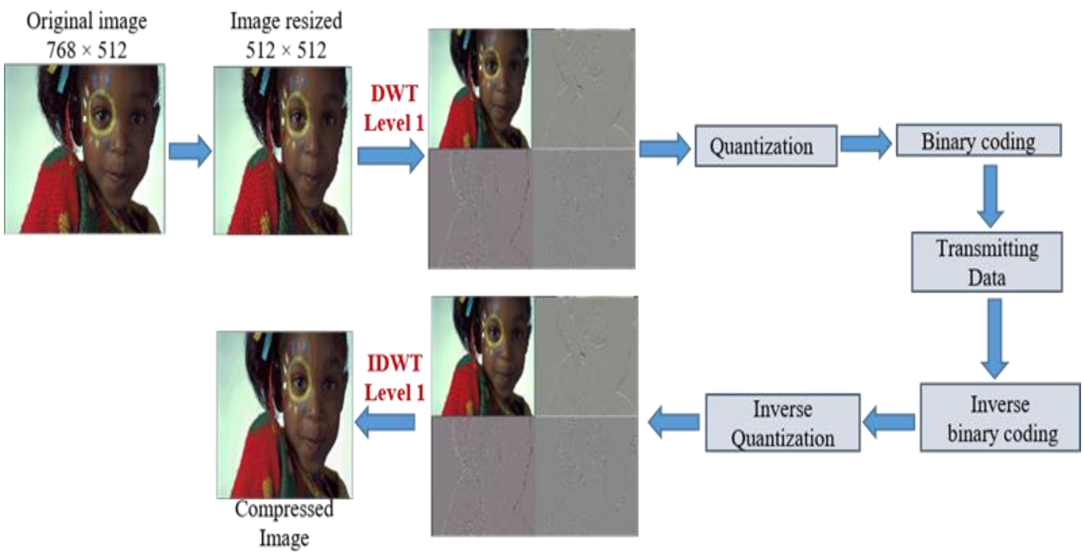

Figure5: Baseline compression process based Wavelet transform.

4. Results and Model Evaluation

As mentioned above, the core of this work is a convolutional autoencoder used to reconstruct the approximate image after decomposing the original image by the wavelet.

This last process allows us to eliminate some details useless and to keep the low-pass sub-image that we put as the input layer for an autoencoder to reconstruct it again. This second step contributes to extract more insignificant data in the approximate image.

The model proposed have been validated through different experimental results that had been carried out on the particular Kodak dataset.

For the experimental purpose, Python 3.10 platform was considered and executed in Pycharm EDI with some basic packages taken like Numpy, Pywt, Matplotlib and others. In addition, the data analysis packages like Keras, Skimage and Tensorflow was taken in an i7 core processor with 2.8 GHz speed and with 16 Go RAM.

We choose to configure our autoencoder architecture using convolutional layers and increasing the number of filters as we go deeper through the network to make the last one more rich in information and to extract more interesting features from the original image.

During the encoder stage, the information flows through the network from one layer to another extracting features and creating maps that output a compressed representation (latent space) used in the decoder stage to replicate the image in input.

As a simple feedforward neural network, the autoencoder training is performed through backpropagation of the error which means that it is trained to minimize reconstruction errors, named function Loss, between the input image and its reconstruct copy. Furthermore, the weights and biases are initialized randomly and iteratively updated during training stage.

From this work, emerge some results summarized in table1 that lists the four metrics coefficients used to evaluate this method. In the first hand, we measure the parameters: MSE, PSNR, CR and SSIM by comparing the original image and its reconstructed copy issue from the scheme proposed and in the second hand we calculate the same metrics between the original image and its compressed version using the baseline method based on wavelet decomposition, quantization and Huffman coding.

Table 1: Evaluation Parameters of compression for both methods: our proposed and baseline.

