A 3D deep learning approach to epicardial fat segmentation in non-contrast and post-contrast cardiac CT images

Thanongchai Siriapisith [Corresp., 1], **Worapan Kusakunniran** [2], **Peter Haddawy** [2, 3]

[1] Faculty of Medicine Siriraj Hospital, Mahidol University, Department of Radiology, Bangkok, Thailand

[2] Faculty of Information and Communication Technology, Mahidol University, Faculty of Information and Communication Technology, Nakhon Pathom, Thailand

[3] University of Bremen, Bremen Spatial Cognition Center, Bremen, Germany

Corresponding Author: Thanongchai Siriapisith
Email address: thanongchai.sir@mahidol.edu

Epicardial fat (ECF) is localized fat surrounding the heart muscle or myocardial and enclosed by a thin-layer membrane of pericardium. Segmenting the ECF is one of the most difficult medical image segmentation tasks. Since the epicardial fat is infiltrated into the groove between cardiac chambers and is contiguous with cardiac muscle, segmentation requires location and voxel intensity. Recently, deep learning methods have been effectively used to solve medical image segmentation problems in several domains with state-of-the-art performance. This paper presents a novel approach to 3D segmentation of ECF by integrating attention gates and deep supervision into the 3D U-Net deep learning architecture. The proposed method shows significant improvement of the segmentation performance, when compared with standard 3D U-Net. The experiments show excellent performance on non-contrast CT datasets with average Dice scores of 90.06%. Transfer learning from a pre-trained model of a non-contrast CT to contrast-enhanced CT dataset was also performed. The segmentation accuracy of contrast-enhanced CT dataset achieved Dice score of 88.16%.

Manuscript Title

A 3D deep learning approach to epicardial fat segmentation in non-contrast and post-contrast cardiac CT images

Thanongchai Siriapisith[1], Worapan Kusakunniran[2], Peter Haddawy[2,3],

[1] Department of Radiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand
[2] Faculty of Information and Communication Technology, Mahidol University, Nakhon Pathom, Thailand
[3] Bremen Spatial Cognition Center, University of Bremen, Germany

Corresponding Author:
Thanongchai Siriapisith[1]
Arun-Amarin, Bangkoknoi, Bangkok, 10700, Thailand
Email address: thanongchai.sir@mahidol.edu
Keyword: epicardial fat, computed tomography, 3D segmentation, deep learning, 3D U-Net, attention gate, deep supervision

Abstract

Epicardial fat (ECF) is localized fat surrounding the heart muscle or myocardial and enclosed by a thin-layer membrane of pericardium. Segmenting the ECF is one of the most difficult medical image segmentation tasks. Since the epicardial fat is infiltrated into the groove between cardiac chambers and is contiguous with cardiac muscle, segmentation requires location and voxel intensity. Recently, deep learning methods have been effectively used to solve medical image segmentation problems in several domains with state-of-the-art performance. This paper presents a novel approach to 3D segmentation of ECF by integrating attention gates and deep supervision into the 3D U-Net deep learning architecture. The proposed method shows significant improvement of the segmentation performance, when compared with standard 3D U-Net. The experiments show excellent performance on non-contrast CT datasets with average Dice scores of 90.06%. Transfer learning from a pre-trained model of a non-contrast CT to contrast-enhanced CT dataset was also performed. The segmentation accuracy on the contrast-enhanced CT dataset achieved a Dice score of 88.16%.

Introduction

Epicardial fat (ECF) is localized fat surrounding the heart muscle and enclosed by the thin-layer pericardium membrane. The adipose tissue located outside pericardium is called pericardial fat that is contiguous with other mediastinal fat (Fig 1). ECF is the source of pro-inflammatory mediators and promotes the development of atherosclerosis of coronary arteries. The clinical significance of the ECF volume lies in its relation to major adverse cardiovascular events. Thus, measuring its volume is important in diagnosis and prognosis of cardiac conditions. ECF volume can be measured in non-contrast CT images (NCCT) with coronary calcium scoring and in contrast-enhanced CT images (CECT) with coronary CT angiography (CCTA). However, accurate measurement of ECF is challenging. The ECF is separated from other mediastinal fat by thin layer pericardium. The pericardium is often not fully visible in CT images, which makes the detection of the boundaries of ECF difficult. ECF can also be infiltrated into grooves between cardiac chambers and is contiguous to the heart muscle. These technical challenges not only make accurate volume estimation difficult but make manual measurement a time-consuming process that is not practical in routine use. Therefore, computer-assisted tools are essential to reduce the processing time for ECF volume measurement.

