2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Automatic Deep Learning Segmentation and Quantification of Epicardial Adipose Tissue in Non-Contrast Cardiac CT scans

Ammar Hoori, *Student Member, IEEE,* Tao Hu, Sadeer Al-Kindi, Sanjay Rajagopalan, David L. Wilson, *Live Member, IEEE*

***Abstract*—An Automatic deep learning semantic segmentation (ADLS) using DeepLab-v3-plus technique is proposed for a full and accurate whole heart Epicardial adipose tissue (EAT) segmentation from non-contrast cardiac CT scan. The ADLS algorithm was trained on manual segmented scans of the enclosed region of the pericardium (sac), which represents the internal heart tissues where the EAT is located. A level of 40 Hounsfield unit (HU) and a window of 350 HU was applied to every axial slice for contrast enhancement. Each slice was associated with two additional consecutive slices, representing the three-channel single input image of the deep network. The detected output mask region, as a post-step, was thresholded between [-190, -30] HU to detect the EAT region. A median filter with kernel size 3mm was applied to remove the noise. Using 70 CT scans (50 training/20 testing), the ADLS showed excellent results compared to manual segmentation (ground truth). The total average Dice score was (89.31%±1.96) with a high correlation of (R=97.15%, p-value <0.001), while the average error of EAT volume was (0.79±9.21).**

***Clinical Relevance*— Epicardial adipose tissue (EAT) volume aids in predicting atherosclerosis development and is linked to major adverse cardiac events. However, accurate manual segmentation is considered tedious work and requires skilled expertise.**

I. Introduction

Epicardial adipose tissue (EAT) is a visceral fat deposit distributed between the pericardium and the heart [1], [2]. Epicardial fat tissue can be detected by CT and can be broadly defined as tissue within the pericardial sac with signal intensity in the fat tissue range (-190 to -130 HU). With careful detection, cardiologists and radiologists can distinguish the pericardium, the sac that enclosed the heart, in non-enhanced CT images and identify the EAT volume by thresholding the internal heart region. However, manual segmentation for EAT in a non-contrast-enhanced CT is a time-consuming task even for expert radiologists. Pericardial tissues are often difficult to distinguish in cardiac CT scans due to the thin layer of the pericardium and the low contrast from surrounding tissues and blood [3].

There are previous reports using deep learning. Commandeur et. al [4], proposed a fully automated approach with two Convolutional Neural Networks (CNNs) to segment EAT and thoracic adipose tissue (TAT). The first CNN detects the heart limits and performs segmentations while the second combines a statistical shape model to detect the pericardium. The study showed a Dice score of 82.3% with a correlation coefficient for automatic vs. manual R=92.4%, for 250 non-contact CT images. Commandeur et. al [5], presented a fully automatic CT multicenter study to segment and quantify EAT using CNN. The proposed method uses a single CNN in two tasks: slice classification and slice segmentation. The classification was meant to recognize whether the slice belongs to the heart or not, while the EAT is segmented in the second task. The method uses three slices as an input patch: the designated slice (k), one prior $(k-1)$, and one post $(k+1)$, while the output is the manual EAT labeling for the middle slice (k). With a dataset of 614 CT non-contrast scans, the method achieved a Dice score of 87.3%, while the correlation coefficient was R=97.4%. Motivated by these three consecutive slice approaches, we concatenate our input patches in three slices, the designated one (k), with two followed slices $(k+1, k+2)$, while the output is the enclosed area of the labeled sac and internal heart tissues of the designated slice (k), preventing a scenario where there is no segmented fat in the first slice of the CT volume. He et. al [6], proposed a 3D deep attention U-Net for segmenting the EAT for 40 CTA images. The high number of slices (with thin slice thickness in CTA images) with an input size of 512×512×32 made it possible to use a 3D deep network, which is not the case in a thick slice non-contrast CT, as in our experiments. However, the method showed a Dice score of 85%. He et. al [7], extended their 3D deep attention U-Net study for CTA images to include 200 patients, which showed an improved Dice score of 88.7%.

In the manual segmentation process, a preprocessing step is performed for all axial slices to prepare the input patches and enhance the training performance. A window/level of [level =40HU, window=350HU] and noise median filter, are applied to each raw CT volume to enhance the contrast of the

Ammar Hoori, is with the Department of Biomedical Engineering (BME), Case Western Reserve University (CWRU), Cleveland, OH 44106 USA (e-mail: ammar.hoori@case.edu).

