A multi-scale convolutional neural network and discrete wavelet transform based retinal image compression

Dalila Chikhaoui[1], Mohammed Beladgham[1], Mohamed Benaissa[2], Abdelmalik Taleb-Ahmed[3]
[1]Information Processing and Telecommunication Laboratory (LTIT), Department of Electrical Engineering, Faculty of Technology-University TAHRI Mohammed Bechar, Bechar, Algeria
[2]Information Processing and Telecommunication Laboratory (LTIT), Department of Electrical Engineering, Faculty of Technology-University Abou Bekr Belkaid Tlemcen, Tlemcen, Algeria
[3]Polytechnic University of Hauts-de-France, Lille University, CNRS, Valenciennes, France

Article Info

Article history:

Received May 19, 2024
Revised Oct 15, 2024
Accepted Oct 30, 2024

Keywords:

CNN
Deep learning
Discrete wavelet transforms
Image compression
Retinal image

ABSTRACT

The different applications of medical images have contributed significantly to the growing amount of image data. As a result, compression techniques become essential to allow real-time transmission and storage within limited network bandwidth and storage space. Deep learning, particularly convolutional neural networks (CNN) have marked rapid advances in many computer vision tasks and have progressively drawn attention for being used in image compression. Therefore, we present a method for compressing retinal images based on deep CNN and discrete wavelet transform (DWT). To further enhance CNN capabilities, multi-scale convolutions are introduced into the network architecture. In this proposed method, multiscale CNNs are used to extract useful features to provide a compact representation at the encoding stage and guarantee a better reconstruction quality of the image at the decoding stage. Based on compression efficiency and reconstructed image quality, a wide range of experiments have been conducted to validate the proposed technique performance compared with popular image compression standards and existing deep learning-based methods. At a compression ratio (CR) of 80, the proposed method achieved an average peak signal-to-noise ratio (PSNR) value of 38.98 dB and 96.8% similarity in terms of multi-scale structural similarity (MS-SSIM), demonstrating its effectiveness.

Corresponding Author:

Dalila Chikhaoui
Information Processing and Telecommunication Laboratory (LTIT), Department of Electrical Engineering
Faculty of Technology-University TAHRI Mohammed Bechar
Bechar, Algeria
Email: chikhaoui.dalila@univ-bechar.dz

1. INTRODUCTION

As part of medical planning in health care systems, medical imaging plays a significant role in digital image processing. Medical imaging has made extensive use of retinal images. The diagnostic process for several eye diseases, including stroke, diabetic retinopathy, and glaucoma, depends heavily on the clarity of retinal images [1]. In addition, the difficulty of diagnosing some retinal diseases also requires advanced technologies that allow multiple specialized medical teams to effectively share, exchange, retrieve, and process data. Retinal images are further used for research purposes to help in the automatic labeling of diseased vessels i.e., automatic diagnostic [2], diabetic retinopathy [3], automatic retinal vessel segmentation [4]. The different applications of retinal images, whether for medical practice or research, have marked an increase in medical data acquired by the latest imaging technologies, in which the size of the images increases as the resolution requirement increases. Thus, storage and transmission systems are affected. Because of the large size of the data and the need for storage capacity, it is essential to compress the data for preservation in Hospital Information Systems (HIS) or picture archiving and communication systems (PACS) for future reference. A variety of mechanisms for efficient data processing are important as storage has become one of the greatest challenges. Moreover, the primary goal of telemedicine is to enable efficient remote analysis of medical images. Systems transmitting large amounts of medical imaging data cause complicated transmission processes. Transmission is mainly managed by reducing bit rates in response to a limited channel bandwidth [5]. Hence, it is important to develop adequate compression techniques that reduce the size of image data while preserving a reasonable level of clinical fidelity in order to overcome the limited bandwidth and storage resources.

At present, a variety of lossy compression techniques for medical images have been developed in the literature. These techniques are primarily divided into two categories: compression techniques based on conventional algorithms (non-deep-learning algorithms) and compression techniques based on deep learning [6]. Generally, conventional compression approaches (non-deep learning algorithms) are realized by combining different transforms jointly with a quantization step and entropy coding method [7], [8]. In addition, conventional approaches have been employed for medical image compression. Hänsgen *et al.* [9], investigated the effect of wavelet compression on automatic analysis tasks and the degradation of retinal image quality caused by different compression ratios (CR). Eikelboom *et al.* [10], the effect of JPEG and wavelet compression techniques on digital retinal images quality has been investigated. Krivenko *et al.* [11], proposed an image coder based on 32×32 pixels blocks discrete cosine transform (DCT) for retinal image compression. Mookiah *et al.* [12], they reported a quantitative assessment of the effects induced by the JPEG image compression algorithm on automatic vessel segmentation in digital retinal images. In the previous studies, as the compression ratio increases, conventional compression methods, such as JPEG, cause blocking artifacts or noise that degrades the quality of the decoded images. Some works proposed to overcome the problem by implementing post-processing or denoising based methods for retinal image processing. For example, Nazari and Pourghassem [13], suggested an approach based on pre-processing vessel extraction, and post-processing to enhance details in retinal images for the extraction of large and thin blood vessels using a 2D Gabor filter followed by linear Hough transformation. Javed *et al.* [14], a technique of edge-based enhancement of retinal images was presented. In this technique the images are processed and analyzed in the JPEG compressed domain to enhance the edges for disease diagnosis perspective. Salih *et al.* [15], presented an effective retinal image compression approach focused on the area of interest (ROI). This approach includes pre-processing with an adaptive median filter, segmentation with enhanced adaptive fuzzy c-means clustering, compression with integer multi wavelet transform, and set partitioning in hierarchical tree, to achieve better image quality.

