Machine Vision Guided 3D Medical Image Compression for Efficient Transmission and Accurate Segmentation in the Clouds

Zihao Liu[1], Xiaowei Xu[2], Tao Liu[1], Qi Liu[1], Yanzhi Wang[3], Yiyu Shi[2], Wujie Wen[1], Meiping Huang[4], Haiyun Yuan[4], Jian Zhuang[4]

[1] Flordia International University, [2] University of Notre Dame, [3] Northeastern University, [4] Guangdong General Hospital

{zliu021,tliu023,qliu020,wwen}@fiu.edu, {xxu8,yshi4}@nd.edu,yanz.wang@northeastern.edu

Abstract

Cloud based medical image analysis has become popular recently due to the high computation complexities of various deep neural network (DNN) based frameworks and the increasingly large volume of medical images that need to be processed. It has been demonstrated that for medical images the transmission from local to clouds is much more expensive than the computation in the clouds itself. Towards this, 3D image compression techniques have been widely applied to reduce the data traffic. However, most of the existing image compression techniques are developed around human vision, i.e., they are designed to minimize distortions that can be perceived by human eyes. In this paper we will use deep learning based medical image segmentation as a vehicle and demonstrate that interestingly, machine and human view the compression quality differently. Medical images compressed with good quality w.r.t. human vision may result in inferior segmentation accuracy. We then design a machine vision oriented 3D image compression framework tailored for segmentation using DNNs. Our method automatically extracts and retains image features that are most important to the segmentation. Comprehensive experiments on widely adopted segmentation frameworks with HVSMR 2016 challenge dataset show that our method can achieve significantly higher segmentation accuracy at the same compression rate, or much better compression rate under the same segmentation accuracy, when compared with the existing JPEG 2000 method. To the best of the authors' knowledge, this is the first machine vision guided medical image compression framework for segmentation in the clouds.

1. Introduction

Deep learning has significantly pushed forward the frontier of automatic medical image analysis [9][28][46][7][9][24][6][12][39]. On the other hand, most deep learning based frameworks have high computation complexities [41][47][45][48][44][16][17]. For example, the number of operations needed by the network by [8] to segment a 3D Computed Tomography (CT) volume would be around 2.2 Tera (10^{12}) , which needs days to be processed on a general desktop computer. In addition, with the advances in medical imaging technologies, the related data has been increasing exponentially for decades [11]. Ponemon Institute survey found that 30% of the worlds data storage resides in the healthcare industry by 2012 [13]. For both reasons, clouds have become a popular platform for efficient deep learning based medical image analysis [22][52][42][43][51].

Utilizing clouds, however, requires medical images to be transmitted from local to servers. Compared with computation time needed to process these images in the clouds, the transmission time is usually higher. For example, the latency to transmit a 3D CT image of size 300MB is about 13 seconds via fixed broadband internet (estimated with 2017 U.S. average fixed broadband upload speed of 22.79 Mbps [25]). On the other hand, it takes no more than 100 milliseconds for 3D-DSN [12] to segment an image through a high-performance cluster of 10 GPUs in cloud [21][10][18]. For slower internet speed, this gap is even bigger.

To tackle this issue, image compression is typically used to prune unimportant information before sending the image to clouds, thus reducing data traffic. The compression time is usually negligible (e.g., 24 milliseconds to compress a 300MB 3D CT image to 30MB using a moderate GPU [23]). There exist many general image compression standards such as JPEG-2000 [4][3], JPEG [38], and MPEG2 [14]. Most of these standards use frequency transformation to filter out information that leads to little visual distortion. In addition to the existing 3D image compression standards, alternative compression methods have been proposed in the literature, most of which modify the existing standards to improve their performance [5][30][29][49]. There are also a few methods for lossless compression of 3D medical images [31][20].

Almost all the existing compression techniques are optimized for the Human-Visual System (HVS), or image quality perceived by humans. However, when we compress im-
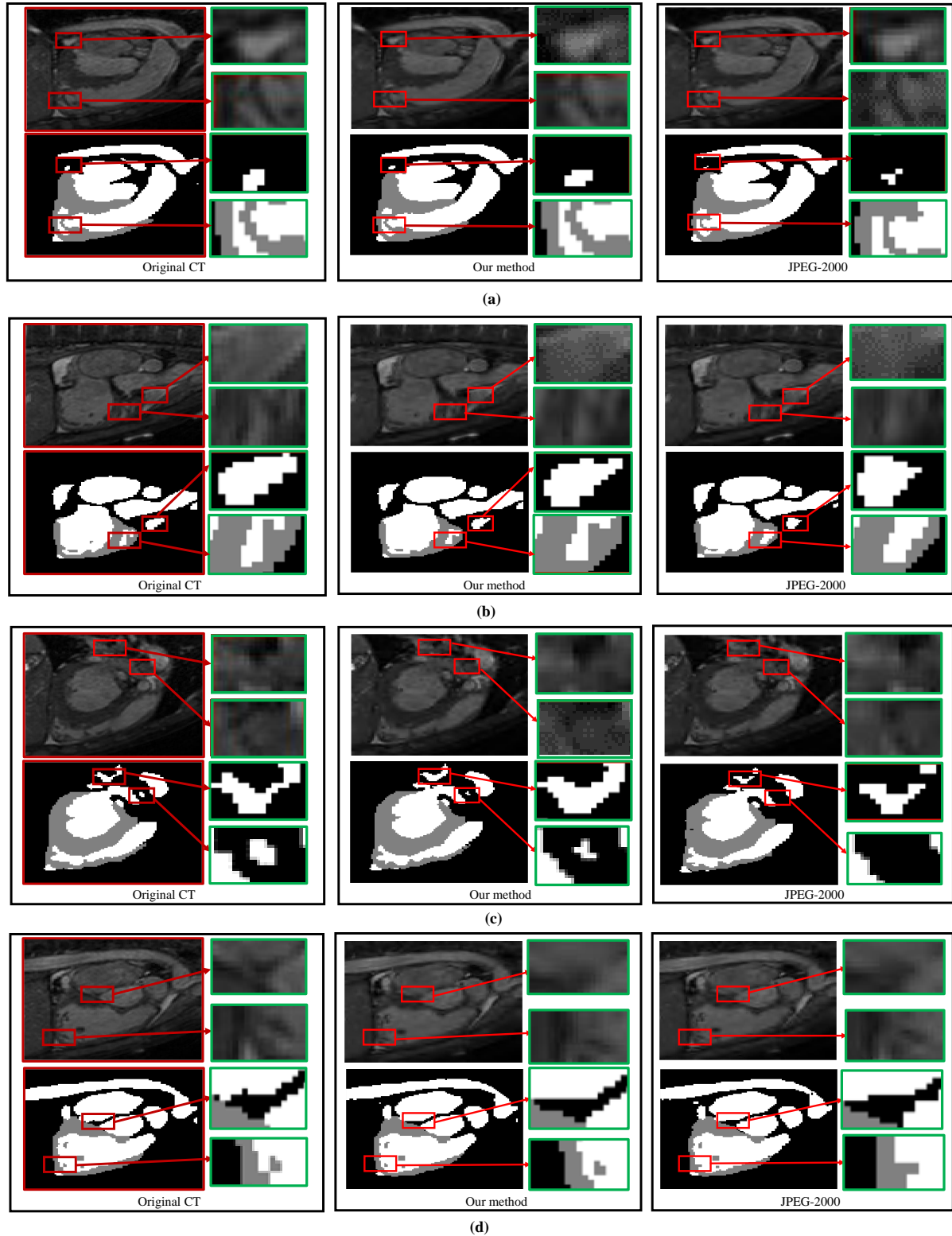
1

Figure 1. Segmentation details of four slices in a CT image in HVSRMR 2016 Challenge dataset [26], compressed using our method and JPEG-2000, and segmented by DenseVoxNet [50]. Many details are missing in the segmentation results from JPEG-2000 compressed images but not in our method. Quantitative comparisons can be found in Section 4.

ages for transmission to the clouds, their quality will not be judged by human vision, but rather by the performance of the neural networks that process them in the clouds. As such, an interesting question that naturally arises is: are the existing compression techniques still optimal for these neural networks, i.e., in terms of "machine visions"? In this paper, we will use 3D medical image segmentation as a vehicle to study this question.