Tao Hu is with the BME, CWRU, Cleveland, OH 44106 USA (e-mail: txh272@case.edu).

Sadeer Al-Kindi, is with Department of Cardiology, University Hospital, Cleveland, OH, 44106, USA (e-mail: Sadeer.Al-Kindi@uhhospitals.org).

Sanjay Rajagopalan, is with Department of Cardiology, University Hospital, Cleveland, OH, 44106, USA (e-mail: sanjay.rajagopalan@uhhospitals.org).

David L. Wilson is with the BME, CWRU, Cleveland, OH 44106 USA and with Department of Radiology, Case Western Reserve University, Cleveland, OH, 44106, USA (e-mail: david.wilson@case.edu).

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

pericardium tissues. Each 2D slice input (k) is associated with its consecutive two slices ($k+1, k+2$), forming a 512×512×3 voxels input patch while the output is the 2D manually labeled mask of the k slice only.

In this paper, we investigate multiple potential improvements to EAT segmentation. We used window-level to preprocess the data and draw attention to the range of CT numbers of interest. In experiments, we compared results with and without the windowing operation. We investigated the role of multiple types of data augmentations (e.g., blur to mimic motion, rotation, and scaling to estimate different patients' characteristics). Image volumes are 512×512 pixel axial images with 40 to 60 image slices. We used 512×512×3 "slab" patches as the inputs input to the deep network. The outputs of the deep learning network are binary masks classifying each pixel either inside or outside the pericardium region as shown in Fig. 1. We also investigated different deep learning techniques (e.g., U-net and SegNet) and compared results. Corresponding EAT volume, Dice score coefficient, Intersection Over Union (IOU) score are calculated to evaluate the segmentation performance. Scatter and Bland-Altman plots with correlation coefficient (R) are presented to understand the association between manual vs. automatic segmentations per patient and per slice.

The experiments involving human subjects described in this paper were approved by the University Hospitals Institutional Review Board.

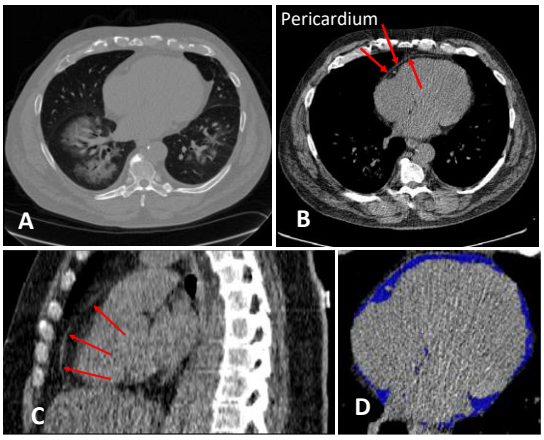

Fig. 1. The expert manual segmentation process of the EAT is illustrated in a step-by-step example. Each 2D axial slice is examined from expert reader as in (A). A window with a level of 40HU and a width of 350 HU was firstly used for each input image to increase the contrast of the pericardium (B). The pericardium (SAC) can be noticed in the inferior region from axial view and sometimes from sagittal view (C). Expert reader draws a contours to distinguish the pericardium and the EAT tissues are recognized as the interior voxels that has fat tissues located inside the contour with thresholding of [-190, -30] as in (D).

II. The Proposed Method

Similar to manual segmentation, preprocessing steps are performed to prepare input patches for the training process and enhance the performance of the network, as shown in Fig. 2. A window/level with a level of 40 HU and a window of 350 HU is firstly applied for each input image to increase the contrast of the pericardium, encouraging the deep learning network to easily capture common contrasted pericardium structural features. Each k labeled image is concatenated with its consecutive two slices ($k+1$ and $k+2$) to generate a