Lately, deep learning methods have been successfully applied to image compression. Those methods have been proposed to benefit from an encoding and decoding module built of convolutional neural networks (CNN). Using CNNs the module enables dimensionality reduction and feature extraction during encoding, and enhanced reconstruction during decoding. Ballé *et al.* [16], presented a deep learning model for image compression by successively applying convolutional linear filters to nonlinear activation functions, while the rounding quantizer was replaced by a uniform quantizer to ensure an uninterrupted training process. Relying on model in [16], other deep learning architectures for image compression have been proposed, such as the one proposed by Cheng *et al.* [17]. In which, the authors introduced residual blocks into the architecture to increase the receptive field and improve compression performance of the model. Those deep learning-based models have outperformed conventional compression methods. Image compression technology based on deep learning was applied to medical images. Kar *et al.* [18], proposed a convolutional autoencoder architecture for medical lossy image compression to preserve diagnostically relevant features during compression. Sushmit *et al.* [19] suggested a convolutional recurrent neural network architecture to learn contextualized features for efficient X-ray image compression. A compression method for retina optical coherence tomography (OCT) images was developed in [20], which uses CNNs and skip connections with quantization to preserve fine structure features between the compression and reconstruction CNNs.

The previously reviewed techniques have revealed some shortcomings that need to be addressed. Starting with conventional compression algorithms which suffered from poor performance at high CR, the image quality was drastically degraded [21]. Therefore, much effort has been focused on improving the performance of these compression approaches using pre-processing and post-processing methods. Despite these performance improvements, these methods involve computationally expensive and time-consuming processes for solving optimal solutions. On the other hand, deep learning-based image compression techniques have demonstrated superior performance. However, their architectures may require deeper CNN or large models, resulting in computations that make the learning process slow. In addition, most of their feature extraction architectures rely on convolutional layers with one convolution each, which may lose some useful features in medical imaging terms. Deep learning models are incompatible with conventional codecs, hence their application is limited.

In response to the above shortcomings, this paper proposes a medical image compression technique based on a low complexity deep learning model and a discrete wavelet transform (DWT) based codec. Motivated by the advantages of CNNs like the ability to automatically detect features and their computationally efficient characteristics [22]. A CNN architecture with a low parameter count is designed for both encoding and decoding, enhanced by multi-scale convolutional layers. The proposed deep learning architecture is integrated with the DWT-based codec for effective performance at high CR, owing to DWT's computational efficiency and compact signal representation of the DWT, which is widely used in image coding [23]. The main contributions of this paper are summarized as follows. Initially, to improve the compression performance, the multi-scale CNN (MS-CNN) on the encoding side derives an optimal compact representation that holds important structural data from the original image. While on the decoding side, the second MS-CNN allows accurate reconstruction of the output image. Second, a DWT-based image codec residing between the encoding MS-CNN and decoding MS-CNN can be effectively utilized by taking a compact representation as input for further compression. Thirdly, we present a learning strategy for the MS-CNNs, which overcomes the problem of training interruption caused by non-differentiable quantization in the DWT codec. As demonstrated by the experimental results, the proposed technique outperforms existing techniques and standard compression techniques in terms of several metrics. Connecting the deep learning model with DWT codec using a compact intermediate representation allows the proposed compression technique to exhibit compatibility with other available image coding standards. To the best of our knowledge, this is the first study to use MS-CNNs to enhance the compression performance of conventional DWT-based codecs and achieve high CR with accurate medical image reconstruction. The succeeding part of the article is structured as follows: a description of the key components of the proposed approach is given in section 2. In section 3 provides a comparison and discussion of simulation results. Finally, conclusion is presented in section 4.

2. METHOD

Data compression based on deep learning is a promising research area. These techniques specialize in many aspects, including training and learning abilities [24]. Lossy compression by reducing dimensionality is one of the major categories of compression based on deep learning techniques, in which performance is comparable to or even better than standard codecs [16], [20]. Dimensionality reduction is accomplished by learning an invertible mapping between the quantized compact representation and the original data. This process relies mainly on deep architectures of CNNs which allow efficient feature extraction and exhibit a good representative ability [25]. Usually, a CNN is cascaded at both the encoding and decoding ends when building these deep learning models. In view of this, our presented lossy compression technique involves two MS-CNNs and a DWT image codec, as shown in Figure 1.

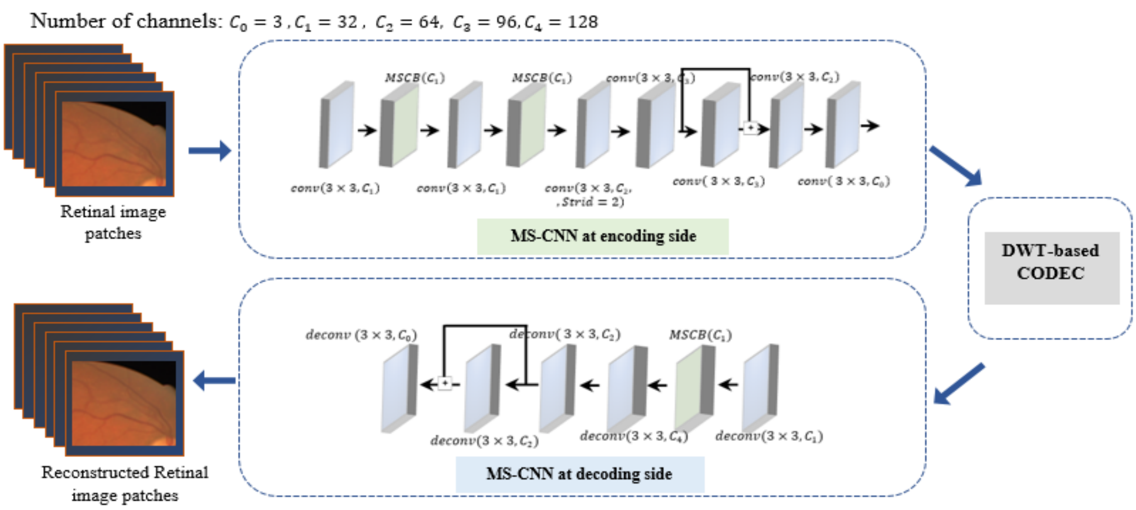
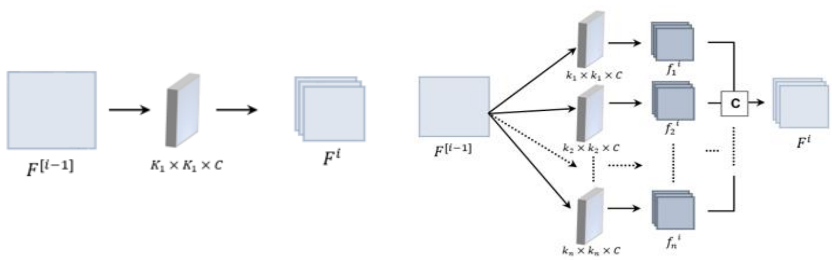