Medical image segmentation extracts different tissues, organs, pathologies, and biological structures to support medical diagnosis, surgical planning and treatments. We adopt JPEG-2000 to compress the HVSMR 2016 Challenge dataset [26], and two state-of-the-art neural networks– DenseVoxNet [50] and 3D-DSN [12] for medical image segmentation. The results for four randomly selected slices are shown in Fig. 1. From the figure we can see that quite significant differences exist between the segmentation results from the original image and the one compressed by JPEG-2000, though visually little distortions exist between the two.

The results may seem surprising at first glance, but it is also fully justifiable. The boundaries in medical images mainly contribute to the high frequency details, which cannot be perceived by human eyes. As such, existing compression techniques will ignore them while still attaining excellent compression quality. Yet these details are critical features that neural networks need to extract to accurately segment an image. Similarly, many low frequency features in a medical image such as brightness of a region are important for human vision guided compression, but not at all for segmentation. In other words, human vision and machine vision are completely different with regard to the segmentation task.

In this paper, we propose a machine vision guided 3D image compression framework tailored for deep learning based medical image segmentation in the clouds. Different from most existing compression methods that take human visual distortion as guide, our method extracts and retains features that are most important to segmentation, so that the segmentation quality can be maintained. We conducted comprehensive experiments on two widely adopted segmentation frameworks (DenseVoxNet [50] and 3D-DSN [12] using the HVSMR 2016 Challenge dataset [26]. Examples on the qualitative effect of our method on the final segmentation results can be viewed in Fig. 1.

The main contributions of our work are as follows:

- We discovered that for medical image segmentation in the clouds, traditional compression methods guided by human vision will result in inferior accuracy, and a new method guided by machine vision is warranted.

- We proposed a method that can automatically extract important frequencies for neural network based image

segmentation, and map them to quantization steps for better compression.

- Experimental results show our method outperforms JPEG-2000 in two aspects: for a same compression rate, our method achieves significantly improved segmentation accuracy; for a same level of segmentation accuracy, it offers much higher compression rate ($3\times$). These advantages demonstrate great potentials for its application in today's deep neural network assisted medical image segmentation.

2. Related Work

2.1. 3D Medical Image Compression

There are many general image compression standards such as JPEG-2000 [4][3], JPEG [38]. Some video coding standards such as H.264/AVC [37],and MPEG2 [14] can also be adopted for 3D image segmentation. Most of these standards use transforms such as Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) for compression while preserving important visual information for humans.

In addition to the existing 3D medical image compression standard, alternative compression methods have been proposed in the literature. Most of the methods modified the existing standards to improve its performance. Bruylants *et al*. [5] adopted volumetric wavelets and entropy-coding to improve the compression performance. Sanchez *et al*. [30] employed a 3-D integer wavelet transform to perform column of interest coding. Sanchez *et al*. [29] reduced the energy of the sub-bands by exploiting the anatomical symmetries typically present in structural medical images. Zhongwei *et al*. [49] improved the compression performance by removing unimportant image regions not required for medical diagnosis. There are a few methods for lossless compression of 3D medical images. Santos *et al*. [31] processed each frame sequentially and using 3D predictors based on the previously encoded frames. Lucas *et al*. [20] further adopted 3D block classification to process the data at the volume level.

Almost all the above methods still adopt the same objective as that used by JPEG-2000, i.e., to minimize human perceived distortions. As shown in the example in Fig. 1, when it comes to the deep learning based segmentation, such a strategy may lead to poor accuracy.

2.2. JPEG-2000 3D image Compression

Our method is also based on JPEG-2000 but modifies its human vision guided objective to one that is guided by the segmentation network. Here we briefly review the details of JPEG-2000 so that later we can explain our work better. Fig. 2 shows the major steps in JPEG-2000 compression:
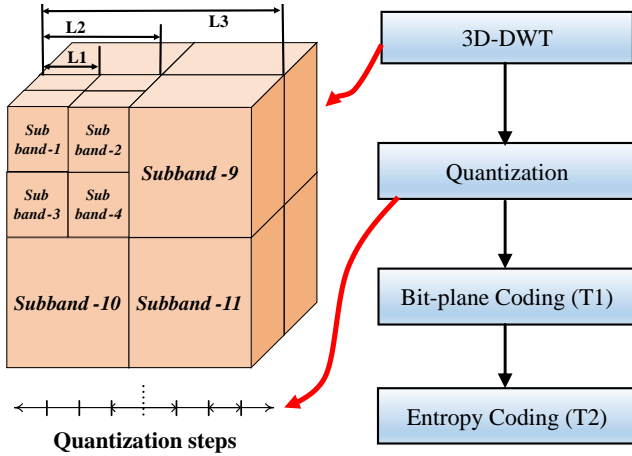
Figure 2. Flow of JPEG-2000 compression method.

First, the 3D discrete wavelet transform (DWT) is applied to an image to decompose it into a multiple-resolution representation in frequency domain [33][2][27]. For example, a 3-D wavelet decomposition leads to three resolution levels (L1, L2, L3). Each resolution level (except L1) is composed of eight subbands: subband 1 to subband 8. The eight lower resolution levels are always generated by progressively applying the 3D DWT process to the upper-left-front block (e.g., subband 1) from the previous resolution level. Then a non-uniform quantization process is applied to each subband based on the number of low pass filters in the subband:

$$x' = \lfloor \frac{x}{QS} \rfloor \qquad (1)$$

where x is the original coefficient after 3D DWT, QS is the quantization step of a subband and x' is the coefficient after quantization.

The rule is that the more low pass filters a subband has the smaller quantization step are applied to the corresponding subband. This is because Human Visual System (HVS) is more sensitive to low pass frequency information, thus less quantization errors in low pass subband. Bit-plane coding and entropy coding mainly perform coding and please interested readers are referred to the related literature [36][32][34][36] for more details.

3. Machine Vision oriented 3D Image Compression

In this section, the details of the proposed machine vision oriented 3D image compression framework for segmentation in the clouds is presented. As shown in Fig. 3, the framework contains two modules: frequency analysis module and mapping module. Compared with original JPEG-2000 compression method, the added two modules can obtain optimized quantization steps (QSs) for better segmentation accuracy. The frequency analysis module extracts frequencies important to segmentation with high statistic indexes (SI) using a machine vision guided frequency model.