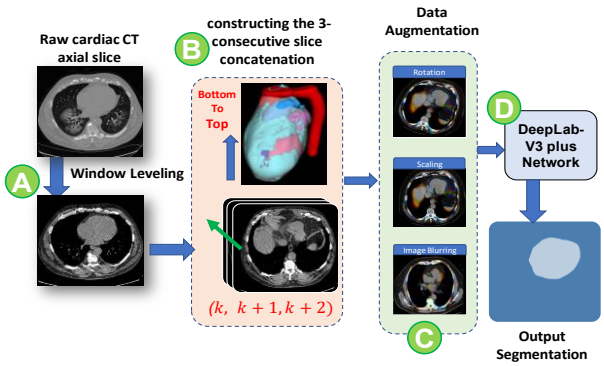

Fig. 2. The full structure of the automatic EAT segmentation training process and the preprocessing steps are illustrated. A window leveling of level of 40 HU and window of 350 HU is shown in (A), then concatenating each slice k with the following ($k+1$, $k+2$) slices to represents the three channels of the training input patches for the deep network, as in (B). Different data augmentations are meant to enrich the deep learning with variations of cases, is also shown in (C). Finally, the DeepLab-v3 plus network is trained with each three sequenced patches with single corresponding mask slice as in (D).

512×512×3 input voxel patch. This concatenation process, on one hand, enriches the surrounding pericardium structures to assist the deep learning network in identifying common features of the sac. On the other hand, this process trains the deep network on the ascending curvature of the sac (as seen in the sagittal view). Intuitively, promoting a unified phenomenon in the input images, hence improving the training performance, a 3-consecutive method is applied to maintain the curvature of the sac. The output mask is the mask of the first slice of each three-consecutive input patch, which we labeled as the entire internal region of the heart tissues beyond the pericardium (sac) contour. Although the manual segmentation considered the sac as the reference to distinguish the EAT from other heart surrounding fats, the deep learning segmentation performs more accurately with larger regions. Sac contours alone represent an imbalanced ratio when it comes to the number of labeled vs. non-labeled voxels. Using the "beyond sac" region enriches the deep learning with a balanced number of voxels, hence improves training. The structure of the preprocessing steps and the full automatic training structure is presented in Fig. 2.

We implemented our deep network model by using transfer learning based on the DeepLab-v3 plus [8] (Resnet-18 as a backbone). The deep network model is a CNN specifically designed for semantic segmentation tasks, which is mainly composed of several important architectures: the backbone network, the Atrous convolution, the Atrous Spatial Pyramid Pooling (ASPP) network, and the decoder section. Traditional deep CNN has a tendency to reduce the spatial resolution of the output feature map as the network goes deeper, and thus is not suitable for semantic segmentation tasks, which require detailed spatial information. In contrast to CNN, the DeepLab-v3 plus applied Atrous convolution, which can adjust the effective field of view for convolution without reducing the size of the output feature map, in the last few blocks of the backbone network. Thus, Atrous convolution can extract denser features at multiple scales while preserving the spatial resolution, which is significant for semantic segmentation. The ASPP is used on the top of the feature map to capture multi-scale object information by applying four parallel

Atrous convolutions with different sampling rates. Batch normalization and image-level features are also incorporated in the ASPP by applying a global average pooling at the last feature map of the backbone and concatenating the corresponding results (contains multi-scale features) with batch normalization [8]. The results are then traversed through a 1×1 convolution with 256 filters to get the final output. Furthermore, a decoder section has been added to gradually recover the spatial information, hence capturing more detailed boundary features, by applying a few 3×3 convolutions to refine the output features obtained from the ASPP with an upsampling factor of 4 [8]. As a post-processing step, the output of the deep network is used as a mask to detect the EAT region. Similar to manual segmentation, we apply noise reduction (3×3×3 median) to reduce artefacts in these low-dose images. We then apply standard fat thresholding [-190 HU, -30 HU] to identify EAT volume within the pericardial sac. The complete preprocessing, augmentation, and training are presented in Fig. 2.

For the training setting, similar to a prior approach [5], the Adam method is used for optimization and the Dice loss is applied as the loss function to maximize the Dice score coefficient. The Dice loss is the cost function used to evaluate the similarity between the resulted mask and the ground truth mask. To enrich the training process, random augmentations are applied during training. Random rotation (-5 to 5 degrees), scaling (0.9 to 1.1), and Gaussian blurring with a standard deviation set to ($\sigma < 2$) are applied for data augmentation. We use a mini-batch strategy with a batch size of 20, while the max number of epochs is set to 30 and the initial learning rate is set to 1e-3. Validation is performed at the end of each epoch. Our proposed deep learning method is further investigated with different experiments as will be illustrated in the next section.