Figure 1. The proposed compression technique overall design

According to our proposed method, the input image will undergo the first MS-CNN residing at the encoding which generates a compact representation that preserves the structural information of the input. The generated compact representation is further encoded, since it allows the DWT based codec to achieve efficient compression with a high CR. On the decoding side, a second MS-CNN is applied in order to produce a more accurate and high-quality reconstructed image. The two networks cooperate to compress images at a very low bit rate while maintaining high quality. Unlike deep learning models with millions of parameters, our method incorporates a DWT-based codec, known for its ideal properties and low computational complexity in image compression tasks [23]. The following subsections provide more details about the key components of the proposed technique, such as the MS-CNN architecture, DWT based codec, the loss functions and training strategy.

2.1. Architecture of the MS-CNNs

Our approach further leverages CNN's robustness, by adding a multi-scale convolutional blocks, which have been previously employed for classification [22] and image super-resolution [23]. The convolutional layers automatically extract local features of input images through the learning process based on the given training dataset. In Figure 2, the main differences between multiscale convolution and basic convolution are illustrated in Figure 2(a) and Figure 2(b), respectively. In general, small size kernels tend to extract features with smaller scales, such as details, while coarse structures respond to kernels with large scales [26]. Therefore, employing multi-scale convolutions with small and large kernel sizes is favorable to guarantee an efficient extraction of the different scale features found in medical images.



(a) (b)

Figure 2. Difference between (a) basic convolutional layer of CNN and (b) multi-scale convolutional layer

The encoding and decoding MS-CNNs architecture details are given in Table 1. The encoding side MS-CNN incorporates basic and multi-scale convolutions and a down-sampling operation to reduce input dimensionality. The multi-scale convolutional block (MSCB) consists of a multi-scale convolutional layer with three parallel convolutions of varying kernel sizes, producing feature maps concatenated along the spectral dimension, as shown in Figure 3. Moreover, a down-sampling operation occurs by using stride convolution in order to contract spatial dimension by a factor of 2. In general, the encoding MS-CNN consists of 2 MSCBs and convolutional layers followed each by a rectified linear unit (ReLU) nonlinear activation function, as shown in Table 1. The architecture involves also skip connections, which have been considered one of the most efficient solutions for training deep networks [24]. The skip connection is clearly shown in Figure 1. The MS-CNN at the decoding side is composed of 5 weight layers and one MSCB, as shown in Table 1. In detail, three types of layers are present in the architecture, notably, deconvolution layers, up-sampling layer and multi-scale deconvolution. The compact representation undergoes the reverse process at the decoding side to secure accurate restoration of the original image.

Table 1. Architecture details of the MS-CNNs at the encoding and decoding sides

MS-CNN at encoding	MS-CNN at decoding
Conv (3×3, 32, stride=1), ReLU	Deconv (3×3, 32, stride=2), ReLU
MSCB (32)	MSCB (64)
Conv (3×3, 32, stride=1), ReLU	Deconv (3×3, 128, stride=1), ReLU
MSCB (32)	Deconv (3×3, 64, stride=1), ReLU
Conv (3×3, 64, stride=2), ReLU	Deconv (3×3, 64, stride=1), ReLU
Conv (3×3, 96, stride=1), ReLU	Deconv (3×3, 3, stride=1), ReLU
Conv (3×3, 96, stride=1), ReLU	
Conv (3×3, 64, stride=1), ReLU	
Conv (3×3, 3, stride=1), ReLU	

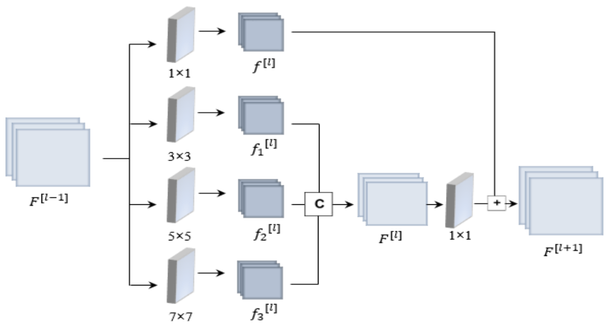


Figure 3. Proposed MSCB

2.2. DWT based codec

The Cohen-Daubechies-Feauveau 9/7 (CDF 9/7) biorthogonal wavelet transform serves as the foundation of the our codec to allow efficient compression. For image decorrelation, Antonini and Barlaud [27], demonstrated the superiority of the biorthogonal wavelet transform 9/7. JPEG-2000 codec and other image coding methods [23], [28], have widely relied on this transform. The CDF family of symmetric biorthogonal wavelets are distinguished by their compact support, biorthogonality, symmetry, and simplicity. According to Table 2, the CDF 9/7 wavelet has 9 coefficients in the low pass filter and 7 coefficients in the high pass filter.

