ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

LEARNED DICOM IMAGE COMPRESSION VIA SUPER-RESOLUTION WITH DILATED RESIDUAL BLOCKS AND PIXEL ATTENTION

OBADA OTHMAN AGHA[1], YAHIA FAREED[2], LOUAY CHACHATI[3]

[1] PhD Candidate, Department of Communication Engineering, Faculty of Electrical & Electronic Engineering, University of Aleppo, Aleppo, Syria;
[2] Professor, Department of Communication Engineering, Faculty of Electrical & Electronic Engineering, University of Aleppo, Aleppo, Syria;
[3] Professor, Department of Electronic Engineering, Faculty of Electrical & Electronic Engineering, University of Aleppo, Aleppo, Syria.
E-mail: [1]Obada_othman_agha@alepuniv.edu.sy, [2]Prof.yahia.fareed@alepuniv.edu.sy, [3]louay.chachati@alepuniv.edu.sy

ABSTRACT

The digital imaging and communications in medicine (DICOM) standard leads global efforts to advance medical imaging. With the growing interest in this field, especially given the rapid developments in telemedicine applications, achieving efficient compression without losing diagnostic accuracy is critical. Therefore, there is an imperative need to employ an advanced hybrid deep-learning model, specifically designed for these medical images, to outperform current methods. The proposed method tests the development of a non-autoregressive model for parallel pixel prediction, avoiding sequential processing. This scheme uses dilated residual blocks (DRBs) to capture long-range dependencies using fewer blocks. It combines depthwise separable convolution (DSC) layers with the pixel attention mechanism to reduce computational complexity while preserving the diagnostic details in the reconstruction task. The method also leverages a discretized mean and scale Gaussian distribution mixture model to achieve efficiency across various DICOM image types. The results demonstrate that the method achieves higher compression ratios, and improves the bpsp metric by 19.26%, 22.89%, and 23.94%, compared to the best competing methods, for each of magnetic resonance imaging (MRI), computed radiography (CR), and computed tomography (CT) images, respectively. Compared to the leading learning-based methods, the system reduces the mean compression time by [16.96 - 19.02] %. At the same time, the high values of PSNR and SSIM metrics demonstrate the ability to ensure high quality. This approach balances compression efficiency, speed, and diagnostic reliability, enhancing DICOM image processing for telemedicine.

Keywords: *DICOM images, Deep learning, Super-Resolution, Images Compression, DSC layers.*

1. INTRODUCTION

Medical imaging is an essential part of telemedicine, especially using the digital imaging and communications in medicine (DICOM) standard, which is widely regarded as the most used and effective standard worldwide. This standard is the foundation through which accurate diagnoses are made, treatment protocols are developed, and patient monitoring is performed [1]. Therefore, achieving effective compression of high-quality DICOM images is essential when processing and exchanging these images between healthcare centers. The importance of using deep neural networks (DNNs) in the context of image compression has recently emerged, divided concerning structure into four categories: innovative, autoencoders, recurrent neural network (RNN) based, and convolutional neural network (CNN) based, where the last structure, in particular, has high-efficiency advantages that have drawn the focus of researchers on development [2]. The usual procedure followed in image compression literature is to assume that the symbol streams are independent and identically distributed (IID) and then revert to parametric models to fit the distribution of such symbols.

Recently, several end-to-end learned schemes using DNNs for image compression have been introduced assuming all symbols are IID. However, due to the insufficient accuracy in establishing the entropy models, these recent methods still have limited performances [3].

DICOM images generally don't conform to the IID assumption. This is primarily due to the inherent correlations among pixels that arise from the anatomical features they depict. Furthermore, DICOM images are characterized by intricate distributions that encompass a variety of structures and textures [1]. By neglecting interdependencies among pixels, some relevant details of primary importance when making an exact diagnosis in medical treatments may be missed. Therefore, the compression algorithms based on this assumption may yield inaccurate representations, which reduce the compression efficiency and introduce artifacts in the reconstructed images. As a result, we evaluate a learnable super-resolution near-lossless compression method, including three parallel scales, that employs hybrid deep learning (DL) models in the feature extraction and prediction tasks while reducing the computational complexity, where a varied dataset comprising magnetic resonance imaging (MRI), computed radiography (CR), and computed tomography (CT) images is utilized.

While previous studies have focused on developing and testing non-learning-based compression algorithms to improve some criteria, such as compression ratio and image quality testing criteria, we aim to build a fast non-autoregressive learning-based compression algorithm to compress various types of medical DICOM images, including their varying bit depth and image dimensions, by achieving a balance that ensures the improvement of the all different criteria simultaneously.

Related studies in the following section can be classified into works that have contributed to the development of learning-based natural image compression algorithms and efforts that have sought to improve DICOM image compression using non-learned methods.

2. RELATED WORK:

Effective efforts have been made in this field, a parallel hierarchical probabilistic framework known as L3C was introduced in [4], which employed a non-autoregressive probability model to achieve lossless compression for natural images, based on an enhanced deep super-resolution (EDSR) model for feature extraction and prediction coupled with an efficient sub-pixel convolutional neural network (ESPCN) model for reconstruction. The work recommended further enhancements by handling specialized image types. SReC [5] introduced an autoregressive super-resolution network to achieve a lossless compression method for natural images. The proposed scheme in [6] introduced a three-tier hierarchical structure derived from L3C model, where a self-supervised clustering module was integrated to proficiently detect and characterize the long-term dependencies inherent in the image. To address the redundancies found within the image pixels, the research [7] proposed an autoregressive model for raw images, utilizing an end-to-end architecture in conjunction with channel-conditioning models. The results of the deep learning techniques comparison conducted by the research [8] showed that both enhanced deep residual neural network (EDRN) and super-resolution generative adversarial network (SRGAN) outperformed the ESPCN model in the field of improving micro-CT images according to several evaluation criteria, where all these models use 16 residual blocks within their structure to achieve a wide receptive field. To improve the spatial resolution in remote sensing image reconstruction, the work [9] employed each of ESPCN and enhanced SRGAN (ESRGAN), an improved version of SRGAN, where the batch normalization layers are removed from residual blocks.

On the other hand, researchers continue their efforts to develop an effective mechanism for compressing DICOM images using methods that don't rely on deep learning. The work [10] developed a mechanism to reduce the redundant information in DICOM images through wavelet compression, achieving an overall compression ratio of 60%. In [11], the researchers tested the results of applying each of the discrete wavelet transform (DWT), discrete cosine transform (DCT), fractal compression algorithm (FCA), and vector quantization algorithm (VQA) only to the non-regions of interest (NROI) in DICOM images, where the DWT algorithm achieved the best results. Emphasizing the need for future employment of artificial intelligence algorithms in this field. The research [12] proposed a transmission system for exchanging DICOM images depending on Golomb-Rice coding and Run-Length Encoding (RLE). The work [13] tested a new compression methodology, attempting to leverage spatial features and pixel concentration patterns for adaptive processing to improve the storage of monochrome DICOM medical files. It demonstrated a 37% increase in compression ratios compared to JP2, which is considered the primary DICOM industry standard. The authors in [14] presented a lossless compression method for medical images using a hybrid of Huffman coding, linear predictive coding, and discrete wavelet. The work [15] introduced a CNN-based compression approach for general medical images and showed relative improvement compared to JPEG. It depended on a systematic iteration of design space exploration (DSE) to reduce the computational complexity.

However, the previous compression methods are limited to testing only a few criteria, such as compression ratio or bpsp. At the same time, other research has focused on testing image quality criteria without comparing compression time or simplifying the structure without affecting the quality of the reconstructed image. Furthermore, reference studies have avoided developing a learning-based compression algorithm for DICOM images due to its high computational complexity. This is what we seek to achieve in this research by balancing the various testing criteria and simplifying the proposed algorithm structure to the simplest possible level.