The mapping module maps these SIs to optimized QSs which are further provided to the quantization module in JPEG-2000 flow for the rest of the processing. Particularly parameter optimization is also performed to find the optimal parameters in the mapping module.

3.1. Frequency Analysis Module

3.1.1 Machine Vision Guided Frequency Model

In this section we build a frequency model that identifies information most useful for segmentation. Assume x_i is a single voxel of a raw 3D image $\mathbf{X}$. x_i can be represented by 3D-DWT at one resolution level in JPEG-2000 compression as:

$$x_i = \sum_{n=0}^{i=N-1} c_i^n \cdot b_i^n \qquad (2)$$

where c_i^n and b_i^n are the 3D-DWT coefficient at matching 3D coordinate i and corresponding basis function at N different subbands, respectively.

For human visual system, the quantization step (QS) for each subband in JPEG-2000 is positively correlated with the number of high pass filters in a subband. For example, the QS of subband 4 is larger than that of subband 2 at the same resolution level. Then larger QS in high frequency subband will increase the distortion of coefficients in this subband. Consequently, it will either directly zero out the associated 3D-DWT coefficient c_i^n or increase the chance to truncate them at rate-distortion optimization process. This is because HVS is less sensitive to high frequency subband, so a high compression rate can be achieved by discarding the high frequency information.

In order to obtain the important frequency for DNN based segmentation, we calculate the gradient of the DNN loss function F with respect to a basis function b_i^n as:

$$\frac{\partial F}{\partial b_i^n} = \frac{\partial F}{\partial x_i} \times \frac{\partial x_i}{\partial b_i^n} = \frac{\partial F}{\partial x_i} \times c_i^n. \qquad (3)$$

Equation (3) indicates that the importance of information at different subbands of a single voxel to DNN is determined by its associated 3D-DWT coefficients (c_i^n) at all subbands. This is quite different from HVS which distorts c_i^n in high frequency subbands (i.e. quantization or truncation). Large c_i^n in high frequency subband will be heavily distorted in JPEG-2000. However, it may carry important information for DNN segmentation, causing accuracy degradation.

3.1.2 Frequency Extraction

In this section, we extract important frequencies based on the above frequency model. Previous studies [35][19] have demonstrated that the distribution of un-quantized 3D-DWT
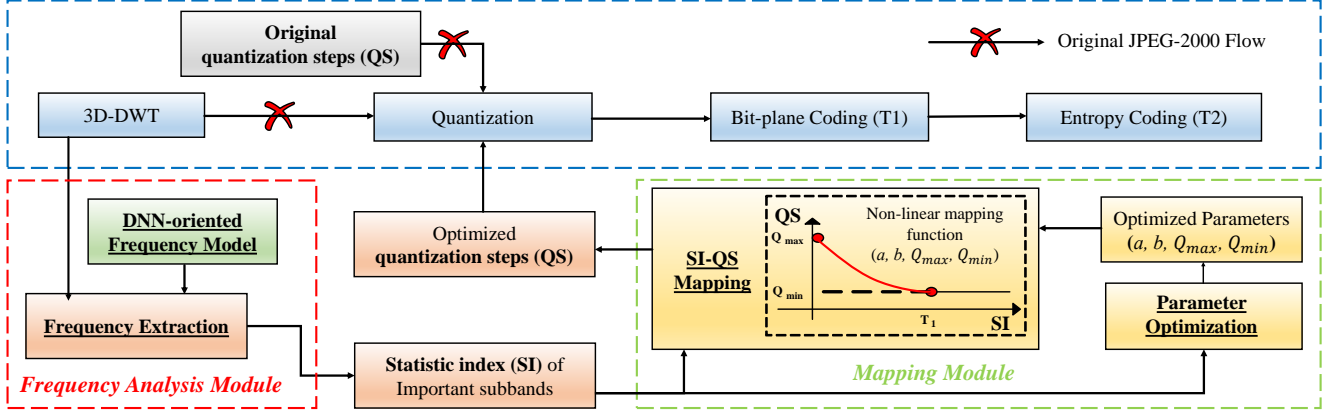
Figure 3. Overview of the proposed DNN-oriented 3D image compression framework.

coefficients in a subband indicates the energy in this subband. Moreover, the distribution of each subband has been proven that they approximately obey a Laplace distribution with zero mean and different standard deviations (δ_n). The larger δ_n a subband has (i.e. more energy in this subband), the more contribution this subband will provide to DNN results. Therefore, δ_n of each subband after 3D-DWT are selected as the SI to represent the importance to DNN. Based on this we propose to conduct the frequency analysis as follows: the number of subbands will be first calculated based on the number of resolution levels at three different dimensions provided by users. After that coefficients that belong to the same subband will be grouped up and reshaped to one dimension. Then the distributions of reshaped coefficients at each subband will be characterized. Finally, the statistical information of each subband, i.e. the standard deviation or SI, will be calculated based on its histogram. The results from this frequency information projection procedure can clearly indicate the importance of each subband to DNN by its SI. With the above discussion, we further analyzed SIs and QSs in JPEG-2000 to show that JPEG-2000 is not optimized for DNNs. We randomly selected two images from HVSMR 2016 dataset labeled as A and B, and then applied our frequency extraction method on them after 3-3-3 3D-DWT. As shown in Fig. 4, some important subbands have large QS which is undesired. For example, subband 2 is less important than subband 3 for image A since $\delta_2 < \delta_3$, however, its QS is much smaller than that of subband 3. The same problem exists for subband 3 and subband 4 with image B. Thus, although lower frequency information is always more important than that of higher frequency in JPEG-2000, it is not the case for segmentation accuracy.

3.2. Mapping Module

3.2.1 SI-QS Mapping

With SIs at each subband, our next step is to find a suitable mapping between SI and QS by well leveraging the intrin-

sic error resilience characteristic of DNN computation. As a result, the segmentation accuracy loss due to increasing compression rate, can be minimized by largely quantilizing the frequency subbands that are less significant to DNN.

In order to precisely model the mapping, we attempt to find a QS curve aligning with most of the SIs. With extensive experiments (we add these experiments in the supplemental material), we observe that the QS-SI points obey a reciprocal function ($y = 1/x$). Thus, we propose a nonlinear mapping (NLM) method to implement nonuniform quantization steps at different subbands:

$$Q_n = \frac{a}{(\delta_n + b)}, \quad s.t. \quad Q_{min} \leq Q_n \leq Q_{max} \quad (4)$$

where Q_n is the quantization step at subband n, Q_{min} and Q_{max} are the smallest and largest QS, and a and b are the fitting parameters.

3.2.2 Parameter Optimization

With the proposed mapping function, parameter optimization is performed to obtain the optimal a, b, Q_{max} and Q_{min} in Equation (4). For a and b, we found that rational functions can fit the relationship between the standard deviation of each subband of an image and the quantization step
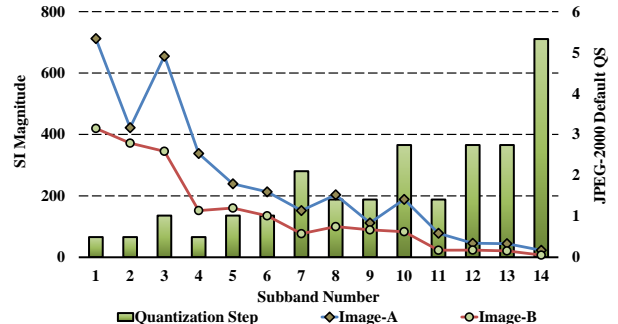

Figure 4. Diverse frequency domain of medical images.