Table 2. Filter coefficients of CDF 9/7 wavelet

i	Low-pass filter	High-pass filter
0	+0.6029490182363579	+1.11508705245700
±1	+0.266864118442875	-0.59127176311425
±2	−0.078223266528990	-0.05754352622850
±3	−0.016864118442875	+0.09127176311425
±4	+0.026748757410810	

The use of the wavelet for compression is dependent on the quantization step. A lossy compression approach uses quantization, which is adjusted to attain the desired CR. The quantized wavelet coefficients are subsequently entropy encoded via arithmetic coding [29]. By coding the most frequent symbols with fewer bits, it is more efficient than coding them all with the same bits number. Entropy encoding, specifically arithmetic encoding, provides lossless compression since the original data can be recovered in the decoder stage without affecting deep learning models. Therefore, we did not include the arithmetic coding in the training of networks in order to minimize unnecessary complexity. Our codec based on DWT can be summarized into the following steps:

– Decomposition of the input compact representation using 2D wavelet transforms (CDF 9/7).
– Quantization of the wavelet coefficients.
– Lossless compression using arithmetic encoding.

2.3. Loss functions and optimization

The objective is to optimize both MS-CNNs to achieve an efficient compression and a better image quality reconstruction. In order to optimize our model, a loss term needs to be minimized over the parameters of the proposed networks. The distortion between the input and reconstructed images represents the loss, and it can be expressed as:

$$L(\theta_i, \theta_j) = \frac{1}{N}\sum_{k=1}^{N}\left\|R(\theta_j, D(E(\theta_i, x_k))) - x_k\right\|^2) \quad (1)$$

in the (1), mean square error (MSE) is used in the loss function as a distortion term, with x_k denoting the input image. $E(.)$ and $R(.)$ indicate of the MS-CNN at the encoding side and MS-CNN at the decoding side, with θ_i, θ_j as their variables, respectively, whereas D denotes the DWT based codec. The input image x_k went through stages of compression, namely, MS-CNN for compact representation and DWT codec, then second MS-CNN for reconstructing the image. However, the rounding function incorporated in the DWT-based codec cannot be differentiated when performing the backpropagation algorithm. To address this issue, the training will be performed in two phases. The first phase involves training both networks without the

DWT codec, whereas the second phase involves finetuning the network on the decoding side taking into consideration the codec.

2.3.1. MS-CNNs training

Assuming a collection of input images x_k undergoes first an MS-CNN to learn an optimum compact representation and reserve the structural information. The reconstruction MS-CNN is then employed to recover the decoded image with high quality, hence the mean squared error loss function used for training can be defined as shown in (2):

$$L_1(\theta_i, \theta_j) = \frac{1}{N}\sum_{k=1}^{N}\left\|R(\theta_j, E(\theta_i, x_k)) - x_k\right\|^2 \quad (2)$$

where θ_i and θ_j denote the trainable variable, whereas N denotes the batch size.

2.3.2. MS-CNN fine-tuning

During MS-CNN reconstruction, the output image is reconstructed in a way that closely replicates the input image. Therefore, the decoded compact representation derived from compression network E then DWT based codec D will be passed through encoding MS-CNN R to learn more accurate reconstruction. The parameter $\hat{\theta}_i$ was fixed while the encoding network parameter θ_j was optimized, the loss function used for fine-tuning the MS-CNN can be formulated as:

$$L_2(\theta_j) = \frac{1}{N}\sum_{k=1}^{N}\left\|R(\theta_j, D(E(\hat{\theta}_i, x_k))) - x_k\right\|^2 \quad (3)$$

2.4. Evaluation metrics

In order to carry out a quantitative assessment of our method's performance, we adopted evaluation metrics based on image quality reconstruction and the efficiency of compression. The reconstructed image quality is evaluated using the peak signal-to-noise ratio (PSNR), MSE and multiscale structural similarity (MS-SSIM). The MSE and PSNR measure distortion between the original and reconstructed images to evaluate visual quality [30], as given in (4) and (5),

$$MSE = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left[X(i,j) - \hat{X}(i,j)\right]^2 \quad (4)$$

$$PSNR = 10\log_{10}\left(\frac{MAX^2}{MSE}\right) \quad (5)$$

where X and $\hat{X}$ are the input and reconstructed images respectively. M, N are the number of rows and columns of the image, while MAX is the maximum value of pixel in the image.

Similarity is a resolution image quality assessment method which computes relative quality scores between a reference reconstructed image [30]. Measurements at different scales can be combined to obtain an overall MS-SSIM evaluation, as shown in (6),

$$MS - SSIM = l_M(x, \hat{x}).\prod_{J=1}^{M} c_j\,(x, \hat{x})s_j(x, \hat{x}) \quad (6)$$

where x and $\hat{x}$ represent the original image and the reconstructed image respectively. M denotes the highest scale, $l_M(x, \hat{x})$, $c_j(x, \hat{x}), and\ s_j(x, \hat{x})$ refer to the luminance, contrast, and structure comparisons at the j-th scale, respectively.

Image compression efficiency is evaluated by CR, bits per pixel (bpp), and space savings (SSs). The CR and bpp give a straight notion of compression degree associated with the amount of data [31], [32]. The SSs is another metric to evaluate the performance of compression technique, it indicates the gained amount of storage space from saving the compressed data [33]. The CR, bpp, and SSs are given in (7)-(9), respectively:

$$CR = \frac{\text{Size of uncompressed image}}{\text{Size of compressed image}} \quad (7)$$

$$bpp = \frac{number\ of\ bits\ in\ the\ compressed\ image}{number\ of\ pixels\ in\ the\ original\ image} \quad (8)$$

$$SSs = \left(1 - \frac{Compressed\ Size}{Uncompressed\ Size}\right) \times 100 \quad (9)$$

3. RESULTS AND DISCUSSION

In this section, the results of our experiments are presented and discussed, showcasing the efficacy of our method. The implementation of the experiments has been carried out on the NVIDIA Tesla K80 GPU provided by Google Colab. Keras with a TensorFlow backend are used to build our network architecture. Table 3 provides a summary of the main simulation parameters of the MS-CNNs that were used in the experiment. For optimization purposes and to minimize the loss functions we used the Adam optimizer [34]. The learning rate is initialized by 1.0E−3 value and reduced using a learning rate scheduler by a factor of 2 based on metric improvement. The networks were trained for 400 epochs and fine-tuned for another 50 epochs.