Therefore, the review of recent research indicates an urgent need to develop a learnable compression algorithm specifically designed for the different types of medical images according to the DICOM standard, one of the most important and widely used standards worldwide, where the modern compression methods struggle to balance achieving a high compression ratio, fast execution, and preserving sensitive image data to prevent misdiagnosis, which is achieved by ensuring high image quality test standards.

3. METHOD:

3.1 Features Extraction:

One of the most important blocks of modern deep learning architecture is the residual block (RB). It avoids problems like vanishing gradients since it allows the network to learn the residual functions indicating the discrepancies between the input and output. The residual architecture was enhanced by eliminating batch normalization layers, resulting in superior performance compared to SRResNet architecture in the context of image super-resolution [16]. Alongside these developments, residual CNN models have been effectively utilized across several computer vision applications.

The work [17] introduced a concatenated residual block (CRB) within the encoding/decoding framework, which involves the sequential linkage of several residual blocks, augmented by supplementary shortcut connections to enhance the network's learning capacity. To perform the feature extraction and prediction tasks, the research [4] used the EDSR model based on residual blocks, utilizing the suggestion to delete the batch normalization layer.

3.1.1 Receptive field increasing:

Receptive field size is a very important factor while building any DNN architecture since, as this size increases, the input data will have more contextual information. This needed increase may be accomplished in several ways. One might opt for a deeper network with more convolution layers where the receptive field grows linearly. Inevitably, this closely relates to the size of the convolution kernel applied in the network. Although the 3×3 kernel size has been widely regarded as the most efficient filter size in CNNs [18], it contributes only a factor of two in the increase of the receptive field per additional layer. Larger kernel sizes would thus be needed, but adding this can have complications because it would increase the number of parameters involved and needlessly overcomplicate the model [19].

Because they can learn complex mappings better, neural network architectures that are bigger and more complex tend to do better at image reconstruction tasks [20, 21]. Nevertheless, contemporary deep networks, such as EDSR, which are used in L3C and other recent studies [19], may not be optimal, since the reliance on numerous conventional convolution layers not only leads to significant computational expenses but also increases the risk of overfitting, while many medical applications related to the DICOM standard require tasks to be performed in real-time. Sub-sampling techniques and the addition of pooling layers achieve this. However, this decreases contextual information, losing spatial accuracy, which makes this method suboptimal for image restoration tasks. Since we want to obtain an output of the same size as the original input, the following layers will use transpose convolution layers for further processing [22].

Dilated convolution (DConv), also called Atrous convolution, refers to a class of convolution layers developed in deep learning that insert gaps between kernel weights. Most especially, for example, a 3x3 filter kernel, when set to a dilation rate of 2, results in a receptive field similar to a kernel of size 5x5, but it uses only 9 parameters. This approach facilitates an expanded field of view while maintaining the same computational cost, eliminating the necessity for multiple convolutional layers or an increase in filter size [23, 24]. Therefore, the dilated convolution layers maintain the size of feature maps and work efficiently in applications that are interested in integrating contextual information according to a larger receptive field at a lower cost [25]. These layers can also be used to remove downsampling layers, thus developing the fields of human motion prediction and semantic segmentation, as in [26].

3.1.2 Dilated residual blocks (DRBs):

When the previous concepts are integrated, a dilated residual block (DRB) employs dilated convolutions within the framework of a residual block. This mechanism enables the network to effectively learn long-range dependencies while preserving the advantages of residual connections, which expands the receptive field, allowing the network to gather contextual information from larger sections of the images. This capability is essential for the compression of images using super-resolution (SR) techniques, where capturing fine details and interrelationships across extensive areas is vital, as in the case of DICOM image compression applications. Moreover, dilated convolutions may provide effective mitigation of blurriness, which are artifacts often introduced by mainstream compression methods. Additionally, the ability to learn efficiently with a reduced number of parameters may facilitate higher compression ratios while preserving superior reconstruction quality.

In terms of architecture, Figure 1a depicts the original residual block utilized in the L3C scheme, which depends on the EDSR model [4]. Figure 1b illustrates the suggested architecture of the fixed-rate dilated residual blocks used in [18], which presented a method for restoring infrared images that reduces the number of RBs from 16, as employed in EDSR and SRResNet architectures [19], to 8. Where PReLU denotes parametric ReLU and D represents the dilation rate, considering that k, C_f, and s denote the parameters of kernel size, channels, and stride, respectively. This increased number of the used layers led to an improvement in the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) evaluation criteria, but it caused the number of the total parameters to be similar to the comparison models [18].

The first part of our proposed model framework aims to reduce the number of used RBs from 16 to 8, utilizing edited DRBs, as shown in Figure 1c. Thus, this approach is used to decrease the computational cost while also increasing the analysis area without losing any information critical to diagnosis, as we apply progressively larger dilation rates (1, 2, 4, 8, and 16) to extract features at different visual scales. Compared to the other activation functions, we choose Leaky ReLU for many reasons. Its capability of alleviating the dying ReLU problem and enhancing the gradient flow can preserve important information and capture intricate feature interactions to achieve more accurate reconstructions [27], especially in the scope of complex data types such as DICOM images.

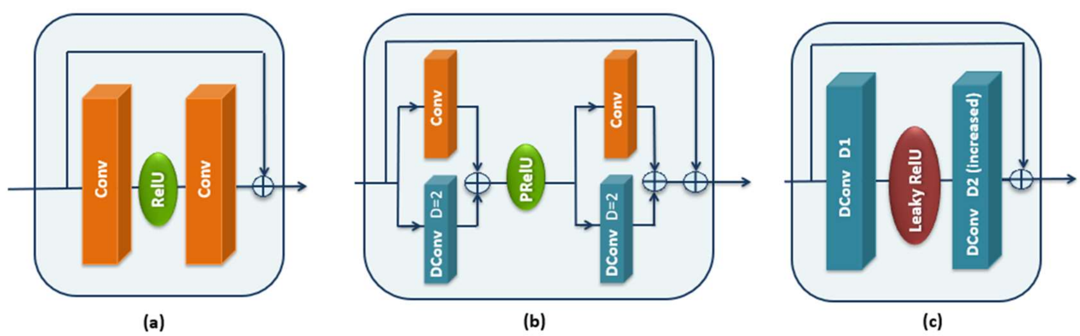

Figure 1. The Architectures Of The Different RB And DRB: (A) Original RB Architecture; (B) DRB Architecture With Fixed Dilation Rate; (C) Our Used DRB Architecture Using Increased Dilation Rate.

3.2 Reconstruction and Upsampling:

SR mechanisms have been a major and ongoing challenge in the field of image-processing, where the main goal is to obtain a high-resolution (HR) image by merging its low-resolution (LR) version. Researchers in academia use similar phrases for SR, such as image upsampling and image upscaling [28]. Advanced deep-learning models with millions of parameters have contributed to many such image processing applications, including SR. These more complicated models indeed tend to be better at high-resolution image reconstructions and perform better concerning objective Image Quality Assessment (IQA) measures. However, such a complex model is very complicated to deploy for real-time applications [29]. Most of the previous network designs for super-resolution have their limits that inhibit their general performance. Generally, these networks take an upsampled coarse estimate obtained by an interpolation-based method, where we refer to such a strategy as the early upsampling scheme. It is computationally expensive to process such high-dimensional input, and multiply-accumulate operations (MACs) increase drastically when the resolution of the input image increases. This is because most of the operations are actually operated on high-dimensional feature maps [19].

The idea was then further refined, allowing for upsampling to be a learnable part of the process, which is normally referred to as late upsampling schemes, where all the computations are done on low-resolution feature maps. It relied on the transposed convolution layers, also called deconvolution layers, as used in the Fast Super-Resolution Convolutional Neural Network (FSRCNN) model [19], where the number and size of filters have been cut down from the super-resolution CNN (SRCNN) model. However, it has a major problem typified by checkerboarding artifacts [28].