Table 1. Segmentation results of our methods and JPEG-2000 using DenseVoxNet and HVSMR2016 dataset. The compression rate is set to 30 for both techniques. The images compressed by ours can be segmented with almost the same accuracy as, or sometimes even better than the original ones, much better than those compressed by JPEG-2000. The segmentation performance of NLM is very close to or even better than that with the original images while is much better than JPEG-2000.

		Original	Ours	JPEG-2000
Myocardium	**Dice**	0.838±0.0334	**0.834±0.0386**	0.816±0.042
	Hausdorff	30.879±7.592	**31±7.940**	33.513±7.566
	ASD	0.673±0.67	**0.652±0.671**	0.722±0.746
Blood Pool	**Dice**	0.915±0.025	**0.914±0.024**	0.912±0.025
	Hausdorff	41.034±9.326	**40.93±9.52**	41.031±9.648
	ASD	0.601±0.455	**0.556±0.432**	0.582±0.453
Compression Rate		1	**∼30x**	**∼30x**
PSNR (dB)		∞	∼35	**∼36**

very well. For Q_{max} and Q_{min}, we examine two corner cases, i.e. upper/lower corner to explore the quantization error tolerance for the most insignificant/significant subband. Then all the parameters in non-linear mapping method can be calculated by substituting pairs (Q_{min}, δ_{max}) and (Q_{max}, δ_{min}) into Equation (4).

Lower Corner Case: we assign the same QS to all the subbands to explore Q_{min}. As long as the error induced by QS in the subband with δ_{max} (the most significant subband to DNN) does not impact the segmentation accuracy, this will also hold true for all the other subbands.

Upper Corner Case: To find Q_{max}, we only vary QS at the subband with δ_{min}, while fixing that of all the other subbands as the same QS–Q_{min}. If the subband with δ_{min} (the least significant subband to DNN) cannot tolerate the error incurred by a Q_{max}, the other subbands cannot either.

4. Evaluation

4.1. Experiment Setup

Our proposed machine vision guided 3D image compression framework was realized by heavily modifying the open-source JPEG-2000 code [1]. This code also served as our baseline–JPEG-2000 for comparison.

Benchmarks: we adopted the HVSMR 2016 Challenge dataset [26] as our evaluation benchmark. This dataset consists of in total 10 3D cardiac MR scans for training and 10 scans for testing. Each image also includes three segmentation labels: myocardium, blood pool, and background.

Evaluation Metrics: We compared our method with the baseline (JPEG-2000) in following two aspects: 1) segmentation results; 2) compression rate. For the segmentation results, we followed the rule of HVSMR 2016 challenge where the results are ranked based on *Dice coefficient (Dice)*. The other two ancillary measurement metrics, i.e. *average surface distance (ASD)* and *symmetric Hausdorff distance (Hausdorff)*, were also calculated for reference. Among the three metrics, a higher Dice represents higher agreement between the segmentation result and the

ground truth, while lower ASD and Hausdorff values indicate higher boundary similarity.

Experiment Methods: To evaluate our methods comprehensively, two state-of-art segmentation neural network models–DenseVoxNet [50] and 3D-DSN [12] were selected. We followed the original settings of the two frameworks at training and testing phases but with compressed images. In the testing phase, since the ground truth labels of the selected dataset are not publicly available, we randomly selected five un-compressed training images for training and the rest compressed five for testing. All our experiments were conducted on a workstation which hosts NVIDIA Tesla P100 GPU and deep learning framework Caffe [15] integrated with MATLAB programming interface.

4.2. Optimal Parameter Selection

In this section, we experimentally find the optimal parameters for Q_{max}, Q_{min}, a and b in Equation (4), following the method discussed in Section 3.2.2.

We tested the two cases as discussed in Section 3.2.2 to find Q_{max}, Q_{min}. We took normalized dice coefficients and Hausdorff distance as segmentation measurements for an 3D cardiac MR scan and adopted the FCN

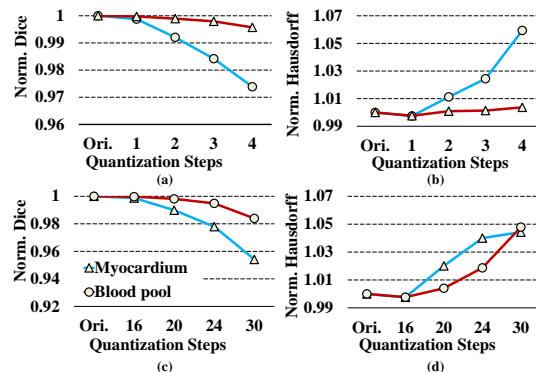

Figure 5. Optimal parameter selection of Q_{max} and Q_{min}.

Table 2. Segmentation results of our methods and JPEG-2000 using 3D-DSN and HVSMR 2016 dataset. The compression rate is set to 30 for both techniques. The images compressed by ours can be segmented with almost the same accuracy as the original ones, and significantly better than those compressed by JPEG-2000.

		Original	Ours	JPEG-2000
Myocardium	**Dice**	0.784±0.059	**0.786±0.059**	0.773±0.058
	Hausdorff	32.345±9.164	**31.002±8.988**	33.041±8.768
	ASD	0.310±0.171	**0.325±0.184**	0.355±0.224
Blood Pool	**Dice**	0.909±0.027	**0.908±0.030**	0.901±0.032
	Hausdorff	38.515±9.59	**38.601±9.951**	39.416±9.932
	ASD	0.235±0.200	**0.223±0.201**	0.230±0.204
Compression Rate		1	**∼30x**	**∼30x**
PSNR (dB)		∞	∼35	**∼36**

model–DenseVoxNet. The measurments for two classes–myocardium and blood pool, are reported. For the lower corner case, as Fig. 5 (a) and (b) show, the two measurments for both labels do not suffer from any degradation only if QS is not larger than 1. Therefore, $Q_{min} = 1$ should be selected as ensure the segmentation results. For the upper corner case, the results are shown in Fig. 5 (c) and (d). The two measurements decrease when the QS at δ_{min} is larger than 16 at both classes, by following a similar trend as the lower corner case. Hence, we chose $Q_{max} = 16$ as the upper bound for our quantization step. Based on Q_{max}, Q_{min} and Equation (4), a and b can be decided accordingly. In our evaluation, we only adopt DenseVoxNet, as an example, to obtain Q_{max}, Q_{min} so as to solve a and b in Equation (4). Then we directly apply it to both DenseVoxNet and 3D-DSN. Note that our method is model agnostic (or rather data specific), since Equation (3) indicates that the importance of subbands can largely rely on DWT coefficients without correlating with DNN model. Therefore, we can use the same tuned parameters in our compression regardless of network structure. This is also one of the advantages of our method.