	MSE		PSNR		CR%		SSIM	
Image Kodak	Proposed method	Baseline method	Proposed method	Baseline method	Proposed method	Baseline method	Proposed method	Baseline method
Kodim01	266.780	364.44	23.869	22.515	15.79	4.27	0.747	0.658
Kodim02	141.924	266.45	26.61	23.874	19.147	4.42	0.83	0.781
Kodim03	109.733	147.796	27.727	26.434	12.373	5.16	0.895	0.813
Kodim04	80.071	124.281	29.096	27.187	16.441	5.79	0.869	0.726
Kodim05	235.378	2587.46	24.413	14.002	9.878	3.31	0.827	0.326
Kodim06	258.018	325.599	24.014	23.004	18.242	5.16	0.784	0.709
Kodim07	102.778	118.073	28.012	27.409	8.549	7.23	0.762	0.839
Kodim08	443.67	692.38	21.664	19.727	11.062	7.53	0.762	0.636
Kodim09	111.337	162.71	27.664	26.016	20.798	6.32	0.887	0.814
Kodim10	163.473	171.623	25.996	24.046	16.199	6.51	0.882	0.791
Kodim11	135.003	224.096	26.827	24.626	16.165	4.49	0.804	0.647
Kodim12	198.662	146.228	25.15	26.48	18.998	6.34	0.859	0.768
Kodim13	424.451	4025.62	21.853	12.082	16.787	3.14	0.722	0.274
Kodim14	154.041	293.215	26.254	23.459	14.185	4.14	0.831	0.637
Kodim15	119.109	293.491	27.371	23.455	17.527	6.53	0.853	0.653
Kodim16	398.825	163.085	22.123	26.006	17.109	5.13	0.801	0.730
Kodim17	69.611	176.769	29.704	25.657	12.793	4.22	0.898	0.657
Kodim18	146.91	1712.34	26.46	15.795	14.023	4.22	0.827	15.795
Kodim19	146.999	196.129	26.458	25.205	18.842	6.34	0.824	0.726
Kodim20	91.749	180.134	28.505	25.575	20.135	6.56	0.885	0.792
Kodim21	184.183	195.475	25.478	25.22	21.13	4.93	0.862	0.788
Kodim22	195.859	216.804	25.211	24.77	16.736	5.04	0.819	0.711
Kodim23	187.225	194.009	25.407	24.982	11.692	7.50	0.919	0.812
Kodim24	360.836	2032.22	22.558	15.051	10.902	4.51	0.809	0.393

As we can see all the metrics results obtained using the proposed approach are greater than those achieved by the baseline method. Indeed, let us give some insight on each one:

- Despite the fact that the PSNR coefficient has been shown to perform poorly, compared to other quality metrics, when it comes to evaluate the quality of images after a lossy compression [57]. In this work, we believe that the evaluation by this parameter is very conclusive since there is no contradiction between the PSNR values obtained and the visual side of the compressed image. In addition, as we can see, the PSNR metric achieve in the case of most images, best values by our proposed method.
- Moreover, the mean square error MSE and the CR parameters between the uncompressed and compressed image are much outperformed using the proposed scheme.
- Another most interest parameter is the structural similarity index SSIM. This coefficient is based on perception model that considers image degradation as perceived change in structural information and takes advantage from tight inter-dependencies between pixels [57, 59]. According to the SSIM values shown in the table1, we can confirm that the measurements achieved by the proposed method greatly exceed the results obtained by the standard image compression technique. This prove that the

compression system scheme proposed yield high-quality compressed image with best values of the SSIM parameter.

There is an exception for two images kodim12 and kodim16. As we can see, PSNR and MSE values are better in the baseline method but in the point of view compressed image visual quality, the result with the proposed method is better as indicated by the following figure 6. We assume that the reason is that with 1000 epochs, the proposed autoencoder architecture achieves only an accuracy of 97% (the lowest value) to reconstruct the approximate image for both kodim12 and kodim16. We suppose that with more epochs we can fix this and improve the metrics results.