III. Experimental Results

A. Dataset and Segmentation Metrics

This study utilized 70 non-contrast cardiac CT scans done for calcium scoring, which was obtained from the University Hospital of Cleveland. The axial slice thickness was 2.5 mm and the 2D slice dimensions were 512×512 pixels per axial slice. A total of 2291 axial slices were used in this study. The dataset was first divided into two subsets: training and testing subjects (50 training/20 testing). Subjects for training were then divided into two subsets: training subset (80%) and validation subset (20%).

The Dice score coefficient was calculated between the automatic method testing output and the ground truth (manual segmentation) to evaluate the performance of the semantic segmentation.

$$Dice(X,Y) = \frac{2 * |X \cap Y|}{|X| + |Y|}, \quad (1)$$

where, X and Y represent the testing output and the ground truth pixels in a slice (or voxels in a volume), $|X \cap Y|$ is the number of overlapping pixels (or voxels) between the predicted EAT segmentation and the ground truth EAT images, while $|X| + |Y|$ represents the total number of pixels (or voxels) in both images (or volumes). We also calculate the Intersection Over Union (IOU) score, also known as the Jaccard Index, which represents the ratio of the overlapped between the automatic and ground truth area region over all the area of the union region.

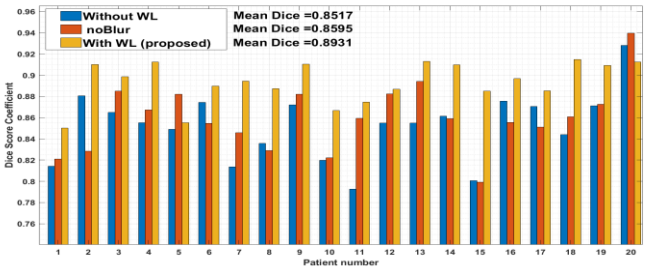

Fig. 3. A Dice score comparison is shown for the EAT segmentation of 20 patients. The proposed method was individually trained using either, without window leveling, without Blur, and with window leveling. The mean dice scores were (85.17%±3.2, 85.95%±3.14, and 89.31%±1.96), respectively. The window leveling provides a higher segmentation results in 18 out of 20 patients over the other two experiments.

$$IOU(X,Y) = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|}. \quad (2)$$

In addition to evaluating the Dice score coefficient and IOU, we compared the automatic EAT volumes collected by the automated method with manual reader segmentation for each test subject. Scatter and Bland-Altman plots were performed for every testing subject, as well as, every slice to evaluate the agreement between the predicted results and the ground truth. The correlation coefficient (R) and its corresponding p-value were calculated to assess the scatter plots, while a 95% confidence interval had been established within 1.96 times the standard deviation for the Bland-Altman analysis.

B. Results and Discussions

The deep learning experiments in this study were performed using a Windows 10 computer with an AMD Ryzen 7 5800X 3.8GHz, 32 GB RAM, 1TB hard disk, GTX 3090 with 24GB GPU. We implemented the code using Matlab 2021a. The manual segmentations were implemented in regular computers using Slicer 3D software, Version 4.111 and the results of manual labeled volumes are saved in DICOM files for an easy association with the original CT volumes.

To perform the automated EAT segmentation, we used transfer learning based on DeepLab-v3 plus. For comparison purposes, the input patches were organized in sequence. The heart input patches are sequentially fed from bottom to top with three consecutive slices for each patch. We investigated three experiments using our deep network model: without using window leveling (without WL), without using blur augmentation (noBlur), and with window leveling (WL, our proposed approach). Using 70 CT scans (50 training/20 testing), the deep network using window leveling showed superiority in Dice results compared to the network with other two experiments in 18 test cases out of the 20 test cases (90%), as shown in Fig. 3. The total average Dice scores per subject for the proposed approach (with window leveling) was 89.31% (ranged from 85.03% to 91.47%) while for the withoutWL approach