Table 3. The main simulation parameters of the experiment

Parameter	Value
Learning rate	1.0E−3 to 1.0E−6
Epochs	450
Batch size	8
Input size	128×128

For experimental purposes, we utilized retinal images from two publicly available datasets. The first, the digital retinal images for vessel extraction (DRIVE) database [35], contains 40 color fundus images with a resolution of 565×584 pixels. We also randomly selected 40 images from the ORIGA-light database [36], which includes 650 high-resolution images from the Singapore Malay Eye Study (SiMES). To prevent memory overflow and optimize the image compression model, each image was cropped to patches of 128×128 pixels and normalized before compression. The dataset was then divided into training, validation, and testing sets, with 80% for training and 20% for validation and testing.

Test dataset images were utilized to evaluate the proposed compression method by comparing it with JPEG, JPEG2000, and existing deep learning-based methods. JPEG and JPEG2000 were selected because of their transform reliance, with JPEG2000 using the CDF 9/7 wavelet also employed in our codec. Among the deep learning methods, Ballé *et al.* [16] was chosen for its state-of-the-art status and lower complexity compared to models such in [17]. To further evaluate the effectiveness of the integrated multi-scale convolution layers in our MS-CNN architecture, we conducted a comparison experiment in which we trained another architecture based on CNN without the multi-scale convolutions that were replaced by sequentially stacked simple convolutions in the architecture based on CNN.

Under high CR, the proposed method's reconstruction was evaluated, with results shown in Table 4. The average MSE, PSNR, CR, and SSs values of the image patches were 6.51, 40.86 dB, 53.03, and 97.95%, respectively. Despite high compression, our method achieved a high PSNR, indicating good retinal image reconstruction quality. Additionally, a space-saving percentage near 100% demonstrates the compression efficiency for systems requiring medical data storage.

Table 4. The performance evaluation of the proposed compression method on retina image patches

Image patches	Measure			
	MSE	PSNR (db)	CR	SSs (%)
Patche 1	2.05	45.02	63.8	98.44
Patche 2	9.3	38.45	56.12	98.22
Patche 3	8.38	38.9	52.13	98.08
Patche 4	10.17	38.06	29.37	96.59
Patche 5	2.67	43.87	63.72	98.43
Mean	**6.51**	**40.86**	**53.03**	**97.95**

We conducted experimental comparisons to assess the proposed method's performance against other compression techniques as detailed in tables below. Tables 5 and 6 demonstrate that our proposed method of compressing medical images is more efficient than the other methods, as evidenced by its achieved maximum PSNR (dB) and MS-SSIM, and the low distortion in MSE. At CR=50, the proposed method outperformed standard methods (JPEG, JPEG 2000) and the deep learning-based approach Ballé *et al.* [16]. Table 5 shows our method achieved superior results, with a 6.89 MSE, 2.32 dB higher PSNR, and 0.8% higher MS-SSIM compared to Ballé *et al.* [16], which itself surpassed JPEG. Compared to JPEG2000, our method showed gains of 2.11 dB in PSNR and 0.7% in MS-SSIM. Additionally, the MS-CNN approach slightly outperformed our method (without MSCB), 1.45 dB PSNR gain, and 0.7% MS-SSIM gain. At CR=80, our method maintained robust performance compared to others. Table 6 reports a 2.6 dB PSNR and 0.79%

MS-SSIM gain over Ballé *et al.* [16], and 3.03 dB PSNR and 1.6% MS-SSIM improvement over JPEG2000. The MS-CNN method also outperformed our method (without MSCB), with 1.89 dB PSNR gain, and 1.09% MS-SSIM gain. This analysis confirms the proposed method's effectiveness in preserving image quality in terms of PSNR and MS-SSIM metrics, especially at high CR (CR=50, CR=80).

Table 5. Performance evaluation of different methods in terms of reconstructed image quality at CR ≈50

Method	Averaged measures		
	MSE	PSNR (dB)	MS-SSIM (%)
Our method	6.89	39.75	97.42
Ballé *et al.* [16]	11.74	37.43	96.63
JPEG2000	11.23	37.64	96.72
JPEG	39.66	32.15	88.27
Our method (without MSCB)	10.08	38.3	96.68

Table 6. Performance evaluation of different methods in terms of reconstructed image quality at CR ≈80

Method	Averaged measures		
	MSE	PSNR (dB)	MS-SSIM (%)
Our method	8.25	38.97	96.75
Ballé *et al.* [16]	15.02	36.37	95.8
JPEG2000	16.72	35.94	95.15
JPEG	218.57	24.77	76
Our method (without MSCB)	12.9	37.08	95.66

The visual quality comparisons at 0.16 bpp are provided in Figure 4. It can be observed that our proposed method attained better subjective quality of reconstructed images in Figures 4(a)-4(b), compared to Ballé *et al.* [16], JPEG200, and JPEG shown in Figure 4(c), Figure 4(d), and Figure 4(e), respectively. In comparison, JPEG marked a low performance with obvious blocking artifacts in the reconstructed retinal images shown in Figure 4(e).