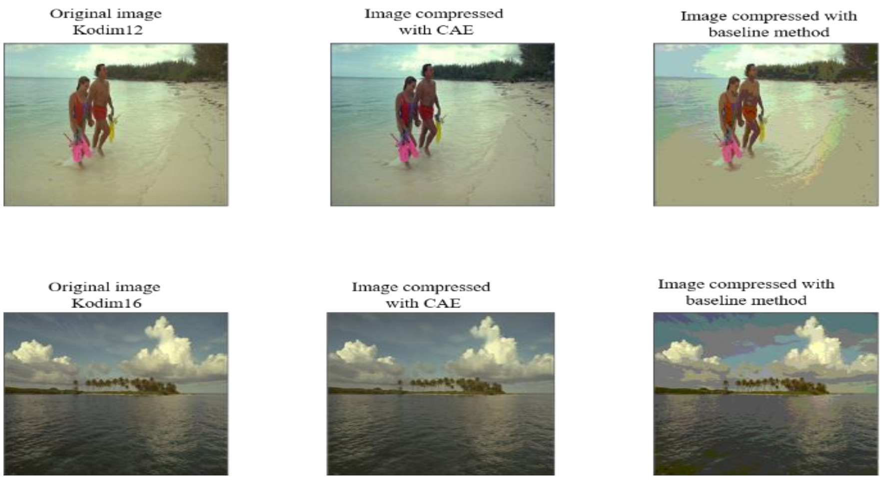

Figure 6: Kodim12 and Kodim16 compressed by proposed CAE and the baseline method.

To clarify more the meaning of the results presented in the table above, here (Figure7) some perceptual examples from the database Kodak. We can see clearly that the compressed image version obtained using the proposed method is very close and similar to the original image. Visually talking, we can confirm that the approach proposed in this paper performs in quality and in terms of metrics, the baseline method based in the classical scheme of a compressed image system. We choose images from various contains, contrasts … to prove the new compression scheme effectiveness.

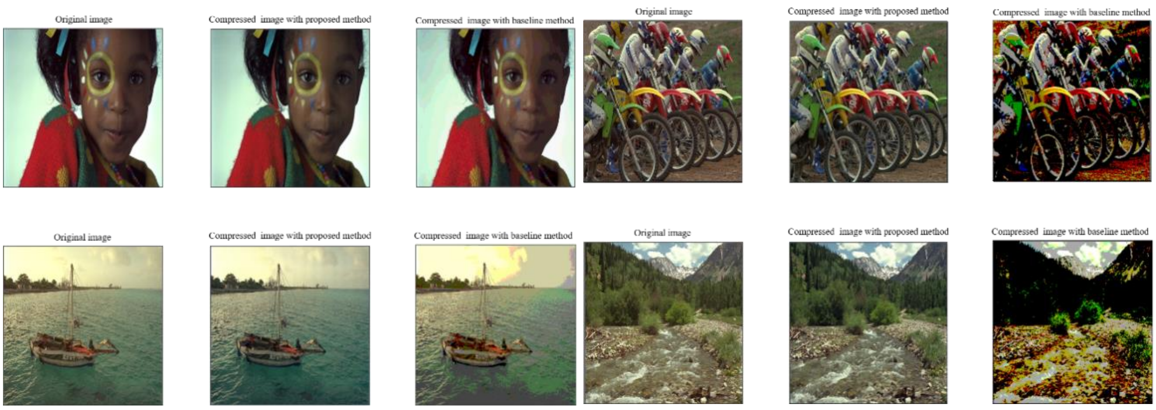

Figure 7: Comparative representation of compressed images: CAE vs baseline method.

5. Conclusions

In this work, we propose a new scheme for image compression based on two powerful techniques namely digital wavelet transform DWT and the unsupervised convolutional autoencoder. The purpose of this combination is to develop an efficient image compression system that takes on count advantages of each technique. In fact, in one side, we have the wavelets a strong tool recently emerged and supported by several research papers as a powerful mean to compress images with a good visual quality and fewer bits. In the other side, we have the convolutional autoencoder a neural network based on convolutional layers that captures recently, more and more significant attention of researchers. To achieve our work we use the database of Kodak that contains 24 uncompressed images commonly used to evaluate new image compression approaches and to evaluate the performance of the model proposed we consider the four known performance metrics: PSNR, MSE, CR and SSIM.