[1] https://www.slicer.org

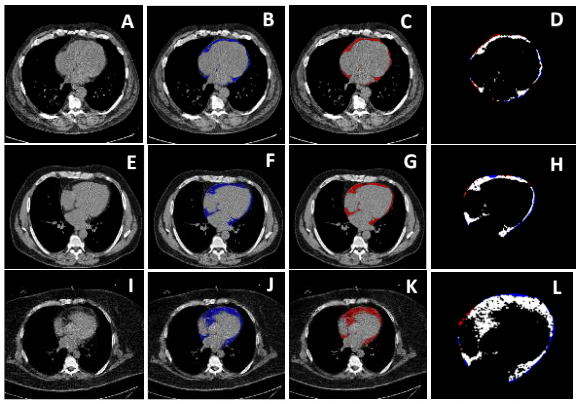

Fig. 4. Segmentation examples of three individuals, images show the axial non-contrast CT view of heart regions, as shown in A, E, and I. The manual EAT segmentations of those individuals are illustrated in blue labelling as in B, F, and I, respectively. The proposed automatic deep learning segmentation results are illustrated denoted with blue in C, G, and K, respectively. Finally, combined regions of manual and automatic labelling are shown in D, H and L, respectively, where red represents manual area, blue represents automatic, and white is the overlapped area.

was 85.17% (ranged from 79.28% to 92.81%), and for the no-Blur approach was 85.95% (ranged from 79.91% to 93.95%). Using the same deep learning structure (DeepLab-v3 plus), keeping the same preprocessing steps, and same testing set, while changing the way of representing the input data, the Dice score results tipped the scale of the window leveling experiments over others. The idea of maintaining the increasing curvature of the three consecutive slices for each input gave the deep network more ease to intelligently detect the sac boundaries for the associated slices. Looking at a special case-patient No. 15, the Dice score for without window leveling was 80% while for noBlur it was 79.8%. However, using window leveling improved the Dice score up to 88.2%, where we can see the effect of window leveling in emphasizing the sac structure for the deep network segmentation.

The deep network is capable of training colorful images (three channels of RGB), however, the CT images are gray-scale images (single channel). We embedded three-consecutive slices in the deep network input slab instead of repeating the same slice in the three channels. This approach provides the deep network with more structure of the sac region. Compared to manual segmentation, experts look into adjacent slices when the current slice's sac is ambiguous. We picked the sequence of the adjacent slices to be the two consecutive ($k+1$, and $k+2$) ones to the labeled one (k), as we start from the bottom of the heart where the first slice has only consecutive slices that have sac.

Fig. 4. illustrates three test individuals of manual vs. automatic (proposed) EAT segmentation results for CT axial slice heart region. The three initially unlabeled slices are shown in A, E, and I, respectively. The manual expert labeled slices are shown in blue in B, F, and J, respectively. The automatically labeled slices are shown red in C, G, and H, respectively. While the manual and automatic are combined in D, H, and L, respectively, with white as the agreement (overlapping) area, blue as the manual only, and red as the automatic only. These subfigures show that the automatic method presents a high agreement with the manual labeling for the untrained test samples, except for some very narrow edges which we strongly believe that expanding the training dataset will much improve the training of our model.

Total individual EAT volume assessment is considered the most important outcome that clinicians look for in EAT evaluation. In this paper, we provided two helpful assessments for the EAT volumes: total per patient and total per slice volume assessments. In a total patient's EAT volume assessment, each test full CT automatic EAT segmentation EAT volume is compared to its corresponding manual labeled using scatter plot and Bland-Altman plot, as shown in Fig. 5.A and Fig. 5.B, respectively. The scatter plot showed a correlation (R=97.15%, p-value <0.001). Likewise, the Bland-Altman plot provided a smaller bias of 0.32 cm^3 and narrower quartile range of [-23.85, 24.49] cm^3. These results are due to a small cohort, which included small to large EAT volume, the noise, and the unclear sac tissues in some slices. And they can be improved more with a larger cohort for the training subjects.

In EAT slice-based study, which - to our knowledge - none have investigated before this paper, provided a detailed per slice segmentation evaluation. Since the deep network is trying to learn the EAT per slice, this study shows the regions where the network suffers from in quartile grouped slices. We demonstrate the automatic results compared to manual test labels in scatter and Bland-Altman plots in Fig. 6.A and Fig. 6.B, respectively. The slices of each test case are categorized into four equal regions based on their location in the total heart slice sequence. These regions are presented in colors as: blue for slices in the bottom 25%, red for the slices between 25%-50%, yellow for the slices between 50%-75%, and blue for slices in the top 25%. The Scatter plot showed a correlation of (R=95.65%, p-value<0.001) while the Bland-Altman plot presented a bias of 0.01 cm^3 with range [-1.08, 1.1] cm^3. Both scatter and Bland-Altman results affirmed the superiority of our method in slice-grouped-wise. Moreover, the subdividing idea emphasized the fact that the top 25% and bottom 25% regional groups have the most deviations among all four regions in the traditional method, while middle regions (25%-75%) show fewer differences. In a region-based Dice score, the results were improved, from (82.27%±5.32, 86.79%±4.34, 88.21%±3.25, and 85.87%±4.7) for without using blur nor window-level to (85.82%±4.65, 88.64%±2.04, 91.87%±2.03, and 91.87%±2.8) using both blur and window-level, for slice regions of (bottom 25%, 25%-50%, 50%-75%, and top 25%), respectively.