4.3. Comparison of Segmentation Accuracy

We first evaluated how our proposed compression framework can improve the segmentation accuracy over the baseline–3D JEPG-2000 using the state-of-the-art segmentation neural network model–DenseVoxNet. For a fair comparison, both our method and 3D JPEG-2000 were implemented at the same compression rate (CR). For illustration purpose, we only report the segmentation accuracy at $CR = 30\times$ (results under other compression rates are summarized in the supplemental material). The mean and standard deviation of the three segmentation measurement metrics–Dice, ASD and Hausdorff, are calibrated from 5 testing images of HVSMR2016 dataset. Note that Dice is the most important metric among the three.

Table 1 reports the segmentation results of the two classes–myrocardium and blood pool for the three methods–original (uncompressed, $CR = 1\times$), ours and

JPEG-2000, under DenseVoxNet. **First**, the default 3D JPEG-2000 exhibits the worst segmentation results at all the three metrics among the three methods. This is as expected, since JPEG-2000 takes the human perceived image quality as the top priority by offering the highest PSNR (36). **Second**, our method, which is developed upon the "machine vision", can beat JPEG-2000 across all three metrics for both classes, with a lower PSNR (~ 35). Impressively, for myocardium, our method can significantly improve Dice, Hausdorff and ASD over JPEG-2000 by 0.018, 2.039, 0.3 on average, respectively. The improvements on blood pool, on the other hand, are relatively limited, given its much higher dice score (0.915 for blood pool v.s. 0.838 for myocardium). **Third**, compared with the original image for both classes, our method only slightly degrades the segmentation results, i.e. $0.001 \sim 0.004$ on average for Dice, but offers a much higher compression rate ($30\times$ v.s. $1\times$). We also observe that the degradation of all three metrics on compressed images of myocardium (w.r.t. original) is always more significant than blood pool, for both our method and JPEG-2000. This is because myoscardium has a lower dice score than blood pool due to the ambiguous border. These results are consistent with the previous work [50].

We would like to emphasize that the achieved performance improvement of our method is very significant for segmentation on the HVSMR 2016 Challenge dataset [40][50] (we also add detailed image by image segmentation results in the supplemental material). Tens of studies performed extensive optimization for segmentation on this dataset. While DenseVoxNet offers the best performance by far [50], compared with other implementations, it still only improves Dice but degrades Hausdorff and ASD. our method, on the other hand, obtains higher performance on all the three metrics on DenseVoxNet. Furthermore, compared with the second-best method [40][50], the average improvement of DenseVoxNet on Dice is 1.2%, while our method can achieve an average improvement of $\sim 1.8\%$ for Myocardium on DenseVoxNet.

We also extended the same evaluations to another state-

of-the-art FCN–3D-DSN, to explore the response of our method to different FCN architectures. As shown in Table 2, the trend of the results are similar to that of DenseVoxNet, except for lower segmentation accuracy. Note this is caused by the neural network structure difference, and DenseVoxNet currently achieves the state-of-the-art segmentation performance. As expected, again, our method significantly outperforms JPEG-2000 at the same compression rate ($30\times$) across all the three metrics, i.e. 0.013 (myoscardium) and 0.007 (blood pool) on average for dice score, while providing almost the same segmentation performance as that of uncompressed version–original ($1\times$). These results clearly show the generalization of our method.

It is also notable that from both tables, the segmentation results from compressed images using our method sometimes even outperform that of original images. This is because compression as frequency-domain filtering also has denoising property. Although the training process attempts to learn comprehensive features, the importance of the same frequency feature may vary from one image to another for a trained DNN. As a result, after compression, the segmentation accuracy of some images may be improved because the unnecessary features that can mislead the segmentation are filtered, as demonstrated in Fig. 1(b) (Our method is better than Original CT). For most images, the segmentation accuracy after compression is still slightly degraded compared with the original images due to minor information loss at high compression rates, though our compression method tries to minimize the loss of important features.

4.4. Comparison of Compression Rate

In this section, we explore to what extent our proposed machine vision-oriented compression framework can improve the compression with regard to the human-visual based 3D JPEG-2000, for medical image segmentation. For a fair compression, we compared the compression rate (CR) of these two methods under the same segmentation accuracy for myocardium using DenseVoxNet. Dice score (0.834) was selected as it is the prime metric to measure the quality of image segmentation. Since the compression rate may vary from one image to another, we chose three representative images from the dataset. As Fig. 6 shows, our method can always deliver the highest compression rate across all the images. On average, it achieves $30\times$ compression rate over the original uncompressed image. Compared with 3D-JPEG 2000, our method can still achieve $3\times$ higher image size reduction, without degrading the segmentation quality. Still taking the example from section 1, we assume the transmission time of a 3D CT image of size 300MB via fixed broadband internet ($22.79Mb$) to cloud is 13s, while the image segmentation computation time on cloud is merely 100ms. Putting these two together, a single image segmentation service time on cloud for our method ($30\times$)

and JPEG-2000 ($10\times$), are 0.53s and 1.4s, respectively, translating into $2.6\times$ speed up.

4.5. Overhead

Our method is built upon 3D-JPEG 2000 by only adding two simple operations: standard deviation calculation for 16 subbands and equation set solution (Equation (4)) with only four variables. Since we reuse the majority of JPEG-2000's function units, the compression and decompression time are at the same level as that of JPEG-2000, e.g., 0.12ms for a 512×512 image [23], which is almost negligible compared with image transmission and segmentation time. Therefore, we expect that our light-weighted machine vision guided 3D image compression framework can find broad applications in medical image analysis.

5. Conclusion

Due to the high computation complexity of DNNs and the increasingly large volume of medical images, cloud based medical image segmentation has become popular recently. Medical image transmission from local to clouds is the bottleneck for such a service, as it is much more time-consuming than neural network processing on clouds. Although there exist a lot of 3D image compression methods to reduce the size of medical image being transmitted to cloud hence the transmission latency, almost all of them are based on human vision which is not optimized for neural network, or rather, machine vision. In this paper, we first present our observation that machine vision is different from human vision. Then we develop a low cost machine vision guided 3D image compression framework dedicated to DNN-based image segmentation by taking advantage of such differences between human vision and DNN. Extensive experiments on widely adopted segmentation DNNs with HVSMR 2016 challenge dataset show that our method significantly beats existing 3D JPEG-2000 in terms of segmentation accuracy and compression rate.
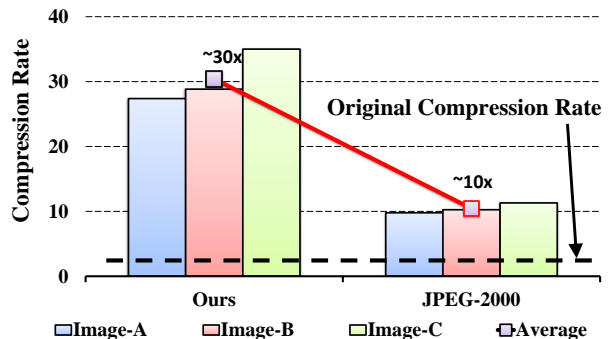

Figure 6. Compression rate comparison of our method v.s. JPEG-2000 under the same segmentation accuracy.