Automated segmentation could potentially make ECF volume estimation more practical on a routine basis. Several approaches based on prior medical knowledge or non-machine learning techniques have been proposed for ECF segmentation, including genetic algorithms, region-of-interest selection with thresholding, and fuzzy c-mean clustering [1-3]. Deep learning techniques have been applied to a wide variety of medical image segmentation problems with great success [4-6]. A recent article [7] demonstrates that deep learning algorithms outperform conventional methods for medical image segmentation in terms of accuracy. But most previous studies involved large solid organs or tumor segmentation [8-10]. The segmentation of relatively

small and complex structures with high inter-patient variability, such as ECF, has been far less successful. Recently, a few deep learning approaches to ECF segmentation have made progress on this problem [11-13]. In this paper, we build upon the previous work by presenting a novel deep learning model for 3D segmentation of ECF.

~~In this paper, we~~ propose a solution of automatic segmentation of ECF volume using a deep learning based approach~~, in~~ non-contrast and contrast-enhanced CT datasets. The NCCT dataset uses coronary calcium scoring and the CECT dataset uses contrast-enhanced coronary CT angiography (CE-CCTA). The model is first learned from scratch on the NCCT dataset with coronary calcium scoring CT. To cover the entire heart, it is scanned in 64 slices with 2.5 mm thickness on each acquisition. Then, the model pre-trained on that NCCT dataset is transferred to the CECT dataset which uses CE-CCTA. The CE-CCTA study is performed in 256 slices with 0.625 mm thickness

One of the key contributions of this paper is ~~to~~ validate the performance of our new developed 3D CNN-based approach on these difficult tasks. Since segmentation of ECF requires utilization of both voxel intensity and location information, we integrate two attention gate (AG) and deep supervision modules (DSV) ~~on~~ a standard 3D ~~U-Net~~. Our proposed model has better performance than the recent state-of-the-art approaches because of the integration of AG and DSV modules. The AG module is used to focus on the target structures by suppressing irrelevant regions in the input image. The DSV module is used to increase the number of learned features by generating a secondary segmentation map combining ~~from~~ different resolution levels of network layers. The second main contribution is the use of transfer learning, taking a model pre-trained on NCCT data, and applying it to CECT data, using only a small amount of data for the re-training. This approach has benefits in clinical applications for both NCCT and CECT data for ECF segmentation. Furthermore, our proposed solution is 3D-based and does not require preprocessing and postprocessing steps, thus it can easily integrate into the clinical workflow of CT acquisition to rapidly generate ECF volume results for the physician in clinical practice.

Related works

Conventional non-~~machine~~ learning methods have been proposed for ECF segmentation. Rodrigues et al. [1] proposed a genetic algorithm to recognize the pericardium contour on CT images. Militello et al. [2] proposed a semi-automatic approach using manual region-of-interest selection followed by thresholding segmentation. Zlokolica et al. [3] proposed local adaptive morphology and fuzzy c-means clustering. However, these conventional methods required many preprocessing steps before entering the segmentation algorithm. The next evolution of ECF segmentations ~~were~~ performed with a ~~machine~~ learning approach. ~~Rodrigues et al. [14] proposed ECF segmentation in CECT images using the Weka library (an open-source collection of machine learning algorithms) with Random-Forest as the classifier. The experiment, performed on 20 patients, yielded a Dice score of 97.7%.~~ Commandeur et al. [13] proposed ECF

segmentation from non-contrast coronary artery calcium computed tomography using ConvNets. They reported the Dice score of 82.3%.

To improve the performance of medical image segmentation, several modifications of U-Net have been proposed. The spatial attention gate has been proposed to focus on the spatial and detailed structure of the important region varying in shape and size [15]. Schlemper et al. [15] demonstrate the performance of the attention U-Net on real-time fetal detection on 2D images and pancreas detection on 3D CT images. He et al. [11] proposed ECF segmentation from CE-CCTA using a modified 3D U-Net approach by adding attention gates (AG). AGs are commonly used in classification tasks [16-19] and have been applied for various medical image problems such as image classification [19-20], image segmentation [15-20], and image captioning [20]. AG ~~is~~ used to focus on the relevant portion of the image by suppressing irrelevant regions [15]. The integration of AG ~~on~~ the standard U-Net [6, 10, 11, 15, 21] or V-Net [10, 22] has been demonstrated to have benefits for region localization.

As mentioned above, the ECF has a complex~~-shaped~~ structure. Some parts contain a thin layer adjacent to the cardiac muscle, which is similar to the microvasculature of the retinal vascular image visualized as small linear structures. In order to improve the performance of segmentation of small structures, several modules have been integrated into the main architecture of U-Net and V-Net such as dense-layer and deep supervision modules [21-25]. The ~~D~~dense-layer [23, 24] has been used to enhance the segmentation result instead of the traditional convolution in the U-Net model. Deep supervision [21, 22, 25] was used to ~~improve~~ local minimal traps during the training. The deep supervision helps to improve model convergence and increase the number of learned features [21]. Kearney et al. [21] showed that addition of deep supervision added to the U-Net model ~~could~~ improve the performance of 3D segmentation in CT image of prostate gland, rectum, and penile bulb.