This was addressed by incorporating a sub-pixel convolution layer, which was proposed based on an ESPCN model. The first implementation of the model was done with the ReLU activation function. It has been modified further to accommodate tanh as the activation function instead [21]. The ESPCN has also been widely adopted in various super-resolution methodologies, like L3C [4], SSNet, SSNet-M [30], and LWSR8 [31] models, which are considered to be pivotal in the domain of super-resolution image reconstruction and have recently demonstrated remarkable success in various applications. However, these models suffer from some limitations, which this research aims to overcome. ESPCN, SSNet-M, and SSNet use the tanh activation function, which suffers from the vanishing gradient and effectively hinders the training process [28]. Besides, the traditional convolution layers are used in the architectures of ESPCN, SRCNN, and FSRCNN, which raises the computational complexity of the networks [29]. For instance, the three-layer ESPCN requires about 18.46 thousand parameters in its second layer only, which is over 75% of the model's total number of parameters. In turn, the U-Net architecture was used in [31]. However, the small depth wasn't allowed to model its ability in each of the feature extraction and image reconstruction stages effectively.

Based on the above, we will use the Leaky ReLU activation function and replace the traditional convolution layers with the depthwise separable convolutions (DSC) layers in the upsampling phase.

3.2.1 Depthwise separable convolutions (DSC) layers:

DSCs efficiently capture spatial features while jointly learning channel dependencies [30]. This is particularly important for DICOM images, where local features such as edges and textures, as well as global contextual information such as anatomical structures, must be preserved to provide a precise reconstruction. DSCs enhance the model's generalization capability and drastically reduce the number of parameters [32]. This reduction is especially vital in the field of medical imaging since often processing power may be limited, and efficiency becomes key. The streamlined architecture of DSCs allows them to enable faster inference times, which becomes a significant criterion for all those applications that require rapid analysis of the images in DICOM format [1]. In the standard convolution layer, the cost of operations is calculated using the Eq. (1), assuming that C_{in} denotes the input channels and C_{out} signifies the output channels [30].

$$\text{Std conv}\ (C_{in}, C_{out}, K) = H \times W \times C_{in} \times C_{out} \times K^2 \quad (1)$$

Where k^2 denotes kernel size, and H*W denotes width and height of the feature map. DSCs have two layers; the first layer is a depthwise convolution in which every input layer has one filter, while the second layer is a pointwise convolution using the 1*1 convolution operation to merge the output of the first layer, as in Eqs. (2) and (3) [30, 32]:

$$\text{DSC depthwise}\ (C_{in}, C_{out}, K) = H \times W \times C_{in} \times K^2 \quad (2)$$

$$\text{DSC pointwise}\ (C_{in}, C_{out}, K) = H \times W \times C_{in} \times C_{out} \quad (3)$$

To realize the effect of applying the DSC layer, let's assume that the DICOM image size is 512*512. Since this is a greyscale image, the input channels C_{in} = 1 while the output channels C_{out} = 32 at a kernel size K = 3. While in the case of a standard convolution, the total operations sum to 75.497.728, for DSC, this is much lower, where the depthwise operations are 2.359.296 and the pointwise operations are 8.388.608. Thus, overall operations for DSC are 10.747.904. Therefore, using DSC instead of standard convolution reduces the number of operations by about 85.75%.

3.2.2 Pixel attention (PA):

Recently, the process of integrating attention mechanisms into the models has effectively contributed to their improved performances. The work [33] introduced a model for improving network representational capacity through dynamic feature recalibration in the channel-spatial dimensions. This model could be applied for image compression since it can be considered the first proposal that takes advantage of channel attention mechanisms to reach strong contrast features. The researchers in [34] developed a model that employs PA to extract the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 3. Comparison Of Upsampling Architectures: (A) Upsampling Using ESPCN Model; (B) Upsampling Using PA And DSC Layers.

informative features, aiming to improve the accuracy of tomato plant disease classification. The results show that PA's performance is better than that of the other alternative mechanisms. While other attention methods rely on building complex attention modules to improve their effectiveness, PA has the potential to enable efficient pixel-wise attention learning with lower computational costs [35].

In the pixel attention mechanism, focusing on the important pixels guided by contextual information improves the quality of the image that is reconstructed. This is particularly useful in medical imaging, where some regions may require more attention because of their clinical significance. The ability to assign higher weights to certain pixels than others allows the model to emphasize the key regions in such a way that diagnostic features in those parts are preserved during upsampling.

This structure returns a features matrix of size (C*H*W) and calculates attention coefficients for every pixel in the feature map. However, it is dependent on computation from the contextual feature extraction. Each output corresponds to the full receptive field with respect to the input [36]. It then normalizes the output from the convolution layer in the interval of [0, 1] using a sigmoid activation function. Then, the weights, which we denote by F(X), are formed and used in an element-wise multiplication with the input X to get a weighted output $\tilde{X}$, as shown in Figure 2, where the salient features are emphasized. Compared to the more traditional methods of layered techniques, the direct integration of features involves fewer parameters in an overall simpler architecture [35].

Figure 2. PA architecture.

The benefits of PA were utilized towards the improvement of the model performance in [36] for super-resolution natural image reconstruction. For this purpose, a two-channel PA mechanism based on the ESPCN model was used with standard convolution layers and a tanh activation function. [37] proposed a fusion network that employs PA for combining semantic with shallow features, enhancing small object detectability against complex backgrounds, with multiple feature selections allowed. The upsampling part of our scheme will exploit the integration between DSC and pixel attention for better artifact reduction that may emanate from the process. This is of prime importance in the compression operation since any form of distortion introduced will surely affect the integrity of the DICOM image. Besides, this is a flexible solution that can be adapted with ease for the different kinds of medical images, considering the large variance of the contents and qualities of the DICOM images. Compared to the original ESPCN model shown in Figure 3a, our used architecture comprises a DSC layer of size (C, C, K), a pixel attention mechanism, and another DSC of size (C, D^2, K), where C is the propagated channels, D is the desired upscale factor (the used value here is 2), and K = 3. Further, it is followed by a pixel shuffle operation to complete the upsampling block, as demonstrated in Figure 3b. In trying to explain how this sub-pixel convolution layer works within the upsampling framework, let's take the case of an input DICOM image of size 512 × 512 × 1; thereafter the layer changes that size from the previously received output of the DSC layer, namely 512 × 512 × D^2, into ((D × 512) × (D × 512) × 1).

3.2.3 Eliminate the use of the Atrous spatial pyramid pooling (ASPP) model:

Atrous spatial pyramid pooling (ASPP) model works through different dilation rates of dilated convolutions, allowing the finding of the receptive field while enabling the model to see the image at multiple detail levels. The previous output is then combined via a final convolution layer as demonstrated in Figure 4. In this way, this strategy effectively combines multiscale information to yield a more elaborated and detailed upsampling image description [38].

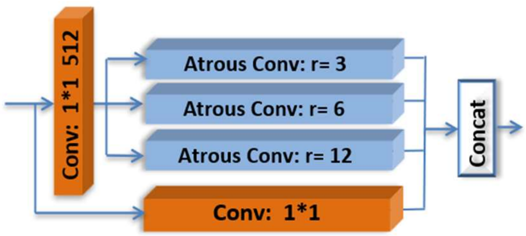

Figure 4. ASPP Model Architecture.

In our scheme, the implementation of dilated convolutions with an increased rate during the feature extraction phase allows the acquisition of contextual information and the learning of multiscale features without the need of downsampling the input image. This approach is particularly important in medical DICOM imaging, where spatial detail preservation is essential for accurate reconstruction. Similarly, applying an increase rate in dilated convolutions during the initial reconstruction phase prior to the upsampling will facilitate the storage of high-resolution data and minimize the possibility of artifacts appearing during the upsampling, making it easier and more effective. Thus, unlike the approach in [4], Upsampling can be performed directly without necessarily incorporating ASPP afterward, freeing up more computation and reducing memory consumption to the barest minimum.

3.3 Architecture:

Figure 5 shows the detailed architecture of the suggested scheme for a single scale, while Figure 6 shows the overall architecture across the three needed scales.