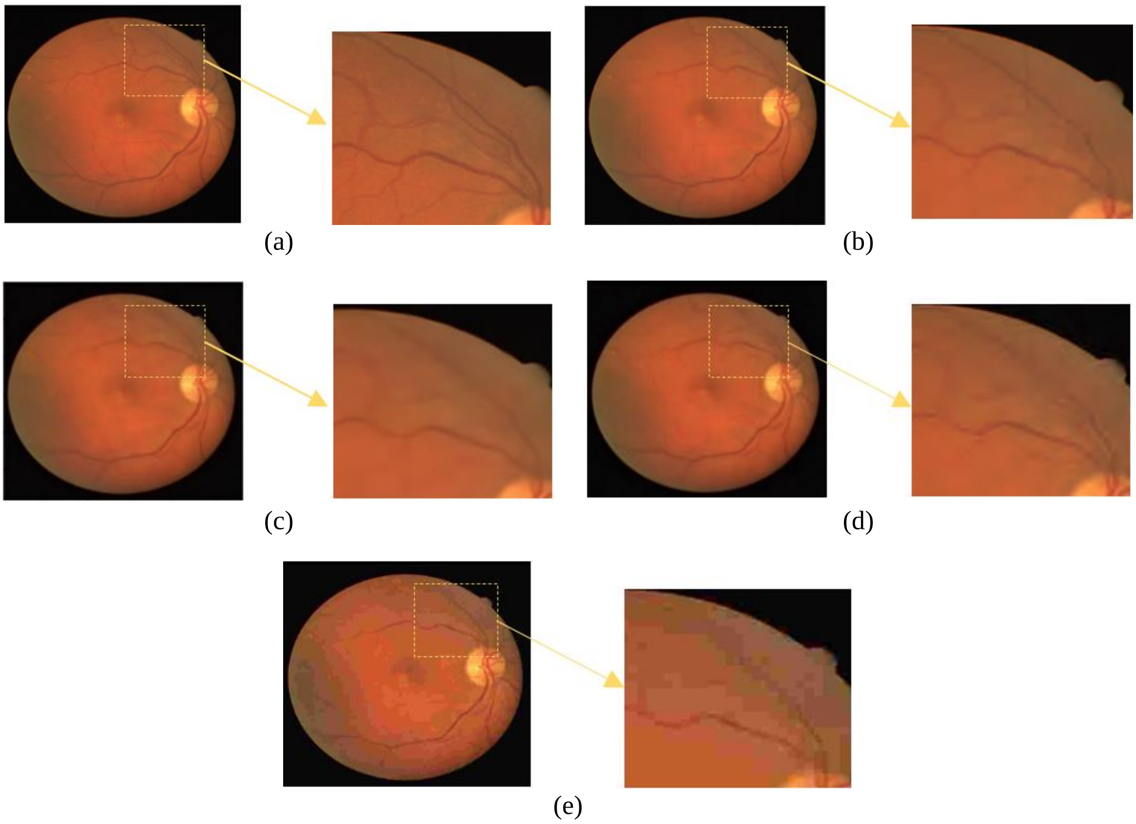


Figure 4. Visual quality comparison between the original and reconstructed images using different methods at 0.16 bpp: (a) original retinal images and their zoomed-in patches, (b) our method, (c) Ballé *et al.* [16], (d) JPEG200, and (e) JPEG

We proposed an efficient compression technique utilizing two MS-CNNs and DWT code. The compression performance was evaluated using various quality metrics. For example, at CR=80 our method achieved the highest PSNR and MS-SSIM values of 38.98 dB and 96.86%, respectively, outperforming even standard compression methods. Particularly, the JPEG2000 which is based on the same CDF 9/7 wavelet transform used in our DWT based codec. Although in our technique deep learning model MS-CNN allowed an enhanced compression performance of DWT codec, it is important to note that our codec is based on a basic compression scheme with a straightforward entropy coding technique, compared to the JPEG2000's sophisticated scheme and entropy coding technique [7], [37]. In contrast, the low performance of JPEG is basically due to the DCT, which is less effective than the wavelet transforms. In comparison to deep learning-based methods, we present in Table 7, a brief and averaged experimental results comparison that includes only learning based compression models. Table 7 compares our method with other learning-based compression models, showing that our method achieved the highest PSNR with the fewest parameters allowing low computational complexity. Although Cheng *et al.* [17] achieved a 0.002 higher MS-SSIM, it required 11.6 million parameters in the architecture. Our method is also 5.2 and 12.1 times faster than Ballé *et al.* [16] and Cheng *et al.* [17], respectively. Thus, it can be said that our proposed method achieves better coding performance and reconstruction quality with low deep learning model complexity and computation time

Table 7. The comparative results of existing learning based methods in terms PSNR, MS-SSIM, number of trainable parameters and computation time at CR=80

Method	PSNR (dB)	MS-SSIM (%)	Numbers of parameters	Execution time (s)
Our method	38.98	96.86	782 694	0.25
Ballé *et al.* [16]	36.38	95.8	2 582 531	1.3
Cheng *et al.* [17]	38.90	97	11 627 916	3.03

The proposed method has demonstrated exceptional performance with promising results, even when compared to complex deep learning-based methods that have a greater number of trainable parameters. The implication of these results is that it offers a superior solution for image data reduction, particularly in scenarios requiring high CR. This advancement could significantly impact various fields that rely on efficient image storage and transmission, such as medical imaging, satellite communications, and digital archiving. The method's ability to maintain good image quality even at high compression levels suggests potential applications in bandwidth-constrained environments or systems with limited storage capacity. Furthermore, this development may lead to improved performance in real-time image processing applications, where data size, quality and computational efficiency are critical factors. In the future, the performance of the proposed method can be enhanced. First, by designing better deep learning architectures and optimization strategies for image compression tasks. Furthermore, the technique's compatibility with conventional codecs, due to its use of a DWT-based codec, opens up possibilities for other conventional codecs to be used in conjunction with it.