While 2D and 3D deep learning approaches have been used for medical image segmentation, 3D approaches have typically shown better performance than the 2D approaches [8, 9, 26]. For example, Zhou et al. [9] demonstrated the better performance of 3D CNN approaches on multiple organs on 3D CT images, when compared to the 2D based method. Starke et al. [8] also demonstrated that 3D CNN achieved better performance on segmentation of head and neck squamous cell carcinoma on CT images. Woo et al. [26] demonstrated that 3D U-Net provided better performance on brain tissue MRI images, compared with 2D U-Net, on a smaller training dataset. Therefore, in this paper we use a 3D CNN for segmenting epicardial fat in cardiac CT images.

Materials & Methods

CNN architecture

The model architecture is based on a 3D U-Net model composed of multiple levels of encoding and decoding paths. The initial number of features at the highest layers of the model is 32. The numbers of feature maps are doubled with each downsampling path. In addition to the

original U-Net architecture, we added an attention gate connecting the encoding and decoding paths and deep supervision at the final step of the network. The model is created on a fully 3D structure at each network level. The final layer is an element-wise sum of feature maps of two last decoding paths. The segmentation map of two classes (epicardial fat and background) is the output layer with threshold ~~of~~ 0.5 to generate the binary classification of the epicardial fat. The architecture of the proposed network is shown in Fig 2.

Starting with the standard 3D U-Net architecture, the attention gate module connects each layer of encoding and decoding paths. The gating signal (g) is chosen from the encoding path and the input features (x) are collected from the decoding path. To generate the attention map, g and x go through a 1x1x1 convolution layer and element-wise sum, followed by rectified linear unit (ReLu) activation, a channel-wise 1x1x1 convolutional layer, batch normalization and sigmoid activation layer. The output of sigmoid activation is concatenated to the input x to get the output of the attention gate module [11, 21].

Deep supervision [10, 22] is the module at the final step of the network where it generates the multiple segmentation maps at different resolution levels, which are then combined together. The secondary segmentation maps are created from each level of decoding paths which are then transposed by 1x1x1 convolution. All feature maps are combined by element-wise sum. The lower resolution map is upsampled by 3D transposed convolution to have the same size as the second-lower resolution. Two maps are combined with element-wise sum then upsampled and added to the next level of segmentation map, until reaching the highest resolution level.

CT imaging data

This experimental study was approved and participant consent was waived by the institutional review board of Siriraj Hospital, Mahidol University (certificate of approval number Si 766/2020). The experimental datasets were acquired from 220 patients with non-contrast enhanced calcium scoring and 40 patients with CE-CCTA. The exclusion criteria were post open surgery of the chest wall. All CT acquisition was performed with the 256-slice multi-detector row CT scanner (Revolution CT; GE Medical Systems, Milwaukee, Wisconsin, United States). The original CT datasets of NCCT and CECT studies were 64 slices in 2.5 mm slice thickness and 256 slices in 0.625 mm slice thickness, respectively. All DICOM images were incorporated into a single 3D CT volume file with preserved original pixel intensity. Due to limitation of GPU memory, the 256 slices of CE-CCTA were pre-process~~ing~~ with rescaling to 64 images in the volume dataset. The final 3D volume dataset in all experiments was 512x512x64. The dataset was raw 12 bits grayscale in each voxel. The area of pericardial fat was defined by fat tissue attenuation inside the pericardium, ranging from -200 HU to -30 HU [14, 27, 28]. The ground-truth segmentation of ECF in all axial slices was performed using the 3D slicer software version 4.10.0 by a cardiovascular radiologist with more than 15 years of experience. No additional feature map or augmentation was performed in the pre-processing step.

Training framework

The experiments were implemented using the ~~p~~pytorch (v1.8.0) deep learning library ~~with Tensorflow backend~~ in Python (v3.6.9). The workflow for network training is illustrated in Fig 3. The training and testing processes were performed on a ~~cuda~~-enabled GPU (Nvidia DGX-A100) with 40 GB RAM. The experiments were divided into three scenarios: model validity assessment, NCCT, and CECE experiments. The parameters were the same for all three experiments. The networks were trained with RMSprop optimizer and mean squared error loss. The training parameters of learning rate, weight decay, and momentum were le-3, le-8 and 0.9, respectively. The initial random seed was set to be 0. The illustration of the experimental framework is shown in Fig 3.

The first experiment was the assessment of the model validity, for which we used 5-fold cross validation. The total dataset consisted of 200 volume-sets (12,800 images), divided into five independent folds. Each fold contained 160 volume-sets (10,240 images) for training and 40 volume-sets (2,560 images) for validation, without repeated validation data between folds. The other 20 volume-sets (1,280 images) were left for testing in second and third experiments~~. The volume matrix of each dataset was 512x512x64 pixels~~. Then the 5-fold cross validation was performed on standard U-Net, AG-U-Net, DSV-U-Net and the proposed method (AG-DSV-U-Net). For each fold of validation, the model with the best training accuracy after 150 epochs was selected for the validation.