3.4 Mixture model:

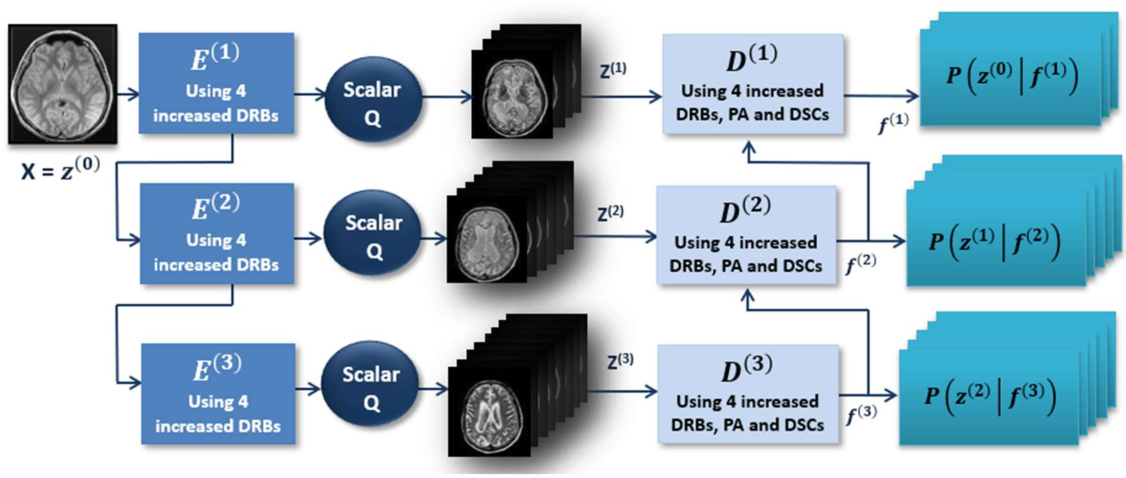

Figure 6. The Overall Architecture Using Three Scales.

The ASPP framework can be integrated at the tail of the reconstruction process to capture a better image context, similar to what has been suggested in [4] as a post-processing procedure after the ESPCN model. The dense ASPP model, which is an edited structure of ASPP, was employed to improve the dense prediction process for image compression in [38]. While Research [39] utilized the ASPP model to improve the glass bottle defect classification task to ensure product quality assurance.

In our scheme, we used adaptive arithmetic coding (AAC) and scalar quantization similar to [4, 6]. As for the mixture model, prior investigations have exhibited diversity in its selection, with numerous studies evaluating various parameter distribution models [40]. For instance, the conventional PixelCNN employs a full 256-way softmax, resulting in significant memory consumption and impracticality for larger images. In contrast, PixelCNN++ introduced distinct logistic mixture probabilities to facilitate expedited training.

Similarly, L3C adopted the logistic mixture model in alignment with the approach of PixelCNN++ [4].

While some previous conventional coding techniques used the Cauchy distribution to model the coefficients, the subsequent test results in [41] showed poor performance. The work [41] tested a mean and scale Gaussian distribution, which showed superior empirical performance compared to the other methods, including zero-mean scalar Gaussian and Laplace distribution models. In addition, the Gaussian mixture model has demonstrated significant efficacy in medical image analysis characterized by indistinct boundaries, as evidenced by the segmentation algorithm introduced in [42], which relies on the probability distributions of both the object and the background present in the image. In accordance with [4], we employ a non-autoregressive predictor $D^{(s)}$ to characterize the joint distribution of the image x alongside the auxiliary hierarchical feature representation after quantization p(x, $z^{(1)}$, ..., $z^{(s)}$) for the scales (s > 0), assuming that $z^{(0)}$ = x. The prediction of p is derived from the features $f^{(s)}$ associated with the preceding scale. Let c represent the channel, while u and v indicate the spatial coordinates. For every scale, we postulate that the components of $z_{cuv}^{(s)}$ are independent concerning u and v, given $f^{(s+1)}$. We define it using Eq. (4) [4].

$$P\left(z^{(s)} \mid f^{(s+1)}\right) = \prod_{c,u,v} P_m \left(z_{cuv}^{(s)} \mid f^{(s+1)}\right) \qquad (4)$$

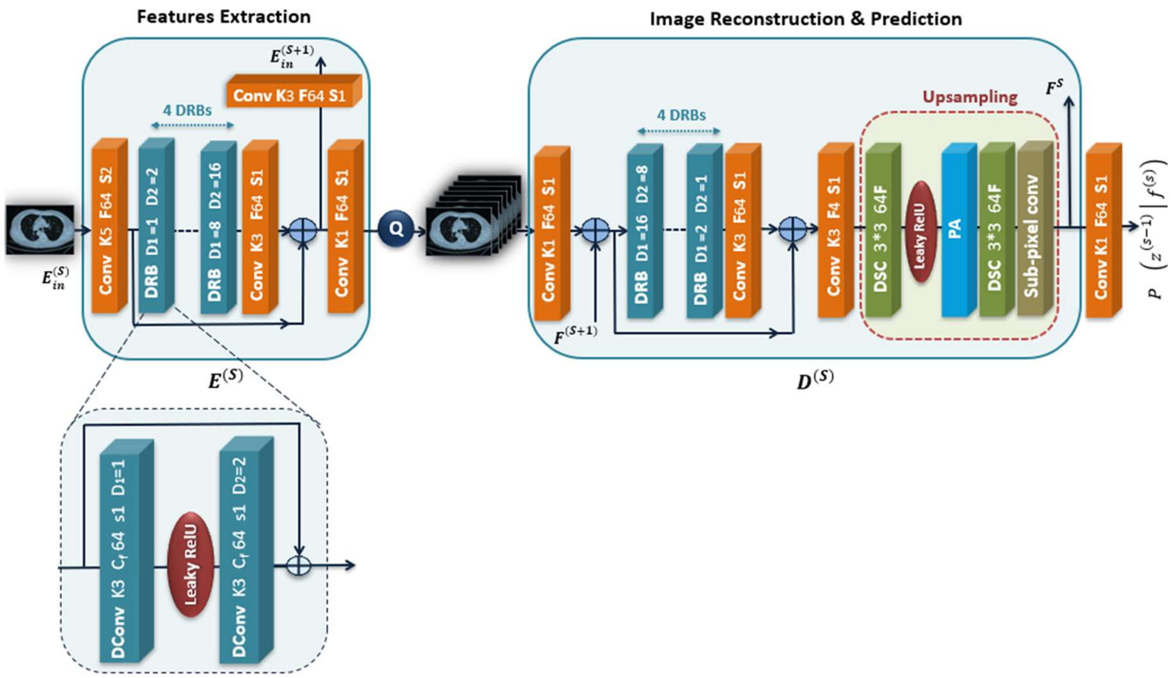

Figure 5. The Proposed Scheme For A Single Scale.

We model the conditional distributions p($z^{(s)}$ | $z^{(s+1)}$, ..., $z^{(s)}$) using the discretized mean and scale Gaussian distribution mixture introduced in the work [43]. The multivariate Gaussian model was utilized in our work instead of the logistic model implemented in L3C and PixelCNN++ as it demonstrated enhanced performance results [41, 43], since it can effectively capture the underlying distribution of pixel values, even in complex medical images with varying textures and structures. The mixture is indicated by P_m and defined as in Eq. (5):

$$P_m\left(z_{cuv}^{(s)} \mid f^{(s+1)}\right) = \sum_{k=1}^{k} \omega_{cuv}^{k} \, N\left(x_{cuv} \mid \mu_{cuv}^{k}, \sigma_{cuv}^{2(k)}\right) \qquad (5)$$