4. CONCLUSION

In this paper, we have introduced a retinal image compression method based on MS-CNNs and DWT. To achieve better image quality at a high CR, two MS-CNNs were connected together, the encoding MS-CNN is employed to generate intermediate compact representation, which maintains the structural information of the original image that will be coded by DWT. Next, the MS-CNN enables high-quality reconstruction and retrieval of the original image at the decoding side. The obtained experimental results confirmed the superiority of the proposed compression method based on different performance metrics. The proposed method attained higher CR (CR=80) while maintaining an acceptable retinal image quality with an average PSNR value of 38.98 dB and MS-SSIM of 96.8%. Hence, contribution to minimizing the data size and saving storage and transmission resources while maintaining visual quality of medical images. Furthermore, our proposed technique exhibits computational efficiency, making it applicable in real-time medical image applications.

REFERENCES

[1] R. Vij and S. Arora, "A systematic survey of advances in retinal imaging modalities for Alzheimer's disease diagnosis," *Metabolic Brain Disease*, vol. 37, no. 7, pp. 2213–2243, Oct. 2022, doi: 10.1007/s11011-022-00927-4.

[2] X. He, Y. Deng, L. Fang, and Q. Peng, "Multi-modal retinal image classification with modality-specific attention network," *IEEE Transactions on Medical Imaging*, vol. 40, no. 6, pp. 1591–1602, Jun. 2021, doi: 10.1109/TMI.2021.3059956.

[3] M. Z. Atwany, A. H. Sahyoun, and M. Yaqub, "Deep learning techniques for diabetic retinopathy classification: a survey," *IEEE Access*, vol. 10, pp. 28642–28655, 2022, doi: 10.1109/ACCESS.2022.3157632.
[4] Y. Yuan, L. Zhang, L. Wang, and H. Huang, "Multi-level attention network for retinal vessel segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 312–323, Jan. 2022, doi: 10.1109/JBHI.2021.3089201.
[5] A. Chaabouni, Y. Gaudeau, J. Lambert, J. M. Moureaux, and P. Gallet, "H.264 medical video compression for telemedicine: Aperformance analysis," *Irbm*, vol. 37, no. 1, pp. 40–48, Feb. 2016, doi: 10.1016/j.irbm.2015.09.007.
[6] X. Liu *et al.*, "Medical image compression based on variational autoencoder," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–12, Dec. 2022, doi: 10.1155/2022/7088137.
[7] D. S. Taubman and M. W. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice*. Boston, MA: Springer US, 2002.
[8] S. P. Raja, "Wavelet-based image compression encoding techniques-a complete performance analysis," *International Journal of Image and Graphics*, vol. 20, no. 2, p. 2050008, Apr. 2020, doi: 10.1142/S0219467820500084.
[9] P. Hänsgen, P. E. Undrill, and M. J. Cree, "The application of wavelets to retinal image compression and its effect on automatic microaneurysm analysis," *Computer Methods and Programs in Biomedicine*, vol. 56, no. 1, pp. 1–10, Apr. 1998, doi: 10.1016/S0169-2607(98)00006-6.
[10] R. H. Eikelboom *et al.*, "Methods and limits of digital image compression of retinal images for telemedicine," *Investigative Ophthalmology and Visual Science*, vol. 41, no. 7, pp. 1916–1924, 2000.
[11] S. Krivenko, V. Lukin, O. Krylova, and V. Shutko, "Visually lossless compression of retina images," in *2018 IEEE 38th International Conference on Electronics and Nanotechnology, ELNANO 2018 - Proceedings*, Apr. 2018, pp. 255–260, doi: 10.1109/ELNANO.2018.8477459.
[12] M. R. K. Mookiah, S. Hogg, T. MacGillivray, and E. Trucco, "On the quantitative effects of compression of retinal fundus images on morphometric vascular measurements in VAMPIRE," *Computer methods and programs in biomedicine*, vol. 202, p. 105969, Apr. 2021, doi: 10.1016/j.cmpb.2021.105969.
[13] P. Nazari and H. Pourghassem, "An automated vessel segmentation algorithm in retinal images using 2D Gabor wavelet," in *Iranian Conference on Machine Vision and Image Processing, MVIP*, Sep. 2013, pp. 145–149, doi: 10.1109/IranianMVIP.2013.6779967.
[14] M. Javed, P. Nagabhushan, B. B. Chaudhuri, and S. K. Singh, "Edge based enhancement of retinal images using an efficient JPEG-compressed domain technique," *Journal of Intelligent and Fuzzy Systems*, vol. 36, no. 1, pp. 541–556, Feb. 2019, doi: 10.3233/JIFS-18859.
[15] N. D. Salih, A. Abid, and C. Eswaran, "Efficient retinal image compression based on modified huffman algorithm," *International Journal of Engineering Research and Technology*, vol. 12, no. 7, pp. 942–948, 2019.
[16] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, "Variational image compression with a scale hyperprior," *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018.
[17] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "Deep residual learning for image compression," *Arxiv*, 2019, [Online]. Available: http://arxiv.org/abs/1906.09731.
[18] A. Kar, S. P. K. Karri, N. Ghosh, R. Sethuraman, and D. Sheet, "Fully convolutional model for variable bit length and lossy high density compression of mammograms," *Arxiv*, 2018, [Online]. Available: http://arxiv.org/abs/1805.06909.
[19] A. S. Sushmit, S. U. Zaman, A. I. Humayun, T. Hasan, and M. I. H. Bhuiyan, "X-Ray image compression using convolutional recurrent neural networks," in *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, May 2019, pp. 1–4, doi: 10.1109/BHI.2019.8834656.
[20] P. Guo, D. Li, and X. Li, "Deep OCT image compression with convolutional neural networks," *Biomedical Optics Express*, vol. 11, no. 7, p. 3543, Jul. 2020, doi: 10.1364/BOE.392882.
[21] F. Liu, M. Hernandez-Cabronero, V. Sanchez, M. W. Marcellin, and A. Bilgin, "The current role of image compression standards in medical imaging," *Information (Switzerland)*, vol. 8, no. 4, p. 131, Oct. 2017, doi: 10.3390/info8040131.
[22] J. Guo and H. Chao, "Building dual-domain representations for compression artifacts reduction," 2016, pp. 628–644.
[23] M. Beladgham, Y. Habchi, M. Ben aissa, and A. Taleb-Ahmed, "Medical video compression using bandelet based on lifting scheme and SPIHT coding: In search of high visual quality," *Informatics in Medicine Unlocked*, vol. 17, p. 100244, 2019, doi: 10.1016/j.imu.2019.100244.
[24] O. H. Nagoor, J. Whittle, J. Deng, B. Mora, and M. W. Jones, "MedZIP: 3D medical images lossless compressor using recurrent neural network (LSTM)," in *Proceedings - International Conference on Pattern Recognition*, Jan. 2020, pp. 2874–2881, doi: 10.1109/ICPR48806.2021.9413341.
[25] L. Theis, W. Shi, A. Cunningham, and F. Huszár, "Lossy image compression with compressive autoencoders," *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 2017.
[26] Y. Wang, G. Wang, C. Chen, and Z. Pan, "Multi-scale dilated convolution of convolutional neural network for image denoising," *Multimedia Tools and Applications*, vol. 78, no. 14, pp. 19945–19960, Jul. 2019, doi: 10.1007/s11042-019-7377-y.
[27] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 205–220, Apr. 1992, doi: 10.1109/83.136597.
[28] M. Beladgham, A. Bessaid, M. Abdelmounaim, and T. Abdelmalik, "Improving quality of medical image compression using biorthogonal CDF wavelet based on lifting scheme and SPIHT coding," *Serbian Journal of Electrical Engineering*, vol. 8, no. 2, pp. 163–179, 2011, doi: 10.2298/SJEE1102163B.
[29] H. Schwarz *et al.*, "Quantization and entropy coding in the versatile video coding (VVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3891–3906, Oct. 2021, doi: 10.1109/TCSVT.2021.3072202.
[30] Q. Hao *et al.*, "High signal-to-noise ratio reconstruction of low bit-depth optical coherence tomography using deep learning," *Journal of Biomedical Optics*, vol. 25, no. 12, Nov. 2020, doi: 10.1117/1.jbo.25.12.123702.
[31] J. Cai, Z. Cao, and L. Zhang, "Learning a single tucker decomposition network for lossy image compression with multiple bits-per-pixel rates," *IEEE Transactions on Image Processing*, vol. 29, pp. 3612–3625, 2020, doi: 10.1109/TIP.2020.2963956.
[32] J. Wang, T. Liu, Q. Liu, X. He, H. Luo, and W. He, "Compression ratio modeling and estimation across error bounds for lossy compression," *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 7, pp. 1621–1635, Jul. 2020, doi: 10.1109/TPDS.2019.2938503.
[33] B. Sujitha, V. S. Parvathy, E. L. Lydia, P. Rani, Z. Polkowski, and K. Shankar, "Optimal deep learning based image compression technique for data transmission on industrial Internet of things applications," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 7, Jul. 2021, doi: 10.1002/ett.3976.
[34] J. L. Ba and D. P. Kingma, "Adam: a method for stochastic optimization," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15, 2015.