The second experiment was to assess segmentation performance by training the network from scratch with the NCCT dataset. The volume matrix of each dataset was 512x512x64 pixels. To compare the performance of segmentation, this experiment was performed with four model architectures: standard U-Net, AG-U-Net, DSV-U-Net and proposed method (AG-DSV-U-Net). The network was ~~trained~~ with ~~a~~ hold-out method, in which a total of 220 volume-sets (14,080 images) were split into 200 volume-sets (12,800 images) for training and 20 volume-sets (1,280 images) for testing. The model output ~~on the training data~~ was collected at the best accuracy of total 300 epochs, named model-A.

The third experiment was to assess segmentation performance ~~in~~ CECT dataset and to evaluate the effectiveness of transferring the learning from NCCT to CECT dataset~~s~~. The pre-train~~ing~~ 3D model (model-A) was trained on large calcium scoring NCCT dataset~~s~~. The key success of the transfer learning ~~on~~ 3D U-Net is to fine-tune only the shallow layers (contracting path) [29] instead of the whole network. This contracting path represents ~~a~~ more low-level feature ~~of~~ the network [29]. The retraining dataset requires only a small amount of data - in our case only 20 volume-sets of CECT data. ~~T~~hese retraining ~~datasets~~ are not from the same cases as used in the pre-trained model. ~~The original volume matrix of each dataset was 512x512x256 pixels. Due to the limitation of GPU memory, the pre-processing step is voxel rescaling from 256 to 64 slices in the z plane.~~ To compare the performance of segmentation, this experiment was performed with four model architectures: standard U-Net, AG-U-Net, DSV-U-Net and proposed method (AG-DSV-U-Net). The network was ~~trained~~ with ~~a~~ hold-out method, in which the total 40 volume-sets (2,560 images) were split into 20 volume-sets (1,280 images) for

training and 20 volume-sets (1,280 images) for testing. The output model is collected at the best training accuracy of total 300 epochs, named model-B.

Performance evaluation

The performance of our proposed CNN segmentation is compared with the performance of the existing methods. The evaluation was quantitatively evaluated by comparison with the reference standard using the Dice similarity coefficient (DSC), Jaccard similarity coefficient (JSC) and Hausdroff distance (HD). An average HD value was calculated using the insight toolkit library of 3D slicer. Differences in the comparison coefficient among the four groups of experiments (standard U-Net, AG-U-Net, DSV-U-Net and AG-DSV-U-Net) were assessed with a paired Student's t-test. P values <0.05 indicated a statistically significant difference. Differences in the comparison between DSC of segmentation result and ECF volume were assessed with Pearson's correlation coefficient. The Pearson's values of < 0.3 indicated poor correlation, 0.3 to 0.7 indicated moderate correlation, and > 0.7 indicated good correlation.

Results

The patient demographics are shown in Table 1. The training dataset of NCCT has an average age of 61.43 years and an average volume of 135.75 ml. The testing dataset of non-contrast CT has a similar distribution, with an average age of 67.80 years and an average volume of 127.59 ml. For the contrast-enhanced dataset, the average ages of training and testing datasets were 65.85 and 60.85 years, respectively. The average volumes of epicardial fat of training and testing datasets were 117.13 and 121.43 ml, respectively.

Five-fold cross validation experiments on our NCCT dataset were used to evaluate the validity and repeatability performance of the proposed method. The dataset was split into training (80%) and validation (10%) for each fold. On each model architecture, the validation data exhibits good results across each fold. The proposed method also demonstrates the best average performance (DSC = 89.02), when compared with other methods ($p<0.05$). (Table 2).

The experimental result of the NCCT dataset is shown in Table 3. The proposed CNN-based method for ECF segmentation on the non-contrast dataset demonstrates excellent results, achieving average DSC, JSC, HD values of 90.06±4.60, 82.42±6.91 and 0.25±0.14, respectively. The baseline of the experiment is the standard 3D U-Net which demonstrates good results with DSC, JSC and HD values of 84.87±5.73, 74.12±8.0.8 and 0.34±0.18, respectively. The segmentation results of the modified U-Net models (AG-U-Net, DSV-U-Net and the proposed method) demonstrate statistically significant improvement compared with the standard U-Net ($p<0.05$). The difference between segmentation results of AG-U-Net and DSV-U-Net is not statistically significant ($p>0.05$). The DSC, JSC, HD values of AG-U-Net are 89.59+4.45, 81.41+6.77 and 0.27+0.12, respectively. The proposed method statistically improved the segmentation result ($p<0.05$) compared with the standard U-Net and AG-U-Net. While the

proposed method is better than DSV-U-Net, the difference is not statistically significant ($p>0.05$). Examples of segmentation results of the proposed method are shown in Fig 4.