References

[1] OpenJPEG JPEG 2000 compression library, http://www.openjpeg.org/. 6

[2] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transform. *IEEE Transactions on image processing*, 1(2):205–220, 1992. 4

[3] M. Boliek. Information technology jpeg 2000 image coding system: Extensions for three-dimensional data. *ISO/IEC 15444-10, ITU-T Rec. T.809*, 2002. 1, 3

[4] M. Boliek. Jpeg 2000 image coding system: Core coding system. *ISO/IEC*, 2002. 1, 3

[5] T. Bruylants, A. Munteanu, and P. Schelkens. Wavelet based volumetric medical image compression. *Signal processing: Image communication*, 31:112–133, 2015. 1, 3

[6] H. Chen, Q. Dou, L. Yu, J. Qin, and P.-A. Heng. Voxresnet: Deep voxelwise residual networks for brain segmentation from 3d mr images. *NeuroImage*, 2017. 1

[7] H. Chen, X. Qi, J.-Z. Cheng, P.-A. Heng, et al. Deep contextual networks for neuronal structure segmentation. In *AAAI*, pages 1167–1173, 2016. 1

[8] J. Chen, L. Yang, Y. Zhang, M. Alber, and D. Z. Chen. Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation. In *Advances in Neural Information Processing Systems*, pages 3036–3044, 2016. 1

[9] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 424–432. Springer, 2016. 1

[10] W. Coomans, R. B. Moraes, K. Hooghe, A. Duque, J. Galaro, M. Timmers, A. J. van Wijngaarden, M. Guenach, and J. Maes. Xg-fast: the 5th generation broadband. *IEEE Communications Magazine*, 53(12):83–88, 2015. 1

[11] I. D. Dinov. Volume and value of big healthcare data. *Journal of medical statistics and informatics*, 4, 2016. 1

[12] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P.-A. Heng. 3d deeply supervised network for automatic liver segmentation from ct volumes. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 149–157. Springer, 2016. 1, 3, 6

[13] J. Gerrity. Health networks - delivering the future of healthcare, 2014. https://www.buildingbetterhealthcare.co.uk/technical/article_page/Comment_Health_networks__delivering_the_future_of_healthcare/94931. 1

[14] I. ITU-T and I. JTC. Generic coding of moving pictures and associated audio information-part 2: video, 1995. 1, 3

[15] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014. 6

[16] W. Jiang, X. Zhang, E. H.-M. Sha, L. Yang, Q. Zhuge, Y. Shi, and J. Hu. Accuracy vs. efficiency: Achieving both through fpga-implementation aware neural architecture search. *arXiv preprint arXiv:1901.11211*, 2019. 1

[17] W. Jiang, X. Zhang, E. H.-M. Sha, Q. Zhuge, L. Yang, Y. Shi, and J. Hu. Xfer: A novel design to achieve super-linear performance on multiple fpgas for real-time ai. In *Proceedings of the 2019 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*, pages 305–305. ACM, 2019. 1

[18] Y. Kang, J. Hauswald, C. Gao, A. Rovinski, T. Mudge, J. Mars, and L. Tang. Neurosurgeon: Collaborative intelligence between the cloud and mobile edge. *ACM SIGPLAN Notices*, 52(4):615–629, 2017. 1

[19] J. Li and R. M. Gray. Text and picture segmentation by the distribution analysis of wavelet coefficients. In *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*, pages 790–794. IEEE, 1998. 4

[20] L. F. Lucas, N. M. Rodrigues, L. A. da Silva Cruz, and S. M. de Faria. Lossless compression of medical images using 3-d predictors. *IEEE transactions on medical imaging*, 36(11):2250–2260, 2017. 1, 3

[21] Y. Ma and Z. Jia. Evolution and trends of broadband access technologies and fiber-wireless systems. In *Fiber-Wireless Convergence in Next-Generation Communication Networks*, pages 43–75. Springer, 2017. 1

[22] M. Marwan, A. Kartit, and H. Ouahmane. Using cloud solution for medical image processing: Issues and implementation efforts. In *Cloud Computing Technologies and Applications (CloudTech), 2017 3rd International Conference of*, pages 1–7. IEEE, 2017. 1

[23] J. Matela et al. Gpu-based dwt acceleration for jpeg2000. In *Annual Doctoral Workshop on Mathematical and Engineering Methods in Computer Science*, pages 136–143, 2009. 1, 8

[24] F. Milletari, N. Navab, and S.-A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *3D Vision (3DV), 2016 Fourth International Conference on*, pages 565–571. IEEE, 2016. 1

[25] R. Molla. Fixed broadband speeds are getting faster whats fastest in your city?, 2017. 1

[26] D. F. Pace, A. V. Dalca, T. Geva, A. J. Powell, M. H. Moghari, and P. Golland. Interactive whole-heart segmentation in congenital heart disease. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 80–88. Springer, 2015. 2, 3, 6

[27] B. Penna, T. Tillo, E. Magli, and G. Olmo. Progressive 3-d coding of hyperspectral images based on jpeg 2000. *IEEE Geoscience and remote sensing letters*, 3(1):125–129, 2006. 4

[28] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241. Springer, 2015. 1

[29] V. Sanchez, R. Abugharbieh, and P. Nasiopoulos. Symmetry-based scalable lossless compression of 3d medical image data. *IEEE Transactions on Medical Imaging*, 28(7):1062–1072, 2009. 1, 3

[30] V. Sanchez, R. Abugharbieh, and P. Nasiopoulos. 3-d scalable medical image compression with optimized volume of interest coding. *IEEE Transactions on Medical Imaging*, 29(10):1808–1820, 2010. 1, 3

[31] J. M. Santos, A. F. Guarda, N. M. Rodrigues, and S. M. Faria. Contributions to lossless coding of medical images

using minimum rate predictors. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 2935–2939. IEEE, 2015. 1, 3

[32] J. W. Schwartz and R. C. Barker. Bit-plane encoding: a technique for source encoding. *IEEE Transactions on Aerospace and Electronic Systems*, (4):385–392, 1966. 4

[33] M. J. Shensa. The discrete wavelet transform: wedding the a trous and mallat algorithms. *IEEE Transactions on signal processing*, 40(10):2464–2482, 1992. 4

[34] A. Skodras, C. Christopoulos, and T. Ebrahimi. The jpeg 2000 still image compression standard. *IEEE Signal processing magazine*, 18(5):36–58, 2001. 4

[35] M. C. Stamm and K. R. Liu. Anti-forensics of digital image compression. *IEEE Transactions on Information Forensics and Security*, 6(3):1050–1065, 2011. 4

[36] D. Taubman. High performance scalable image compression with ebcot. *IEEE Transactions on image processing*, 9(7):1158–1170, 2000. 4

[37] I. Telecom et al. Advanced video coding for generic audiovisual services. *ITU-T Recommendation H. 264*, 2003. 3

[38] G. K. Wallace. The jpeg still picture compression standard. *IEEE transactions on consumer electronics*, 38(1):xviii–xxxiv, 1992. 1, 3

[39] T. Wang, J. Xiong, X. Xu, and Y. Shi. Scnn: A general distribution based statistical convolutional neural networkwith application to video object detection. In *The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI19)*, 2019. 1

[40] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum. Dilated convolutional neural networks for cardiovascular mr segmentation in congenital heart disease. In *Reconstruction, Segmentation, and Analysis of Medical Images*, pages 95–102. Springer, 2016. 7

[41] X. Xu, Y. Ding, S. X. Hu, M. Niemier, J. Cong, Y. Hu, and Y. Shi. Scaling for edge inference of deep neural networks. *Nature Electronics*, 1(4):216, 2018. 1