We further investigated the use of different deep learning techniques with both with and without window leveling input arrangement, as shown in Table I. The study includes the training/validating on the same 50 CT scans and testing on the same 20 CT scans using U-Net [9](with CNN as internal blocks),

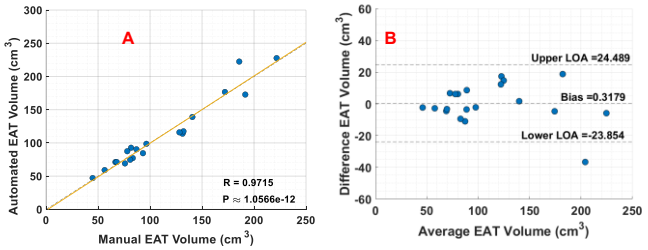

Fig. 5. Experiments plots of EAT volumetric evaluation for the testing subjects between manual and automatic methods per patient. The scatter plot of manual vs. automated method is shown in (A) while the Bland-Altman plot is shown in (B). The one outlier in (B) had an unusual automatic segmentation which could be easily identified and corrected. These differences are due to the large EAT volume, the noise, and the unclear sac tissues in some slices. And they can be improved with more training subjects.

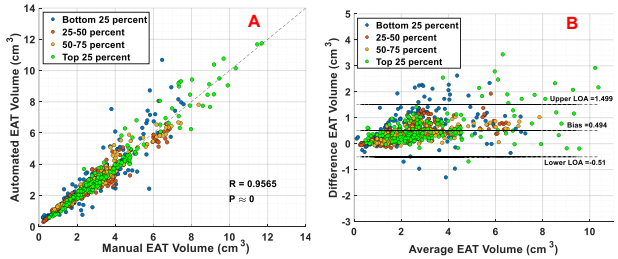

Fig. 6. Experiments scatter (A) and Bland-Altman (B) plots of EAT volumetric evaluation for the testing subjects between manual and automatic methods per slice. Slices are categorized into quartile groups based on their location in total heart slices and identified in colors as: blue for slices in the bottom 25%, red for the slices between 25%-50%, yellow for the slices between 50%-75%, and blue for slices in the top 25%. From plots the most deviations are in the top and bottom 25% slices while the middle 50% slices are more accurately segmented.

SegNet [10] (with VggNet-16 internal blocks), SegNet (with VggNet-19 internal blocks), and our proposed method, the DeepLab-v3 plus (using ResNet-18). We used the same three consecutive concatenations for all experiments. Our proposed method provides the highest average Dice score of (89.31%± 1.95), and the highest average IOU score of (80.74%±3.16), with the second least average, EAT volume error of (0.79%±9.21). In a general overview, one can also notice that all the window leveling-based deep network results are better than without window leveling-based results. This emphasized the upper hand of the proposed method associated with the suggested DeepLab-v3 plus deep network.

A comprehensive comparison with the recently reported results from different algorithms, cohorts, and imaging modalities is shown in Table II. Our proposed reverse model showed the best Dice score among comparable non-contrast CT modalities and the best correlation coefficient of (R=98.52%) among all methods. The only method that showed closer performance was presented by He et. al [7], which had 200 subjects with CTA modality. CTA has a thinner slice thickness, hence 5-times the total number of slices than CT, while CTA has a contrast agent that improves the detection of blood vessels. With a slice thickness of 2.5 mm, the inter-slice textural details are missing, while no-contrast CT makes it more difficult to segment the pericardium, however, our method performed as well as a CTA study. The correlation coefficient can be improved with larger cohorts, where R is less sensitive to individual errors.