The experimental result of the proposed method on the CECT dataset is shown in Table 4. This transfer learning approach achieved average DSC, JSC and HD values of 88.16±4.57, 79.10±6.75 and 0.28±0.20, respectively ~~(Table 4)~~. The segmentation result of the proposed method demonstrates statistically significant improvement~~,~~ when compared with other methods ($p<0.0~~.~~5$). The segmentation results of the standard 3D U-Net and DSV-U-Net demonstrate good similar performance ($p>0.05$). The segmentation results of the standard 3D U-Net and DSV-U-Net are statistically significantly better than AG-U-Net ($p<0.05$). Examples of segmentation results of transfer learning with the proposed methods are shown in Fig 5.

Discussion

Segmentation of ECF is a difficult image segmentation task because of the thin layer and complex structures at the outer surface ~~sulci~~ of the heart. The ECF is also variable in distribution, depending on body habitus. In general, obese patients have larger amounts of ECF than do thin patients. Segmentation of ECF is more challenging than the segmentation of other cardiac structures.

Most CNN approaches work on 2D images whereas in clinical practice, 3D volume segmentation is used [30]. The 2D-based CNN approaches such as ResNet and VGG are not applicable for 3D datasets. The model architectures for 2D CNN and 3D CNN are different [8, 9, 26]. 3D CNN has an advantage over 2D-CNN by extracting both spectral and spatial features simultaneously, while 2D CNN can extract only spatial features from the input data [4]. For this reason in general, the 3D CNN is more accurate than 2D [4, 31]. 2.5D CNN has been developed to solve the memory consumption problem of 3D models [32]. 2.5D CNN has at least three approaches [32, 33]. The first is a combination of output of 2D CNN in three orthogonal planes (axial, coronal and sagittal) with majority voting. The second is 2D CNN with 3 or 5 channels from adjacent 3 or 5 slices. Third is 2D CNN with randomly oriented 2D cross sections. In the final step, 2.5D segmentation requires an additional post-processing step to generate 3D output [34]. ~~Recent studies demonstrated that the 3D CNN provides a higher accuracy for image segmentation, when compared with the 2D CNN [8, 9, 26]. However,~~ the 3D CNN requires more resources and time for the model training~~. For~~ the best performance, we use the 3D CNN in our implementation. The best performing methods for 3D volume segmentation of medical data are U-Net and V-Net. V-Net has more trainable parameters in its network architecture. Recent experimental comparisons of U-Net and V-Net on medical data have not shown statistically significant differences in performance [11, 35]. However, U-Net is less complex and easier to modify so that additional modules can be ~~used to integrate to~~ the standard U-Net in order to improve the performance.

Several state-of-the-art approaches for CNN-based segmentation of ECF have recently been proposed. Commandeur et al. [13] proposed the first CNN-based method for ECF

segmentation in a non-contrast 2D CT dataset using a multi-task convolutional neural network called ConvNets. They reported a Dice score of 82.3% for the segmentation result. He et al. [11] proposed another CNN-based method on a 3D CECT dataset using AG integrated into 3D U-Net. The segmentation result was reported to have a DSC of 88.7% [11]. We repeated the experiment by implementing the AG in 3D U-Net on our NCCT dataset by hold-out test. The amount of ~~hold-out train on our dataset~~ (200 volume-sets) is more than the one used in the previous article (150 volume-sets) [11]. Our 4 layer AG-U-Net method demonstrates significant improved performance with DSC of 89.59%, as compared with 3 layer AG-U-Net of 86.54% ($p<0.05$). That should be due to more layers of our network. In our implementation, ~~our~~ AG-U-Net has ~~deeper~~ convolutional layers (4 layers) and removes sigmoid at the end of the network. However, the AG integration provides significantly better performance ($p<0.05$) as compared with standard 3D U-Net (DSC of 84.87%). To the best of our knowledge, our experiment uses the largest volume size of ~~the~~ dataset (512x512x64). ~~We try to improve the accuracy of the segmentation by modifying the standard U-Net architecture.~~ We introduce a novel approach to 3D segmentation of ECF by integrating both AG and DSV modules into all layers of 3D U-Net deep learning architecture. The AGs are commonly used in natural image analysis and natural language processing [36, 37], which can generate attention-awareness features. The AG module is beneficial for organ localization, which can improve organ segmentation [15]. During CNN training, AG is automatically learned to focus on the target without additional supervision [38]. The AG module can improve model accuracy by suppressing feature activation in irrelevant regions of an input image [15]. The AG module is used to make connections between encoding and decoding paths on the standard U-Net. The DSV module is used to deal with the vanishing gradient problem in the deeper layer of CNN [10, 25]. The standard approach provides the supervision only at the output layer. But the DSV module propagates the supervision back to the earlier layer by generating a secondary segmentation map combining from different resolution levels. The losses of this segmentation map ~~is~~ weighted and added to the final loss function ~~that~~ can effectively increase the performance [39]. The DSV module is used by adding into the decoding path of 3D U-Net. The AG-DSV modules had been implemented in previous work [10, 22] for kidney [22] tumor segmentation (Kidney Tumor Segmentation Challenge 2019), as well as for liver [10] and pancreas [10] tumor segmentation (Medical Decathlon Challenge 2018).