[42] X. Xu, F. Lin, A. Wang, X. Yao, Q. Lu, W. Xu, Y. Shi, and Y. Hu. Accelerating dynamic time warping with memristor-based customized fabrics. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 37(4):729–741, 2018. 1

[43] X. Xu, F. Lin, W. Xu, X. Yao, Y. Shi, D. Zeng, and Y. Hu. Mda: A reconfigurable memristor-based distance accelerator for time series mining on data centers. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2018. 1

[44] X. Xu, Q. Lu, T. Wang, Y. Hu, C. Zhuo, J. Liu, and Y. Shi. Efficient hardware implementation of cellular neural networks with incremental quantization and early exit. *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, 14(4):48, 2018. 1

[45] X. Xu, Q. Lu, T. Wang, J. Liu, C. Zhuo, X. S. Hu, and Y. Shi. Edge segmentation: Empowering mobile telemedicine with compressed cellular neural networks. In *Proceedings of the 36th International Conference on Computer-Aided Design*, pages 880–887. IEEE Press, 2017. 1

[46] X. Xu, Q. Lu, L. Yang, S. Hu, D. Chen, Y. Hu, and Y. Shi. Quantization of fully convolutional networks for accurate biomedical image segmentation. *Preprint at https://arxiv. org/abs/1803.04907*, 2018. 1

[47] X. Xu, T. Wang, Q. Lu, and Y. Shi. Resource constrained cellular neural networks for real-time obstacle detection using fpgas. In *2018 19th International Symposium on Quality Electronic Design (ISQED)*, pages 437–440. IEEE, 2018. 1

[48] X. Xu, D. Zeng, W. Xu, Y. Shi, and Y. Hu. An efficient memristor-based distance accelerator for time series data mining on data centers. In *2017 54th ACM/EDAC/IEEE Design Automation Conference (DAC)*, pages 1–6. IEEE, 2017. 1

[49] Z. Xu, J. Bartrina-Rapesta, I. Blanes, V. Sanchez, J. Serra-Sagristà, M. García-Bach, and J. F. Muñoz. Diagnostically lossless coding of x-ray angiography images based on background suppression. *Computers & Electrical Engineering*, 53:319–332, 2016. 1, 3

[50] L. Yu, J.-Z. Cheng, Q. Dou, X. Yang, H. Chen, J. Qin, and P.-A. Heng. Automatic 3d cardiovascular mr segmentation with densely-connected volumetric convnets. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 287–295. Springer, 2017. 2, 3, 6, 7

[51] T. Zhao, R. J. Taylor, G. Li, S. Hu, and J. Wu. Cloud-based medical image processing system with anonymous data upload and download, Oct. 8 2013. US Patent 8,553,965. 1

[52] T. Zhao, R. J. Taylor, G. Li, J. Wu, and C. Jia. Cloud-based medical image processing system with access control, Mar. 25 2014. US Patent 8,682,049. 1

Supplementary Material: Machine Vision Guided 3D Medical Image Compression for Efficient Transmission and Accurate Segmentation in the Clouds

Zihao Liu[1], Xiaowei Xu[2], Tao Liu[1], Qi Liu[1], Yanzhi Wang[3], Yiyu Shi[2], Wujie Wen[1],

Meiping Huang[4], Haiyun Yuan[4], Jian Zhuang[4]

[1] Flordia International University, [2] University of Notre Dame, [3] Northeastern University,

[4] Guangdong General Hospital

{zliu021,tliu023,qliu020,wwen}@fiu.edu, {xxu8,yshi4}@nd.edu,yanz.wang@northeastern.edu

In this supplementary material, we introduce four mapping functions and perform performance comparison in Section 1. In Section 2, we present more results with compression rate of 40. In Section 3, we then provide the image by image segmentation results to further demonstrate that the improvement by our method is statistically significant.

1 Mapping Function Exploration

We explored four types of mapping functions–reciprocal mapping (RM) (depicted as 'ours' in our paper), linear mapping (LM), discrete mapping (DM) and power mapping (PM), to search the best mapping between statistic index (SI) and quantization step (QS). Then we compared segmentation accuracy and compression rates using these mapping methods in Section 1.4.

1.1 Linear Mapping

We first assumed the relationship between SI and QS obeyed a linear mapping (LM) function:

$$Q_n = a \cdot \delta_n + b, \quad s.t. \quad Q_{min} \leq Q_n \leq Q_{max}, \tag{1}$$

where Q_n is the quantization step at subband n, Q_{min} and Q_{max} are the smallest and largest QS, and a and b are the fitting parameters.

1.2 Discrete Mapping

We further developed a seemingly more intuitive discrete mapping (DM) method defined as:

$$Q_n = \begin{cases} Q_{min} & \delta_n > T_1 \\ 2 \cdot Q_{min} & T_2 < \delta_n \leq T_1 \\ 4 \cdot Q_{min} & T_3 < \delta_n \leq T_2 \\ Q_{max} & \delta_n \leq T_3 \end{cases} \tag{2}$$

where T_1, T_2 and T_3 indicate the average intervals between the largest and smallest δ_n.

Table 1: Segmentation results of four candidate mapping methods (LM, DM, PM and RM (depicted as "ours" in our paper)) using DenseVoxNet and HVSMR2016 dataset.

		LM	DM	PM	RM (Ours)
Myocardium	**Dice**	0.771±0.0438	0.821±0.043	0.834±0.0366	**0.834±0.0386**
	Hausdorff	36.823±10.079	32.372±8.551	31.012±7.911	**31±7.940**
	ASD	0.822±0.846	0.669±0.761	**0.651±0.724**	0.652±0.671
Blood Pool	**Dice**	0.903±0.026	0.911±0.027	0.913±0.026	**0.914±0.024**
	Hausdorff	43.007±11.358	41.792±9.796	41.014±9.769	**40.93±9.52**
	ASD	0.664±0.514	0.576±0.439	0.556±0.453	**0.556±0.432**
Compression Rate		**∼40x**	**∼35x**	**∼22x**	**∼30x**

1.3 Power Mapping

We adopted reciprocal mapping function in our paper (as Ours) to get QS for each subband from corresponding SIs. We also considered another non-linear mapping function, power mapping (PM),

$$Q_n = a \cdot \delta_n^b, \tag{3}$$

where a and b are fitting parameters. This mapping methods introduced more non-linearity between Q_n and δ_n.

1.4 Results Comparison

We compared segmentation accuracy and compression rate of these three mapping methods with the method in our paper. The experiment setup is the same as that in our paper for fair comparison. As shown in Table 3, we can observe that non-linear mapping methods (DM, PM and RM (ours)) can always achieve significant higher segmentation accuracy than linear mapping (LM) at all three segmentation measurement metrics, which demonstrates the advantage of the non-linear relationship between SI and QS. For the four mapping methods, it can be noticed that usually high accuracy corresponds to low compression rate except for RM (Ours). *This clearly indicates that our adopted RM is the best option to balance the accuracy and compression rate, e.g. offering almost the best accuracy at a relatively high compression rate.*

Figure 1: Mapping method comparison for result illustration.