IV. Conclusion

We conclude in this study that our window leveling 3-consecutive-slice-based method detects the sagittal curvature of the sac of the heart to support the axial slices, hence improving the deep learning in all presented deep network architectures. The use of window leveling in our method outperforms the traditional (without WL) method in (90%) of the testing set. Compared to alternative deep methods, the proposed method shows dice score improvement with averages of 22.19%, 1.13%, and 0.35% over U-net, SegNet(Vgg-16), and SegNet (Vgg-19), respectively. The scatter and Bland-Altman plots prove that the proposed method is highly correlated compared to manual segmentation with less bias and fewer outliers. In this work, even with relatively small sample size, our method outperformed other approaches. Indeed, our results were comparable to studies in larger cohorts and favorably compared with other studies using CT contrast and higher resolution.

TABLE I
Results Of Different Deep Segmentation Techniques Using The Same Training/Testing Cohort.

Deep Segmentation (Deep network)	Window leveling	Average Dice Score(%)	Average IOU Score (%)	Average EAT volume error (%)
U-Net (CNN)	Without	34.31±10.92	21.26±8.75	-53.52±12.28
U-Net (CNN)	With	67.12±5.58	50.76±6.22	-46.64±7.01
SegNet(VGG-16)	Without	86.16±2.66	75.78±4.07	2.03±11.56
SegNet(VGG-16)	With	88.18±2.20	78.93±3.58	9.18±9.45
SegNet(VGG-19)	Without	86.78±3.19	76.77±4.95	**-0.46±9.79**
SegNet(VGG-19)	With	88.96±2.04	80.17±3.24	1.92±7.20
DeepLab-v3 Plus (ResNet-18)	Without	85.17±3.20	74.30±4.88	13.94±10.05
DeepLab-v3 Plus (ResNet-18)	With	**89.31±1.96**	**80.74±3.16**	0.79±9.21

TABLE II
Results Of Different Segmentation Studies Using Different Cohorts With Different Modalities.

Study	Cohort population/ modality	Average Dice Score(%)	Correlation Coefficient R (%)
Commanduer [4]	250/ CT*	82.3	92.4
Commanduer [5]	614/ CT	87.3	**97.4**
He [6]	40 /CTA	85	-
He [7]	200/ CTA	88.7	94.9
Ours (with WL)	70/ CT	**89.31±1.96**	97.2

* All CT modalities in this table are thick slice non-contrast CT scans, while the CTA scans are thin slice with contrast agent for easy vessel detection and clinical functionality evaluation.

References

[1] N. Aslanabadi *et al.*, "Epicardial and Pericardial Fat Volume Correlate with the Severity of Coronary Artery Stenosis," *J. Cardiovasc. Thorac. Res.*, vol. 6, no. 4, pp. 235–239, Dec. 2014

[2] E. Nagy, A. L. Jermendy, B. Merkely, and P. Maurovich-Horvat, "Clinical importance of epicardial adipose tissue," *Arch. Med. Sci.*, vol. 4, pp. 864–874, 2017.

[3] Z. Zhao *et al.*, "Automatic Segmentation of Visible Epicardium Using Deep Learning in CT Image," in *Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery*, vol. 1074, Y. Liu, L. Wang, L. Zhao, and Z. Yu, Eds. Cham: Springer International Publishing, 2020, pp. 577–584.

[4] F. Commandeur *et al.*, "Deep Learning for Quantification of Epicardial and Thoracic Adipose Tissue From Non-Contrast CT," *IEEE Trans. Med. Imaging*, vol. 37, no. 8, pp. 1835–1846, Aug. 2018.

[5] F. Commandeur *et al.*, "Fully Automated CT Quantification of Epicardial Adipose Tissue by Deep Learning: A Multicenter Study," *Radiol. Artif. Intell.*, vol. 1, no. 6, p. e190045, Nov. 2019.

[6] X. He *et al.*, "Automatic epicardial fat segmentation in cardiac CT imaging using 3D deep attention U-Net," in *Medical Imaging 2020: Image Processing*, Houston, United States, Mar. 2020, p. 84.

[7] X. He *et al.*, "Automatic segmentation and quantification of epicardial adipose tissue from coronary computed tomography angiography," *Phys. Med. Biol.*, vol. 65, no. 9, p. 095012, May 2020.

[8] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," in *Computer Vision – ECCV 2018*, vol. 11211, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham: Springer International Publishing, 2018, pp. 833–851.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241.

[10] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.