The experiment demonstrated that our proposed method (AG-DSV-U-Net) achieves excellent performance with average and max DSC values of 90.06% and 95.32%, respectively. Our proposed method also shows a significant improvement of performance (90.06%), when compared with the previous state-of-the-art network (86.54%) on the same dataset ($p<0.05$). The example of the results was shown in Fig 4. While one might expect the segmentation performance to improve with fat volume of the dataset, unexpectedly, the statistical analysis demonstrates that there is poor correlation between segmentation performance and fat volume (Pearson's correlation 0.2).

The 3D volume size of our dataset is larger (512x512x64) compared with the previous work (512x512x32) [11]. For comparative analysis on different numbers of slices and image

resolution of dataset, the previous work demonstrates that a 40-slice of volume dataset achieves 1% higher DSC than 32-slice and 24-slice [11]. However, the training time is also increased. We set up additional experiments to test the effect of different numbers of slices and image resolution on segmentation performance with our proposed model (AG-DSC-U-Net).The ~~number of~~ training, testing datasets and hyperparameters are the same as defined in the NCCT experiment. The experiment of effect of the number of slices was performed by rescaling the slices with 64, 32 and 16 slices and fixed image resolution with 512x512 pixels. The segmentation results ~~of 64, 32 and 16 slices~~ are DSC 90.06%, 81.76%, 78.93%, respectively. The experiment to determine the effect of different image resolution scales was performed by rescaling the resolution with 512x512, 256x256 and 128x128 pixels, with the number of slices fixed at 64. The segmentation results ~~of 512x512, 256x256 and 128x128 resolution~~ are DSC 90.06%, 86.19% and 83.73% respectively. The 512x512 image resolution and 64 slices ~~still~~ give the best performance, with significant improvement over lower resolution ($p<0.05$). More slices and higher image resolution of the dataset let the network extract more spatial information that can help to improve segmentation accuracy. Furthermore, because the ECF ~~somewhere~~ is a thin layer along the ~~sulcus of the~~ heart contour, more spatial resolution will improve segmentation accuracy. To give the best performance, we ~~choose the~~ 64 slices for our implementation which is a perfect fit with the original NCCT dataset, having 64 images in each dataset. In the CECT, the original CT dataset had 256 slices and needed to be rescaled to 64 slices. Due to ~~this~~ limitation of the proposed model and current GPU architecture, the voxel size of train and test datasets cannot be extended beyond 64 slices. The other limitation of this study is the size of the dataset: 220 volume-sets for NCCT experiment and 40 volume-sets for CECT experiment. ~~However, the experiment demonstrates the excellent result of the testing.~~

In clinical practice, the cardiac CT scan can be performed in NCCT or CECT or ~~both studies~~. For this reason, the ECF can also be either segmentation from NCCT or CECT dataset. To the best of our knowledge, ours is the first implementation of ECF segmentation on NCCT and CECT datasets. In our experiment, we start to train with the NCCT dataset (200 volume-sets). We use the concept of transfer learning to re-train with a similar dataset by taking a small amount of the dataset (Fig 3). We re-train the pre-trained NCCT model with a small amount of CECT data (20 volume-sets). We test the model with additional testing of 20 volume-sets. The experimental result ~~achieves~~ good performance with a DSC value of 88.16%. The performance ~~result~~ is also significantly better than the standard U-Net and AG-U-Net (Table 4). ~~O~~ur proposed re-trained model demonstrates ~~a~~ good performance as compared with the previous training from scratch (88.7%) [9].

Future studies could include investigations in more data diversity from multiple CT venders, larger patient variation, and testing the model across different healthcare centers. Further investigation ~~in~~ clinical correlation between CNN segmentation of ECF volume and occurrence of cardiovascular disease ~~would be also~~ interesting research question~~s~~.

Conclusions

In the paper, we have introduced a CNN-based approach for ECF segmentation using integration of AG and DSV modules into the standard 3D U-Net. ECF segmentation is one of the most difficult medical image segmentation tasks. We trained the NCCT dataset from scratch and re-trained on a CE CT dataset from the pre-trained NCCT model. We successfully improved the performance of ECF segmentation in both NCCT and CECT datasets, when compared with the previous state-of-the-art methods. It is expected that this proposed method has potential to improve the performance in other difficult segmentation tasks. This concept of training and retraining models can be also applied to other medical image segmentation problems.

Acknowledgements

References

[1]. Rodrigues ÉO, Rodrigues LO, Oliveira LSN, Conci A, Liatsis P. Automated recognition of the pericardium contour on processed CT images using genetic algorithms. Computers in Biology and Medicine. 2017;87:38-45.