The above phenomenon can be further explained with the illustration in Figure 1. The quantization steps in LM are always larger than non-linear mapping methods (PM and RM) at different SIs, thus resulting with a higher compression rate but a lower segmentation accuracy due to the more significant distortions of important features to DNN. Compared with RM (ours), PM has a steeper slope and a smaller QSs, thus decreasing compression rate (22× v.s. 30×). However, as we can see from Table 3, the segmentation accuracy is not improved w.r.t. RM. This is because the additional information saved by PM has almost no contributions to DNN segmentation, i.e. redundant data. Although DM is similar with RM in terms of the magnitude of quantization steps (QS),
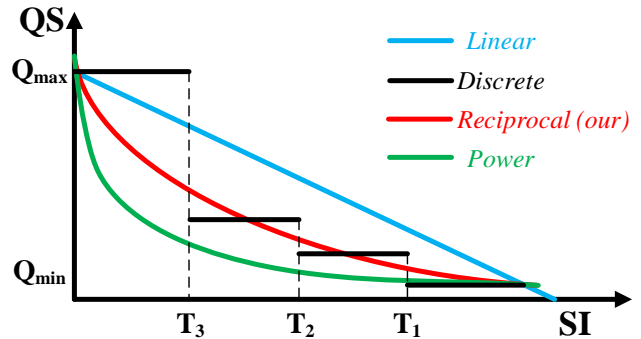
2

Table 2: Segmentation results of our method and JPEG-2000 using DenseVoxNet and HVSMR2016 dataset at different compression rates (30× and 40×).

		Ours	JPEG2000	Ours	JPEG2000
Myocardium	Dice	**0.834±0.0386**	0.816±0.042	0.808±0.043	0.802±0.039
	Hausdorff	**31±7.940**	33.513±7.566	34.112±8.551	34.506±7.962
	ASD	**0.652±0.671**	0.722±0.746	0.767±0.791	0.802±0.824
Blood Pool	Dice	**0.914±0.024**	0.912±0.024	0.910±0.023	0.909±0.025
	Hausdorff	**40.93±9.52**	41.031±9.648	41.192±9.706	41.448±9.671
	ASD	**0.556±0.432**	0.564±0.414	0.576±0.439	0.582±0.453
Compression Rate		∼30x	∼30x	∼40x	∼40x

it only exploits QS with coarse-grained intervals. Therefore, DM can achieve a higher compression rate but lower segmentation accuracy due to the non-optimized (coarse-grained) quantization steps.

Therefore, we finally select RM for our compression framework which can achieve a high compression rate with the best segmentation accuracy.

2 Results with Different Compression Rates

The comparison of our method and JPEG-2000 with different compression rates is shown in Table 2. At 30× compression rate, our method has remarkable segmentation accuracy improvement compared with JPEG-2000. However, at 40× compression rate, the segmentation accuracy is improved. One reasonable explanation is that too aggressive compression will cause serious information loss resulting with low performance. When the compression rate is 30x, the necessary information for segmentation is still retained. However, when it comes to 40x, this information is seriously dropped which causes a large accuracy degradation. This phenomenon is common in image compression, and compression methods can only drop unnecessary information within their capacities for particular domains. Otherwise, performance loss will always happen.

3 Image by Image Segmentation Results Comparison

To better illustrate that our improvement is statistically significant, we further report the image by image segmentation results of the HVSMAR dataset by using our method and JPEG-2000 under two different DNN structures–DenseVoxNet and 3D-DSN. Table 3 and 4 clearly show that our method wins in almost every case w.r.t. JPEG2000 for Dice, Hausdorff and ASD characterized from various Myocardoum and Blood Pool images under the same compression rate for both networks. Moreover, compared with the original uncompressed images, our method always offers the similar or even better results due to filtering the unnecessary image features that can mislead the segmentation.

Table 3: Image by image segmentation results of original (uncompressed), our method, JPEG-2000 using DenseVoxNet and HVSMR2016 dataset. The compression rate of our method and JPEG-2000 is 30. The images compressed by ours can be segmented with almost the same accuracy, or sometimes even better than original ones, significantly beating those compressed by JPEG-2000.

	Myocardoum								
	Dice			Hausdorff			ASD		
	Original	Ours	JPEG-2000	Original	Ours	JPEG-2000	Original	Ours	JPEG-2000
Img. 1	0.876	**0.876**	0.850	19.647	**19.197**	22.782	0.344	**0.327**	0.380
Img. 2	0.786	**0.774**	0.755	37.229	**37.815**	39.434	1.426	**1.453**	1.53
Img. 3	0.849	**0.855**	0.848	32.419	**31.516**	31.718	1.371	**1.3**	1.526
Img. 4	0.851	**0.846**	0.839	27.258	**28.083**	29	0.122	**0.089**	0.083
Img. 5	0.831	**0.818**	0.791	37.842	**38.615**	41.109	0.099	**0.088**	0.088
	Blood Pool								
	Dice			Hausdorff			ASD		
	Original	Ours	JPEG-2000	Original	Ours	JPEG-2000	Original	Ours	JPEG-2000
Img. 1	0.898	**0.9**	0.897	25.278	**24.434**	24.515	1.316	**1.247**	1.226
Img. 2	0.9	**0.9**	0.898	45.200	**44.878**	44.922	0.323	**0.311**	0.341
Img. 3	0.953	**0.954**	0.953	41.219	**41.677**	41.158	0.552	**0.532**	0.553
Img. 4	0.895	**0.897**	0.895	49.669	**48.363**	48.518	0.683	**0.62**	0.649
Img. 5	0.927	**0.919**	0.911	43.806	**45.31**	46.776	0.127	**0.11**	0.139

Table 4: Image by image segmentation results of original (uncompressed), our method, JPEG-2000 using 3D-DSN and HVSMR 2016 dataset. The compression rate of our method and JPEG-2000 is 30. The images compressed by ours can be segmented with almost the same accuracy, or sometimes even better than original ones, significantly beating those compressed by JPEG-2000.

	Myocardoum								
	Dice			Hausdorff			ASD		
	Original	Ours	JPEG-2000	Original	Ours	JPEG-2000	Original	Ours	JPEG-2000
Img. 1	0.829	**0.829**	0.801	20.273	**20.149**	22.045	0.446	**0.476**	0.607
Img. 2	0.688	**0.692**	0.703	39.712	**38.652**	38.026	0.401	**0.386**	0.39
Img. 3	0.799	**0.802**	0.781	35.567	**34.929**	35.185	0.451	**0.5**	0.526
Img. 4	0.834	**0.839**	0.836	25.199	**24.880**	25.160	0.093	**0.091**	0.088
Img. 5	0.773	**0.769**	0.748	40.976	**40.902**	44.744	0.157	**0.171**	0.164
	Blood Pool								
	Dice			Hausdorff			ASD		
	Original	Ours	JPEG-2000	Original	Ours	JPEG-2000	Original	Ours	JPEG-2000
Img. 1	0.903	**0.905**	0.896	23.324	**22.782**	25.239	0.568	**0.559**	0.57
Img. 2	0.873	**0.866**	0.863	45.837	**46.551**	46.755	0.229	**0.235**	0.239
Img. 3	0.949	**0.950**	0.949	40.410	**40.386**	40.841	0.221	**0.179**	0.197
Img. 4	0.915	**0.913**	0.910	35.931	**36.905**	38.730	0.88	**0.076**	0.079
Img. 5	0.906	**0.900**	0.883	47.074	**47.403**	51.254	0.068	**0.066**	0.067