Basis on our hypothesis, the uncompressed image is decomposed in level 1 using the DWT algorithm in order to eliminate the unnecessary details and to keep only the approximate image, process that deletes redundant information from virgin version of the image. Next, we configure and implement an autoencoder network to reconstruct the approximate image; this second step helps to remove all irrelevant information that remains. To accomplish successfully this work, we opted for a deep architecture with four convolutional layers in each part and a maxpooling function as pooling process. From this proposed idea emerged results, which demonstrates that the best performances are obtained using the proposed scheme. The purpose of the work presented in this paper is to demonstrate that with the same decomposition level using wavelet transform and thanks to the deep learning autoencoder, this method can achieve significantly higher visual quality and good compression parameters when compared to a classical image compression technique. In fact, in present work we attempt to show that with minimal level decomposition we can outperform traditional compression approach based on wavelet transform such as JPEG2000. We weren't preoccupied by the real time running as our goal was to create a scheme that allow the storage space reduction in order to save a lot of images. Nevertheless, in areas where time is crucial factor to take into account like medicine or aerial surveillance this idea remains to be improved. Even in this case, we assume that using parallel computing with the aid of multiple suitable GPU, this proposal can undoubtedly succeed. Despite all this, we proved that our compression scheme.

based on convolutional autoencoder combined to wavelet transform outperforms the traditional image compression mainly in terms of similarity, as the compressed version of images is visually indistinguishable from uncompressed ones. Our proposed method have limitations in its present form. Firstly, the proposed autoencoder architecture is not suitable for grayscale images. In fact, during experiment process, we tried to apply our CAE to images with one channel and the reconstructing accuracy was very low under 34%. Secondly, we use only one type of wavelet basis (Bior1.1), when others possibly perform better.

As future work, we will manage to overcome this limitations and to improve the proposed system performance by decomposing original image in high levels trying others wavelets, and configuring deeper autoencoder as well as taking into consideration the real running time.

References

[1] L. Xiang and J. Shihaho, " Neural Image Compression and Explanation," *arXiv: 1908.08988v2 [cs.CV]* December 2020.

[2] K. He, X. Zhang, S. Ren and J. Sun, " Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR). pp. 770-778, 2016.

[3] J. Devlin, M.-W.Chang, K, Lee and K. Toutanova, " Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding," arXiv preprint arXiv: 1810.04805, 2018.

[4] E. Battenberg, J. Chen, R. Child, A. Coates, Y. Gaur, Y. Li, H. Liu and al. "Exploring Neural Transducers for End-To-End Speech Recognition," arXiv preprint arXiv: 1707.07413, 2017.

[5] M. Alam, M.D. Samad, L. Vidyaratne, A. Glandon and K.M. Iftekharuddin, " Survey on Deep Neural Networks in Speech and Vision Systems," Neurocomputing 417 , 302–321, 2020.

[6] G. De-Las-Heras, J. Sánchez-Soriano and E. Puertas, " Advanced Driver Assistance Systems (ADAS) Based on Machine Learning Techniques for the Detection and Transcription of Variable Message Signs on Roads," Sensors, 21, 5866, 2021.

[7] A. Krizhevsky, I. Sutskever, and G. Hinton, " ImageNet Classification with Deep Convolutional Neural Networks, " in NIPS, 2012.

[8] A.M. Kramer, " Nonlinear Principal Component Analysis using Autoassociative Neural Networks, ". AIChE Journal. 37 (2): 233–243, 1991.

[9] P. Vincent, H. Larochelle, I, Lajoie, Y. Bengio and P-A. Manzag, " Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion, " Journal of Machine Learning Research. 11: 3371–3408, 2010.