[2]. Militello C, Rundo L, Toia P, Conti V, Russo G, Filorizzo C, Maffei E, Cademartiri F, La Grutta L, Midiri M, Vitabile S. A semi-automatic approach for epicardial adipose tissue segmentation and quantification on cardiac CT scans. Computers in Biology and Medicine. 2019;114.

[3]. Zlokolica V, Krstanović L, Velicki L, Popović B, Janev M, Obradović R, Ralević NM, Jovanov L, Babin D. Semiautomatic Epicardial Fat Segmentation Based on Fuzzy c-Means Clustering and Geometric Ellipse Fitting. Journal of Healthcare Engineering. 2017;2017.

[4]. Singh SP, Wang L, Gupta S, Goli H, Padmanabhan P, Gulyás B. 3D Deep Learning on Medical Images: A Review. Sensors. 2020;20(18).

[5]. Kim M, Yun J, Cho Y, Shin K, Jang R, Bae H-J, Kim N. Deep Learning in Medical Imaging. Neurospine. 2019;16(4):657-68.

[6]. Hesamian MH, Jia W, He X, Kennedy P. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. Journal of Digital Imaging. 2019;32(4):582-96.

[7]. Renard F, Guedria S, Palma ND, Vuillerme N. Variability and reproducibility in deep learning for medical image segmentation. Scientific Reports. 2020;10(1):13724.

[8]. Starke S, Leger S, Zwanenburg A, Leger K, Lohaus F, Linge A, Schreiber A, Kalinauskaite G, Tinhofer I, Guberina N, Guberina M, Balermpas P, von der Grün J, Ganswindt U, Belka C, Peeken JC, Combs SE, Boeke S, Zips D, Richter C, Troost EGC, Krause M, Baumann M, Löck S. 2D and 3D convolutional neural networks for outcome modelling of locally advanced head and neck squamous cell carcinoma. Sci Rep. 2020;10(1):15625.

[9]. Zhou X. Automatic Segmentation of Multiple Organs on 3D CT Images by Using Deep Learning Approaches. In: Lee G, Fujita H, editors. Deep Learning in Medical Image Analysis : Challenges and Applications. Cham: Springer International Publishing; 2020. p. 135-47.
[10]. Turečková A, Tureček T, Komínková Oplatková Z, Rodríguez-Sánchez A. Improving CT Image Tumor Segmentation Through Deep Supervision and Attentional Gates. Frontiers in Robotics and AI. 2020;7(106).
[11]. He X, Guo BJ, Lei Y, Wang T, Fu Y, Curran WJ, Zhang LJ, Liu T, Yang X. Automatic segmentation and quantification of epicardial adipose tissue from coronary computed tomography angiography. Physics in Medicine and Biology. 2020;65(9).
[12]. Rodrigues ÉO, Pinheiro VHA, Liatsis P, Conci A. Machine learning in the prediction of cardiac epicardial and mediastinal fat volumes. Computers in Biology and Medicine. 2017;89:520-9.
[13]. Commandeur F, Goeller M, Betancur J, Cadet S, Doris M, Chen X, Berman DS, Slomka PJ, Tamarappoo BK, Dey D. Deep Learning for Quantification of Epicardial and Thoracic Adipose Tissue from Non-Contrast CT. IEEE Transactions on Medical Imaging. 2018;37(8):1835-46.
[14]. Rodrigues ÉO, Morais FFC, Morais NAOS, Conci LS, Neto LV, Conci A. A novel approach for the automated segmentation and volume quantification of cardiac fats on computed tomography. Computer Methods and Programs in Biomedicine. 2016;123:109-28.
[15]. Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, Rueckert D. Attention gated networks: Learning to leverage salient regions in medical images. Medical Image Analysis. 2019;53:197-207.
[16]. Wu Y, Zhang X, Xiao Y, Feng J. Attention Neural Network for Water Image Classification under IoT Environment. Applied Sciences. 2020;10(3).
[17]. Sharmin S, Chakma D. Attention-based convolutional neural network for Bangla sentiment analysis. AI & SOCIETY. 2021;36(1):381-96.
[18]. Fei H, Zhang Y, Ren Y, Ji D. Optimizing Attention for Sequence Modeling via Reinforcement Learning. IEEE Transactions on Neural Networks and Learning Systems. 2021:1-10.
[19]. Kelvin X, Jimmy B, Ryan K, Kyunghyun C, Aaron C, Ruslan S, Rich Z, Yoshua B. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. 2015/06/01: PMLR; 2015. p. 2048-57.
[20]. Zhao B, Feng J, Wu X, Yan S. A survey on deep learning-based fine-grained object classification and semantic segmentation. International Journal of Automation and Computing. 2017;14(2):119-35.
[21]. Kearney V, Chan JW, Wang T, Perry A, Yom SS, Solberg TD. Attention-enabled 3D boosted convolutional neural networks for semantic CT segmentation using deep supervision. Physics in Medicine & Biology. 2019;64(13):135001.
[22]. Tureckova A, Turecek T, Komínková Z, Rodríguez-Sánchez A, editors. KiTS challenge: VNet with attention gates and deep supervision2019.