Considering that w, μ, and σ represent weights, mean, and variance parameters, respectively, while N represents the Gaussian density function. In contrast to [41, 43], we opted not to employ a uniform noise distribution to approximate quantization during the training phase. Although that can enhance training stability in certain contexts, it is unsuitable to employ it alongside the used adaptive arithmetic coding because of the associated memory complexity [4]. Additionally, the risks of introducing artifacts, degradation, and adverse clinical implications surpass its advantages in the medical imaging field. Therefore, it is essential to prioritize the maintenance of image quality and clinical precision when handling DICOM images. Where the features extractors $E^{(s)}$ define the representations $z^{(s)} = F^{(s)}$(x), and p(x, $z^{(1)}$, ..., $z^{(s)}$) represents an outcome derived from the discretized Gaussian mixture model with parameters established using $f^{(s)}$, which are calculated based on the predictor $D^{(s)}$, the loss of training N samples can be defined as the following Eq. (6):

$$L\left(E^{(1)},\ldots,E^{(S)},D^{(1)},\ldots,D^{(S)}\right) = -\sum_{i=1}^{N} log\, p\left(x_i \mid F_i^{(1)},\ldots,F_i^{(S)}\right) + \sum_{S=1}^{S} log\, p\left(F_i^{(s)} \mid F_i^{(s+1)},\ldots,F_i^{(S)}\right) \quad (6)$$

4. EXPERIMENTS:

To understand the difference in the mechanisms of applying previous research, Table 1 compares the methods in terms of the dataset, evaluation measures, and implementation environment. These methods don't rely on DL techniques, except for research [15], which was trained on general medical images and isn't specific to the DICOM standard.

The implementation in our research was done using Python 3.11 and the models were built via PyTorch 1.11.0, taking advantage of Pydicom 2.1.1 library to extract the pixel matrix of the DICOM file. The experimental was done using Google-Colab (RAM: 13G, GPU: Nvidia-K80). This study utilized three open-source datasets with different dimensions and types, the first is siim_dicom_images dataset [44], which consists of 12089 computed radiography (CR) images, the second is Brain MRI dataset [45], which consists of 21000 MRI images, and the third is from The Cancer Imaging Archive (TCIA) [46], which includes 3954 CT scans. The datasets were split into 90% for training and 10% for testing. A padding process with zero values was utilized to handle the dimensions if needed. The patch size is 128*128 and the used optimizer is Adam.

5. RESULTS AND DISCUSSION:

We tested compression ratio, compression time, and bits per sub-pixel (bpsp) metrics on GPU Nvidia-K80 using 100 DICOM images each of 1024*1024 CR [44], 512*512 MRI [45], and 512*512 CT [46] from the test set.

5.1 Compression ratio:

Figure 7 shows the mean compression ratio of different industry standards and third-party formats on the test set. Where JPEG2000-Lossless (JP2), Run Length Encoding (RLE), and ZIP Deflate are

Table 1. Comprehensive analysis of DICOM images compression Methods

Paper	year	Method	DICOM Dataset	Performance measures	Implementation
[10]	2020	Wavelet compression.	12 DICOM images.	Compression ratio and PSNR = [24.74 – 28.04] dB.	-
[11]	2024	DCT, DWT, FCA, and VQA.	COVID19/Stanford University (CT images).	Best NROI CR = [37.91 – 68.91] and Best NROI PSNR = [101.93 – 127.29] dB.	MATLAB 2018 (Core i7 3687U and 4 GB RAM)
[12]	2023	Golomb-Rice coding and Run-Length Encoding.	Angio, X-ray, MR-1, CT and MR-2.	Compression ratio = [2.14 – 4.46]	-
[13]	2023	Fractal pixel traversal coupled with delta and entropy coding.	TCIA - DICOM lung CT scans	Compression ratio = 2.42	Python 3.8
[14]	2023	Huffman coding, linear predictive coding and discrete wavelet.	DICOM images (ultrasound, MRI, and CT).	PSNR = [43.05, 44.03, and 47.13] dB and SSIM = [0.9566, 0.9835, and 0.9859] respectively.	MATLAB (8 GB RAM and Intel i7 10th)
[15]	2023	CNN based architecture.	General medical (not DICOM) datasets: CBIS-DDSM and InBreast.	PSNR = 29.44 dB and SSIM = 0.9779	-

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

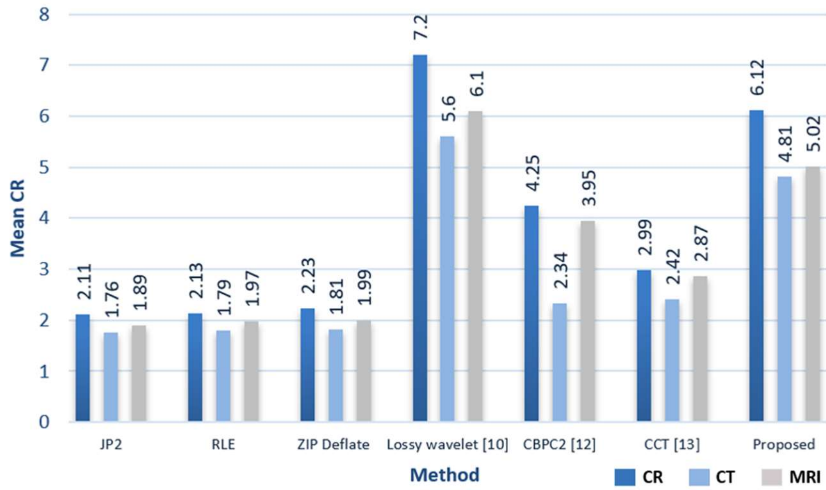

Figure 7. Mean compression ratio of the compression methods.

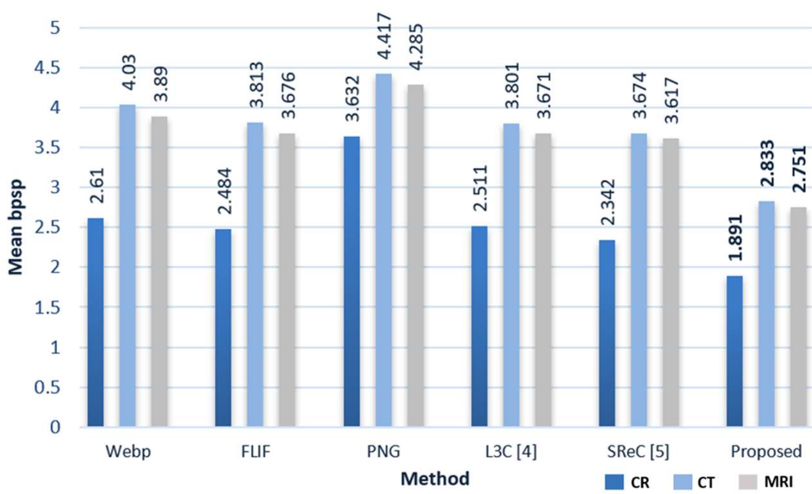

Figure 9. Comparison of mean value of bpsp metric.

integrated compression formats in the DICOM standard [13]. The results of applying each of CCT, which refers to the open-source CompaCT method [13], CBPC2, which depends on Golomb-Rice and Run-Length Encoding [12], and lossy Wavelet compression [10], are also shown. The results weren't compared with the proposed algorithms in [11] despite their superior values, due to their implementation on NROI only. The results clearly show the proposed scheme's superiority over other methods, except for the lossy wavelet compression [10], which achieves low quality, as observed through the low PSNR values in Table 1.

5.2. Compression time (CT):

While many studies don't use compression time as an evaluation metric due to the variability in working mechanisms, Figure 8 compares the mean compression time for the different methods. We can observe that traditional methods outperform learning-based methods in terms of speed, especially the RLE algorithm. The proposed method reduces compression time compared to learning-based methods by lightening the number of residual blocks and using DSC layers. It improves compression time by 19.02%, 17.76%, and 16.96% on average for CR, CT, and MRI images, respectively, compared to the SReC algorithm, which outperforms L3C.