[10] G.E, Hinton, A. Krizhevsky A and S.D. Wang, " Transforming auto-encoders, " In International Conference on Artificial Neural Networks, 14 (pp. 44-51). Springer, Berlin, Heidelberg, Jun 2011.

[11] A. Géron, " Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, " Canada: O'Reilly Media, Inc. pp. 739–740, 2019.

[12] L. Cheng-Yuan, H. Jau-Chi and Y. Wen-Chie," Modeling word perception using the Elman network," Neurocomputing. 71 (16–18): 3150, 2008.

[13] L. Cheng-Yuan, C. Wei-Chen, L. Jiun-Wei and L. Daw-Ran, "Autoencoder for words," Neurocomputing. 139: 84–96, 2014.

[14] P. Diederik, Welling, Max and Kingma,. "An Introduction to Variational Autoencoders ," Foundations and Trends in Machine Learning. 12 (4): 307–392. arXiv:1906.02691, 2019.

[15] S.G. Mallat, "A theory for multi resolution signal decomposition: The wavelet representation," IEEE Trans. Pattern Anal. Machine Intell, vol. 11, pp. 674-693, 1989.

[16] M. Rabbani and P.W. Jones, " Digital Image Compression Techniques,"vol.TT07, SPIE Press Book, Bellingham, Washington, USA, Feb 1991

[17] M. Antonini, M. Barlaud, P. Mathieu, I. Daubechies, " Image coding using wavelet transform, " IEEE Trans. Image Process. 1 (2), pp. 205-220, 1992.

[18] A. Said and W.A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees, " IEEE, Trans. Circuits syst. Video Technol. 6 (3), pp. 243-250, 1996.

[19] Q. Zang, " Wavelet Network in Nonparametric Estimation, " IEEE Trans. Neural Networks, 8(2), pp. 227-236, 1997.

[20] K. Ratakondu, and N. Ahuja, " Loless Image Compression with Multiscale Segmentation, " IEEE, Transactions on Image Processing, vol. 11, No. 11, pp. 1228-1237, 2002.

[21] G.K. Wallace, " The JPEG still picture compression standard," IEEE Trans. On Consumer Electron., vol. 38 (1), pp. xviii-xxxiv, Feb. 1992.

[22] W.B. Pennebaker and J. L. Mitcell, "JPEG: Still Image Data Compression Standard," New York: Van Nostrand Reinhold, 1993.

[23] M. Rabbani and R. Joshi, "An Overview of the JPEG2000 Still Image Compression Standard," Signal Processing: Image Communication Journal, Volume 17, Number 1, October 2001.

[24] A. Skodras, C. Charilaos and T. Ebrahimi, " The JPEG 2000 Still Image Compression Standard, " IEEE Signal Processing magazine. September 2001.

[25] D. Taubman and P. Marcellin, "JPEG2000: Image Compression Fundamentals, Practice and Standards," Kluwer Academic Publishers, 2001.

[26] D. Taubman, and M. Marcellin, "JPEG2000 Image Compression Fundamentals, Standards and Practice: Image Compression Fundamentals, Standards and Practice," Springer Science & Business Media. ISBN 9781461507994, 2012.

[27] J. Liang and R. Talluri, "Tools for robust image and video coding in JPEG 2000 and MPEG-4 Standards," in Proc. SPIE Visual Communications and Image Processing Conf. (VCIP), San Jose, CA, Janvier 1999.

[28] M.W. Marcellin, M. Gormish, A. Bilgin, and M. Boliek, "An overview of JPEG 2000," in Proc. IEEE Data Compression Conf., Snowbird, UT, pp. 523-541, Mar. 2000,

[29] D. Santa Cruz and T. Ebrahimi, "An analytical study of the JPEG 2000 functionalities," in Proc. IEEE Int. Conf. Image Processing (ICIP 2000), Vancouver, Canada, 10-13, vol. II, pp. 49-52, September. 2000.