[35] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, Apr. 2004, doi: 10.1109/TMI.2004.825627.
[36] Z. Zhang *et al.*, "ORIGA-light : an online retinal fundus image database for glaucoma analysis and research," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10*, Aug. 2010, pp. 3065–3068, doi: 10.1109/IEMBS.2010.5626137.
[37] O. Watanabe, A. Uchida, T. Fukuhara, and H. Kiya, "An encryption-then-compression system for JPEG 2000 standard," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Apr. 2015, vol. 2015-August, pp. 1226–1230, doi: 10.1109/ICASSP.2015.7178165.

BIOGRAPHIES OF AUTHORS

Dalila Chikhaoui is a Ph.D. student in Telecommunications Systems at University TAHRI Mohamed of Bechar. Graduated with a master's degree diploma in Telecommunications systems from the University of Bechar, Algeria. Her current research interest includes image processing, medical image compression, deep learning and machine learning. She can be contacted at email: chikhaoui.dalila@univ-bechar.dz.

Mohammed Beladgham was born in Tlemcen, Algeria. He received his master's degree in signals and systems, followed by a Ph.D. in Electronics from the University of Tlemcen in Algeria in 2012. He is currently a professor University TAHRI Mohamed of Bechar, Algeria. His research interests include video and image processing, biomedical imaging, wavelets transforms and optimal encoders, and deep learning. He can be contacted at email: beladgham.mohammed@univ-bechar.dz.

Mohamed Benaissa was born in France. He obtained his engineering degree in electronics, his Magister degree in signals and systems from the University of Tlemcen and the PhD degree in electrical engineering from the University of Bechar, Algeria. He is currently a professor at the Faculty of Technology in Tlemcen, Algeria. His research interests include technology and applications of embedded electronics, telecommunications applications, and signal and image processing. He can be contacted at email: moh.benaissa@gmail.com.

Abdelmalik Taleb-Ahmed was born in Roubaix, France, in 1962. He received his Postgraduate degree and Ph.D. in Electronics and Microwaves from the University of Science and Technology of Lille, France, in 1989 and 1992, respectively. He is currently a Professor at the Polytechnic University of Hauts-de-France, and does his research at IEMN UMR CNRS 8520. His research interests include signal and image processing. He can be contacted at email: abdelmalik.taleb-ahmed@uphf.fr.