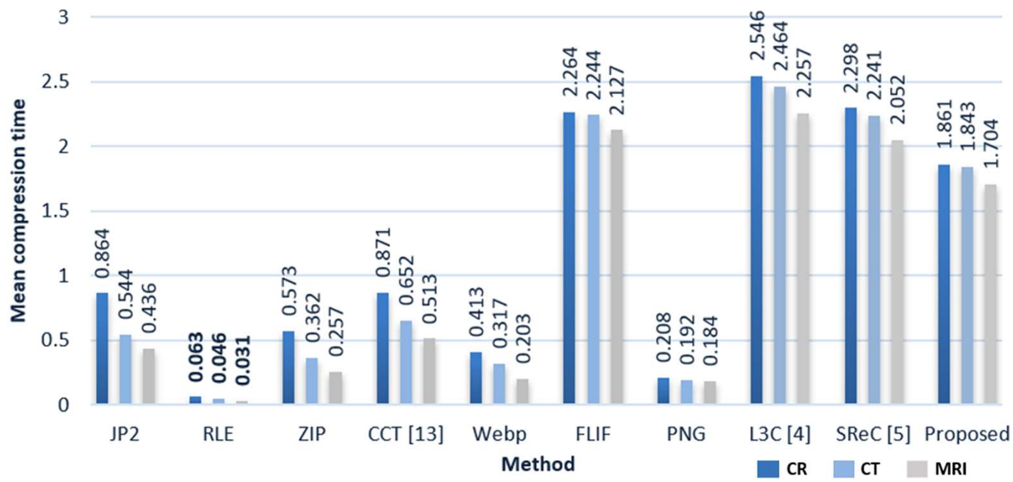

Figure 8. Mean compression time of the compression methods.

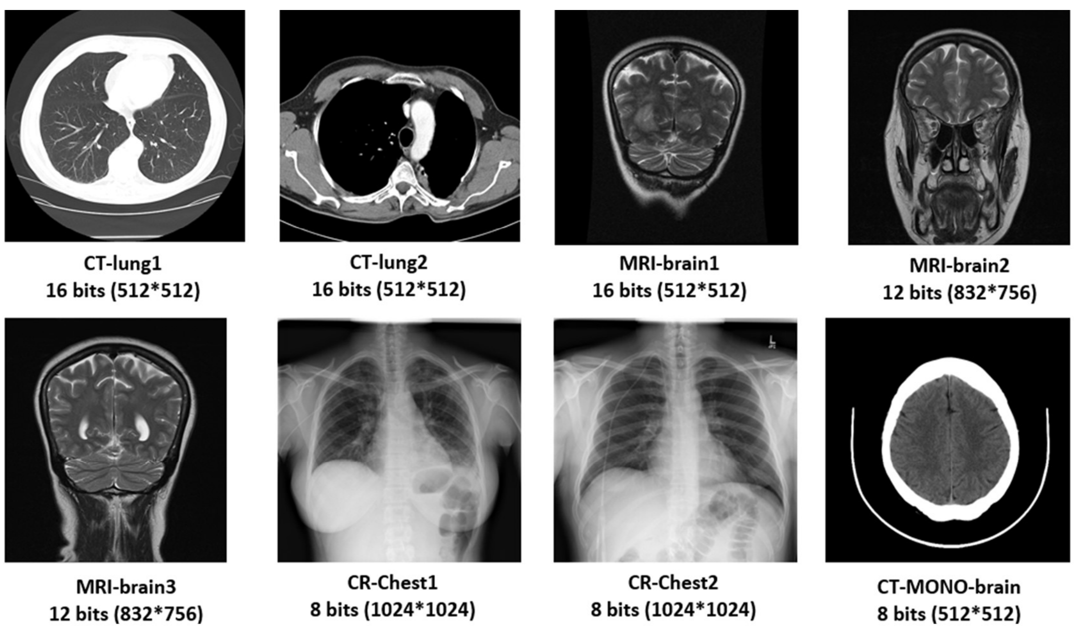

Figure 10. DICOM Images With Different Types, Dimensions, And Bit Depth.

5.3 Bits per sub-pixel (BPSP):

Bpsp is a state-of-the-art benchmark for evaluating image reconstruction mechanisms, especially in modern deep learning-based compression methods [4, 5, 6, 7]. Figure 9 compares this metric with the open-source learning-based L3C [4] and SReC [5] results.

Similar to [4, 5, 6, 13], Figure 9 also demonstrates the results of the engineered codecs, including Webp, PNG, and FLIF as external benchmarks, which aren't commonly used with medical images. The results show that the proposed scheme outperforms its nearest competitor (SReC) by approximately 19.26% for CR images, while it outperforms CT and MRI images by an average of 22.89% and 23.94%, respectively.

In order to test the image quality metrics on a set of images with a variety of dimensions and bit depths, Figure 10 shows a set of DICOM images of different parts of the body.

5.4 Peak signal-to-noise ratio (PSNR) and mean square error (MSE):

PNSR is the most preferred metric for measuring the image quality according to the computer vision tasks. It makes use of the maximum pixel value M,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and MSE metric between the original image I and the compressed image Î [2], as the following Eqs. (7 - 8):

$$PSNR = 10log_{10} (\frac{M^2}{MSE}) \quad (7)$$

$$MSE = \frac{1}{t} \sum_{i=1}^{t} \left(I(i) - \hat{I}(i)\right) \quad (8)$$

The results of testing DICOM images according to MSE and PSNR are included in Table 2.

5.5 Structural similarity index measure (SSIM):

Another essential standard in this field is SSIM, which relies on the human visual system and the perceptual similarity of image structure. This metric quantifies how the image structure is similar to the original image [2] and can be calculated using Eq. (9):

$$SSIM(I, \hat{I}) = \frac{2\,\mu_I \mu_{\hat{I}} + k1}{\mu_I^2 + \mu_{\hat{I}}^2 + k1} \cdot \frac{2\sigma_{I\hat{I}} + k2}{\sigma_I^2 + \sigma_{\hat{I}}^2 + k2} \quad (9)$$

Where μ and σ^2 are the mean and variance of each of the original and compressed images, respectively, and $\sigma_{I\hat{I}}$ is the covariance between these two images. Also, k_1 and k_2 are constants added to prevent instability when the denominator approaches zero. The SSIM values for the tested DICOM images are shown in Table 2.

Table 2. The values of PSNR, MSE, and SSIM for various tested DICOM images.

DICOM image	Type	PSNR (dB)	MSE	SSIM
CT-lung1	CT	52.23	0.38	0.9963
CT-lung2	CT	53.36	0.29	0.9946
MRI-brain1	MRI	52.19	0.39	0.9897
MRI-brain2	MRI	51.44	0.46	0.9891
MRI-brain3	MRI	50.96	0.52	0.9874
CR-chest1	CR	47.92	1.04	0.9837
CR-chest2	CR	48.14	0.98	0.9841
CT-MONO-brain	CT	50.51	0.57	0.9891

Except for [11], which achieved the highest PSNR values by applying the compression to NROI only, the results in Table 2 show the proposed method's superiority over the average values obtained by other existing techniques [10, 14, 15], which are shown in Table 1. Figure 11 demonstrates the advancement of the suggested scheme according to the PSNR/SSIM values of a test sample compared to the open-source compression methods.

These results indicate that the proposed method achieves high efficiency in compressing DICOM images compared to the existing methods, as demonstrated by the compression ratio and bpsp values, while reducing compression time compared to the learning-based compression methods. Image quality metrics also demonstrate high image restoration accuracy across a wide and varied range of medical image types, ensuring accurate patient diagnosis. However, although the proposed algorithm achieves superiority in terms of compression efficiency and quality standards, non-learned compression algorithms still relatively outperform in terms of compression time. Also, although the training process included the most important types of medical images according to the DICOM standard, other types can be included in the training process, as with the case of some low-resolution images.