[30] D. Santa Cruz, M. Larsson, J. Askelof, T. Ebrahimi, and C. Christopoulos, "Region of interest coding in JPEG 2000 for interactive client/server applications," in Proc. IEEE Int. Workshop Multimedia Signal Processing, Copenhagen, Denmark, p. 389-384, September 1999.

[31] A.T. Kouanou, D. Tchiotsop, R. Tchinda and Z.D. Tansaa, "A Machine Learning Algorithm for Image Compression with application to Big Data. Architecture: A Comparative Study," British Biomedical Bulletin. British Biomedical Bulletin. Vol.7 No. 1:316, 2019.

[32] Q. Zang, and A. Beneveniste, "Wavelet networks," IEEE Tans. Neural Networks, 7(1), pp. 889-898, 1992.

[33] S. Osowski, R. Waszczuk, and P. Bojarczak, "Image Compression Using Feed Forward Neural Networks- Hierarchical Approach," Lecture Notes in Computer Science, Book chapter, Springer – Verlag, 3497, pp. 1009-1015, 2006.

[34] Q. Zang, "Wavelet Network in Nonparametric Estimation," IEEE Trans. Neural Networks, 8(2), pp. 227-236, 1997.

[35] A.V. Singh, and K.S. Murthy, "Neuro-Wavelet based Efficient Image Compression Using Vector Quantization," International Journal of Computer Applications (0975-08887), vol. 49-N°.3, July 2012.

[36] K. Ahmadi, A.Y. Javaid, and E. Salari, "An efficient Compression Scheme based on Adaptive Thresholding in Wavelet Domain using Particle Swarm Optimization," signal processing: image communication 32, 2015, pp.33-39.

[37] V. Krishnanaik, G.M. Someswar, K. Purushotham and A. Rajaiah, "Implementation of Wavelet Transform, DPCM and Neural Network for Image Compression," International Journal of Engineering and Computer science ISSN: 2319-7242, vol. 2, issue. 8, pp. 2468-2475, August 2013.

[38] T. Denk, K. Perhi, and V. Cherkassky, "Combining Neural Network and the Wavelet Transform for Image Compression," Proceeding of Intl Conf, pp. 637-640, 1993.

[39] K. Dimililer and A. Khashman, "Image Compression using Neural Networks and Haar wavelet," Transaction on Signal Processing, ISSN: 1790-5052, vol. 4, issue: 5, May 2008.

[40] A.K. Alexandridis and A.D. Zaprani, "Wavelet Neural Networks: A Pratical Guide," neural networks

42, pp. 1-27, 2013.

[41] L. Theis and M. Bethge, " Generative Image Modeling Using Spatial LSTMs," Advances in Neural Information Processing Systems 28, 2015.

[42] L. Theis, A. van den Oord, and M. Bethge, " A Note on the Evaluation of Generative Models, " International Conference on Learning Representations, 2016.

[43] 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," International Conference on Learning Representations, 2016a.

[44] G. Toderici, D. Vincent, N. Johnston, S. J. Hwang, D. Minnen, J. Shor, and M. Covell, " Full Resolution Image Compression with Recurrent Neural Networks, ". Computer Vision and Pattern Recognition, arXiv:1608.05148v1, 2016b.

[45] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, "Variational image compression with a scale hyperprior," in Proc. of International Conference on Learning Representations, 2018.

[46] M. Tschannen, E. Agustsson, and M. Lucic, "Deep generative models for distribution-preserving lossy compression," in Proc. of Advances in Neural Information Processing Systems, 2018.

[47] F. Yang , L. Herranz, J. van de Weijer, J. A. Iglesias Guitian , A. M. Lopez and G. Mozerov, "Variable Rate Deep Image Compression With Modulated Autoencoder," IEEE Signal Processing Letters, Vol. 27, 2020.