6. CONCLUSION:

In this paper, we tested a proposed learned SR architecture for DICOM image compression. To learn the essential global structures as well as crucial details for medical diagnosis, we developed a non-autoregressive predictor with a three scales parallel hierarchy for all pixels, aiming to characterize the joint distribution of the image alongside the auxiliary feature representation. While current learning-based compression methods use the original residual blocks, we replaced them with edited dilated blocks, using increased dilation rates to capture long-term features. This reduced the number of these blocks by half, without affecting the feature extraction task, and significantly decreased the computational complexity and overall compression time. In addition, combining DSC layers with a pixel attention mechanism in the upsampling task, instead of the ESPCN model or other methods that rely on traditional convolutional layers, contributed to a significant 85.75% reduction in the number of operations per layer, while preserving critical diagnostic features in the image using PA. Also, using a discretized mean and scale Gaussian distribution mixture model instead of the logistic model utilized in the compared methods improved the SSIM and PSNR values across various DICOM images, in terms of dimensions, bit depth, and image type. All the above modifications made it possible to delete the ASPP model or its enhanced version (Dense ASPP) at the end of the image reconstruction task, which simplified the proposed scheme and accelerated its implementation.

The results demonstrated the effectiveness of the proposed system according to the quality and testing criteria, achieving higher compression ratios, and outperforming the best current techniques in terms of the bpsp metric, with improvement ratios of 19.26%, 22.89%, and 23.94% for CR, CT, and MRI images,

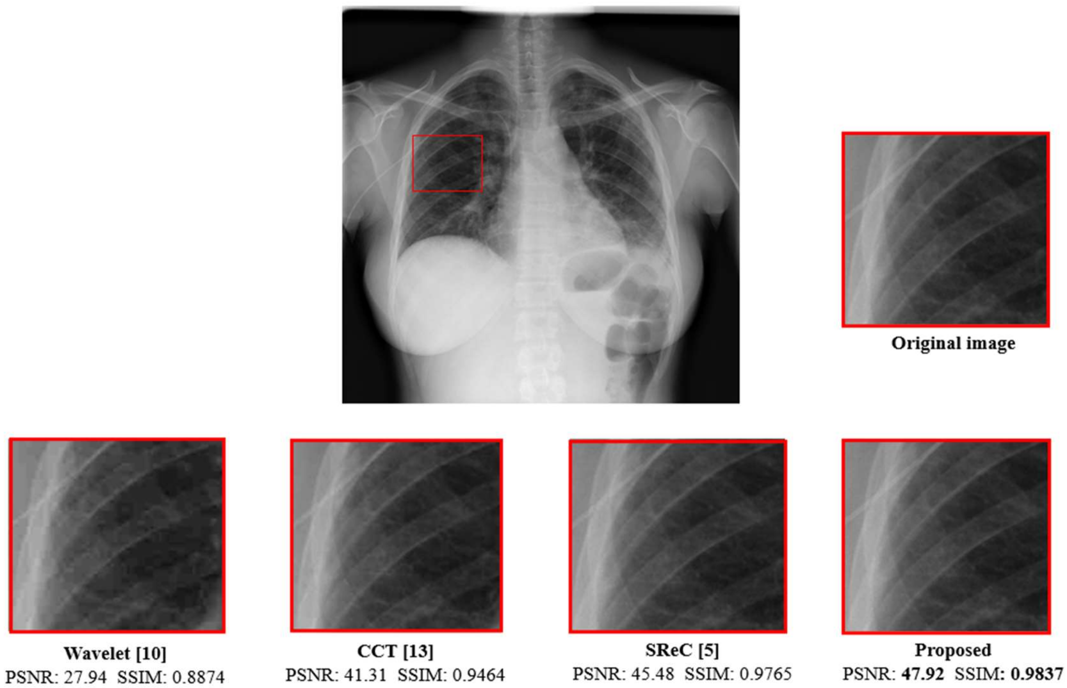

Figure 11. PSNR and SSIM values of testing sample [44] for different compression method.

respectively. Regarding the total compression time, while traditional methods generally outperform learning-based methods, the proposed system reduced the overall compression execution time by approximately 17% to 19% for various types of DICOM images, compared to the competing learned approaches. Related to PSNR and SSIM metrics, the results showed high-quality values across different image dimensions, types, and bit depths, demonstrating the ability to ensure accurate diagnosis.

In the future, it is possible to develop the proposed scheme by using low-cost multivariate mixture models for learned medical image compression or improving the upsampling task via employing another hybrid attention mechanisms, such as squeeze-and-excitation (SE) blocks and convolutional block attention module (CBAM). It is also possible to expand its use to include ultrasound DICOM images and explore the potential deployment of the proposed mechanism on medical IoT devices for telemedicine use.

AUTHOR CONTRIBUTIONS

Obada Othman Agha: Conceptualization, Writing an original draft, Software, Formal analysis, Reviewing and editing; Yahia Fareed: Investigation, Supervision, Methodology, Reviewing and editing, Project administration; Louay Chachati: Resources, Research design, Data analysis.

REFERENCES

[1] Larobina M. Thirty years of the DICOM standard. Tomography. 2023 Oct 6;9(5):1829-38. https://doi.org/10.3390/tomography9050145

[2] Bourai NE, Merouani HF, Djebbar A. Deep learning-assisted medical image compression challenges and opportunities: systematic review. Neural Computing and Applications. 2024 Jun;36(17):10067-108. https://doi.org/10.1007/s00521-024-09660-8

[3] Fu H, Liang F. Learned image compression with generalized octave convolution and cross-resolution parameter estimation. Signal Processing. 2023 Jan 1;202:108778. https://doi.org/10.1016/j.sigpro.2022.108778

[4] Mentzer F, Agustsson E, Tschannen M, Timofte R, Gool LV. Practical full resolution learned lossless image compression.

InProceedings of the IEEE/CVF conference on computer vision and pattern recognition 2019 (pp. 10629-10638). DOI: 10.1109/CVPR.2019.01088

[5] Cao S, Wu CY, Krähenbühl P. Lossless Image Compression through Super-Resolution. arXiv e-prints. 2020 Apr:arXiv-2004. https://doi.org/10.48550/arXiv.2004.02872

[6] Lan Y, Qin L, Sun Z, Xiang Y, Sun J. GOLLIC: Learning global context beyond patches for lossless high-resolution image compression. arXiv preprint arXiv:2210.03301. 2022 Oct 7. https://doi.org/10.48550/arXiv.2210.03301

[7] Wang R, Liu J, Sun H, Katto J. Learned lossless image compression with combined channel-conditioning models and autoregressive modules. IEEE Access. 2023 Jul 3;11:73462-9. DOI: 10.1109/ACCESS.2023.3291591

[8] Soltanmohammadi R, Faroughi SA. A comparative analysis of super-resolution techniques for enhancing micro-CT images of carbonate rocks. Applied Computing and Geosciences. 2023 Dec 1;20:100143. https://doi.org/10.1016/j.acags.2023.100143

[9] Duan Y, Liu Z, Li M. Super-resolution reconstruction of sea surface pollutant diffusion images based on deep learning models: a case study of thermal discharge from a coastal power plant. Frontiers in Marine Science. 2023 Jun 9;10:1211981. https://doi.org/10.3389/fmars.2023.1211981

[10] Shakya A, Ramola A, Tandon Y, Vidyarthi A. A Matrix-based Mathematical Model for Redundant Data Reduction from DICOM Images. In5th International Conference on Next Generation Computing Technologies (NGCT-2019) 2020 Feb 27. http://dx.doi.org/10.2139/ssrn.3545074

[11] Shakya AK, Vidyarthi A. Comprehensive study of compression and texture integration for digital imaging and communications in medicine data analysis. Technologies. 2024 Jan 24;12(2):17. https://doi.org/10.3390/technologies12020017

[12] Pervan B, Tomic S, Ivandic H, Knezovic J. MIDOM—a DICOM-based medical image communication system. Applied Sciences. 2023 May 15;13(10):6075. https://doi.org/10.3390/app13106075

[13] Khan T. CompaCT: Fractal-Based Heuristic Pixel Segmentation for Lossless Compression of High- Color DICOM Medical Images. arXiv preprint arXiv:2308.13097. 2023 Aug 24. https://doi.org/10.48550/arXiv.2308.13097