[48] Y. Chuxi, Y. Zhao and S. Wang, " Deep Image Compression in the Wavelet Transform Domain Based on High Frequency Sub-Band Prediction," IEEE Computer Science. April 2019.

[49] H. Ma, D. Liu, N. Yan, H. Li and F. Wu, "End-to-End Optimized Versatile Image Compression With Wavelet-Like Transform," IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 44 issue: 3- 2020.

[50] T. Williams and R. Li, " An Ensemble of Convolutional Neural Networks Using Wavelets for Image Classification," Journal of Software Engineering and Applications. Vol.11 No.02, 2018.

[51] A. Paul, A. Kundu, N. Chaki and al, " Wavelet enabled convolutional autoencoder based deep neural network for hyperspectral image denoising," Multimed Tools Appl 81, 2529–2555, 2022.

[52] Q. Feng, Q. Yin and P. Guo, "Image Recognition With Haar Wavelet and Pseudoinverse Learning Algorithm Based Autoencoders," Journal of Physics: Conference Series, Volume 2278, 2022 6th International Conference on Machine Vision and Information Technology (CMVIT 2022) Feb. 25, 2022 Online.

[53] Q. Zhu, H. Wang and R. Zhang, "Wavelet Loss Function for Auto-Encoder," in IEEE Access, vol. 9, pp. 27101-27108, 2021.

[54] H. Luo, Y. Yan Tang, R.P. Biuk-Aghai, X. Yang, L. Yang and Y. Wang, "Wavelet-based extended morphological profile and deep autoencoder for hyperspectral image classification," International Journal of Wavelets, Multiresolution and Information Processing. Vol. 16, No. 03, 1850016, 2018.

[55] Q. Huynh-Thu and M. Ghanbari, M, "Scope of validity of PSNR in image/video quality assessment". Electronics Letters. 44 (13): 800. 2008.

[56] Q. Huynh-Thu and M. Ghanbari, "The accuracy of PSNR in predicting video quality for different video scenes and frame rates". Telecommunication Systems. 49 (1): 35–48. Janv 2012.

[57] Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli, "Image Quality Assessment: from error Visibility to Structural Similarity,". IEEE Trans Image Process;13:600-12, 2004.

[58] Z. Wang, E.P. Simoncelli and A.C. Bovik, "Multiscale structural similarity for image quality assessment," Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, Vol. 2. pp. 1398–1402, 2004.

[59] R. Rozema, H.T. Kruitbosch, B. van Minnen, B. Dorgelo, J. Kraeima and PMA Van Ooijen, " Structural Similarity Analysis of Midfacial Fractures—a feasibility study, " Quant Imaging Med Surg, 12(2): 1571–1578, Feb 2022.

[60] G.P. Renieblas, A.T. Nogués, A.M. González, N. Gómez-Leon and E.G. Del Castillo, "Structural Similarity Index Family for Image Quality Assessment in Radiological Images, " J Med Imaging (Bellingham);4:035501, 2017.

[61] P. Parsania1 and P. Virparia, "A Review: Image Interpolation Techniques for Image Scaling, " International Journal of Innovative Research in Computer and Communication Engineering. Vol. 2, Issue 12, December 2014.

[62] F. Farnoush, " Learning Activation Functions in Deep Neural Networks, " Ecole Polytechnique, Montreal (Canada) ProQuest Dissertations Publishing, 10957109, 2017.

[63] C.E. Nwankpa, W. Ijomah, A. Gachagan and S. Marshall, " Activation Functions: Comparison of Trends in Practice and Research for Deep Learning," arXiv:1811.03378v1 [cs.LG] 8 Nov 2018.

[64] P.D. Kingma and J. Lei Ba, ADAM: "A Method For Stochastic Optimization," arXiv:1412.6980v9 [cs.LG] 30 Jan 2017.

[65] S. Ioffe and C. Szegedy, " Batch normalization: Accelerating deep network training by reducing internal covariate shift," Machine Learning, arXiv:1502.03167[cs.LG], 2015.