[14] Latha HR, Ramaprasath A. HWCD: A hybrid approach for image compression using wavelet, encryption using confusion, and decryption using diffusion scheme. Journal of Intelligent Systems. 2023 Jan 30;32(1):2022905. https://doi.org/10.1515/jisys-2022-9056

[15] Raj A, Sathish R, Sarkar T, Sethuraman R, Sheet D. Designing deep neural high-density compression engines for radiology images. Circuits, Systems, and Signal Processing. 2023 Feb;42(2):643-82. https://doi.org/10.1007/s00034-022-02222-0

[16] Hassan M, Illanko K, Fernando XN. Single Image Super Resolution Using Deep Residual Learning. AI. 2024 Mar 21;5(1):426-45. https://doi.org/10.3390/ai5010021

[17] Fu H, Liang F, Lin J, Li B, Akbari M, Liang J, Zhang G, Liu D, Tu C, Han J. Learned image compression with gaussian-laplacian-logistic mixture model and concatenated residual modules. IEEE Transactions on Image Processing. 2023 Apr 3;32:2063-76. DOI: 10.1109/TIP.2023.3263099

[18] Falahkheirkhah K, Yeh K, Confer MP, Bhargava R. DRB-net: Dilated residual block network for infrared image restoration. InInternational Symposium on Visual Computing 2022 Oct 3 (pp. 104-115). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-20716-7_9

[19] Ooi YK, Ibrahim H. Deep learning algorithms for single image super-resolution: a systematic review. Electronics. 2021 Apr 6;10(7):867. https://doi.org/10.3390/electronics10070867

[20] Abdou MA. Literature review: Efficient deep neural networks techniques for medical image analysis. Neural Computing and Applications. 2022 Apr;34(8):5791-812. https://doi.org/10.1007/s00521-022-06960-9

[21] Ağalday MF, Çinar A. Improvement of a Subpixel Convolutional Neural Network for a super-Resolution Image. Applied Sciences. 2025 Feb 25;15(5):2459. https://doi.org/10.3390/app15052459

[22] Wang Z, Ji S. Smoothed dilated convolutions for improved dense prediction. InProceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data

mining 2018 Jul 19 (pp. 2486-2495). https://doi.org/10.1145/3219819.3219944

[23] Salehi A, Balasubramanian M. DDCNet: Deep dilated convolutional neural network for dense prediction. Neurocomputing. 2023 Feb 28;523:116-29. https://doi.org/10.1016/j.neucom.2022.12.024

[24] Khalfaoui-Hassani I, Pellegrini T, Masquelier T. Dilated convolution with learnable spacings. arXiv preprint arXiv:2112.03740. 2021 Dec 7. https://doi.org/10.48550/arXiv.2112.03740

[25] Takahashi N, Mitsufuji Y. Densely connected multi-dilated convolutional networks for dense prediction tasks. InProceedings of the IEEE/CVF conference on computer vision and pattern recognition 2021 (pp. 993-1002). https://doi.org/10.48550/arXiv.2011.11844

[26] Yuan Y, Fu R, Huang L, Lin W, Zhang C, Chen X, Wang J. Hrformer: High-resolution vision transformer for dense predict. Advances in neural information processing systems. 2021 Dec 6;34:7281-93. https://doi.org/10.48550/arXiv.2110.09408

[27] Hanoon FS, Alasadi AH. A modified residual network for detection and classification of Alzheimer's disease. International Journal of Electrical and Computer Engineering (IJECE). 2022 Aug 1;12(4):4400-7. DOI: 10.11591/ijece.v12i4.

[28] Hamida A, Alfarraj M, Zummo SA. ESC-PAN: An Efficient CNN Architecture for Image Super-Resolution. IEEE Access. 2023 Jul 20;11:76120-31. DOI: 10.1109/ACCESS.2023.3297144

[29] Anwar S, Khan S, Barnes N. A deep journey into super-resolution: A survey. ACM computing surveys (CSUR). 2020 May 28;53(3):1-34. https://doi.org/10.1145/339046

[30] Hung KW, Zhang Z, Jiang J. Real-time image super-resolution using recursive depthwise separable convolution network. IEEE Access. 2019 Jul 16;7:99804-16. DOI: 10.1109/ACCESS.2019.2929223

[31] Li B, Wang B, Liu J, Qi Z, Shi Y. s-lwsr: Super lightweight super-resolution network. IEEE Transactions on Image Processing. 2020 Aug 13;29:8368-80. DOI: 10.1109/TIP.2020.3014953

[32] Lu Y, Jiang M, Wei L, Zhang J, Wang Z, Wei B, Xia L. Automated arrhythmia classification using depthwise separable convolutional neural network with focal loss. Biomedical Signal Processing and Control. 2021 Aug https://doi.org/10.1016/j.bspc.2021.102843

[33] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. InProceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 7132-7141). https://doi.org/10.48550/arXiv.1709.01507

[34] Sunil CK, Jaidhar CD. Tomato plant disease classification using multilevel feature fusion with adaptive channel spatial and pixel attention mechanism. Expert Systems with Applications. 2023 Oct 15;228:120381. https://doi.org/10.1016/j.eswa.2023.120381

[35] Zhao H, Kong X, He J, Qiao Y, Dong C. Efficient image super-resolution using pixel attention. InComputer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16 2020 (pp. 56-72). Springer International Publishing. https://doi.org/10.1007/978-3-030-67070-2_3

[36] Zhang W, Song A. Super-resolution image reconstruction network based on dual-channel pixel attention mechanism. InJournal of Physics: Conference Series 2022 Feb 1 (Vol. 2170, No. 1, p. 012005). IOP Publishing. DOI: 10.1088/1742-6596/2170/1/012005

[37] Xiao J, Guo H, Yao Y, Zhang S, Zhou J, Jiang Z. Multi-scale object detection with the pixel attention mechanism in a complex background. Remote Sensing. 2022 Aug 16;14(16):3969. https://doi.org/10.3390/rs14163969

[38] Othman Agha O, Fareed Y, Chachati L. Proposed High Resolution Method for Medical DICOM Images Compression using Deep Learning. Research journal of Aleppo university. 2022.

[39] Wang H, Zhao F, An X, Zhao Y, Li K, Guo Q. TIANet: A Defect Classification Structure Based on the Combination of CNN and ViT. Electronics. 2025 Apr 9;14(8):1502. https://doi.org/10.3390/electronics14081502

[40] Mishra D, Singh SK, Singh RK. Deep architectures for image compression: a critical review. Signal Processing. 2022 Feb 1;191:108346. https://doi.org/10.1016/j.sigpro.2021.108346

[41] Cheng Z, Sun H, Takeuchi M, Katto J. Learned lossless image compression with a hyperprior and discretized gaussian mixture likelihoods. InICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2020 May 4 (pp. 2158-2162). IEEE. DOI: 10.1109/ICASSP40776.2020.9053413

[42] Riaz F, Rehman S, Ajmal M, Hafiz R, Hassan A, Aljohani NR, Nawaz R, Young R, Coimbra M. Gaussian mixture model based probabilistic modeling of images for medical image segmentation. IEEE Access. 2020 Jan 17;8:16846-56. DOI: 10.1109/ACCESS.2020.2967676

[43] Cheng Z, Sun H, Takeuchi M, Katto J. Learned image compression with discretized gaussian mixture likelihoods and attention modules. InProceedings of the IEEE/CVF conference on computer vision and pattern recognition 2020. https://doi.org/10.48550/arXiv.2001.01568

[44] Thakur, A., siim_dicom_images dataset, 2019. [Online]. Available:https://www.kaggle.com/datasets/abhishek/siim-dicom-images

[45] MedAI Data Hub, Brain MRI Dataset for Medical Imaging Research, 2025.[Online]. Available:https://www.kaggle.com/datasets/medaidatahub/brain-mri-dataset-for-ai

[46] D Goldgof et al., Qin lung CT Version 2 (The Cancer Imaging Archive, 2017), doi:10.7937/K9/TCIA.2015.NPGZYZBZ.