[23]. Liang Z, Liu H, Zhao X, Yu L. Segmentation of Retinal Vessels Based on DenseNet-Attention-Unet Model Network. Proceedings of the 2nd International Conference on Industrial Control Network And System Engineering Research; Kuala Lumpur, Malaysia: Association for Computing Machinery; 2020. p. 111–7.
[24]. Wu C, Zou Y, Zhan J. DA-U-Net: Densely Connected Convolutional Networks and Decoder with Attention Gate for Retinal Vessel Segmentation. IOP Conference Series: Materials Science and Engineering. 2019;533:012053.
[25]. Chen-Yu L, Saining X, Patrick G, Zhengyou Z, Zhuowen T. Deeply-Supervised Nets. 2015/02/21: PMLR; 2015. p. 562-70.
[26]. Woo B, Lee M, editors. Comparison of tissue segmentation performance between 2D U-Net and 3D U-Net on brain MR Images. 2021 International Conference on Electronics, Information, and Communication (ICEIC); 2021 31 Jan.-3 Feb. 2021.
[27]. Shahzad R, Bos D, Metz C, Rossi A, Kirişli H, van der Lugt A, Klein S, Witteman J, de Feyter P, Niessen W, van Vliet L, van Walsum T. Automatic quantification of epicardial fat volume on non-enhanced cardiac CT scans using a multi-atlas segmentation approach. Medical Physics. 2013;40(9):091910.
[28]. Kazemi A, Keshtkar A, Rashidi S, Aslanabadi N, Khodadad B, Esmaeili M, editors. Automated Segmentation of Cardiac Fats Based on Extraction of Textural Features from Non-Contrast CT Images. 2020 25th International Computer Conference, Computer Society of Iran, CSICC 2020; 2020.
[29]. Amiri M, Brooks R, Rivaz H. Fine-Tuning U-Net for Ultrasound Image Segmentation: Different Layers, Different Outcomes. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control. 2020;67(12):2510-8.
[30]. Milletari F, Navab N, Ahmadi S, editors. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 Fourth International Conference on 3D Vision (3DV); 2016 25-28 Oct. 2016.
[31]. Han L, Chen Y, Li J, Zhong B, Lei Y, Sun M. Liver segmentation with 2.5D perpendicular UNets. Computers & Electrical Engineering. 2021;91:107118.
[32]. Zhang C, Hua Q, Chu Y, Wang P. Liver tumor segmentation using 2.5D UV-Net with multi-scale convolution. Computers in Biology and Medicine. 2021;133:104424.
[33]. Minnema J, Wolff J, Koivisto J, Lucka F, Batenburg KJ, Forouzanfar T, van Eijnatten M. Comparison of convolutional neural network training strategies for cone-beam CT image segmentation. Computer Methods and Programs in Biomedicine. 2021;207:106192.
[34]. Han Y, Li X, Wang B, Wang L. Boundary Loss-Based 2.5D Fully Convolutional Neural Networks Approach for Segmentation: A Case Study of the Liver and Tumor on Computed Tomography. Algorithms. 2021;14(5).
[35]. Ghavami N, Hu Y, Gibson E, Bonmati E, Emberton M, Moore CM, Barratt DC. Automatic segmentation of prostate MRI using convolutional neural networks: Investigating the impact of network architecture on the accuracy of volume measurement and MRI-ultrasound registration. Medical Image Analysis. 2019;58:101558.

[36]. Wang F, Jiang M, Qian C, Yang S, Li C, Zhang H, Wang X, Tang X, editors. Residual Attention Network for Image Classification. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 21-26 July 2017.
[37]. Anderson P, He X, Buehler C, Teney D, Johnson M, Gould S, Zhang L. Bottom-Up and Top-Down Attention for Image Captioning and VQA. ArXiv. 2017;abs/1707.07998.
[38]. Oktay O, Schlemper J, Folgoc LL, Lee MJ, Heinrich M, Misawa K, Mori K, McDonagh SG, Hammerla N, Kainz B, Glocker B, Rueckert D. Attention U-Net: Learning Where to Look for the Pancreas. ArXiv. 2018;abs/1804.03999.
[39]. Kayalibay B, Jensen G, Smagt PVD. CNN-based Segmentation of Medical Imaging Data. ArXiv. 2017;abs/1701.03056.

Figure 1

The example CT dataset of epicardial fat

(A) is non-contrast (B) is post-contrast CT images. The pericardium is a thin layer of membrane surrounding the heart (arrow). The epicardial fat is fat along outer surface of heart and inside the pericardium (*).

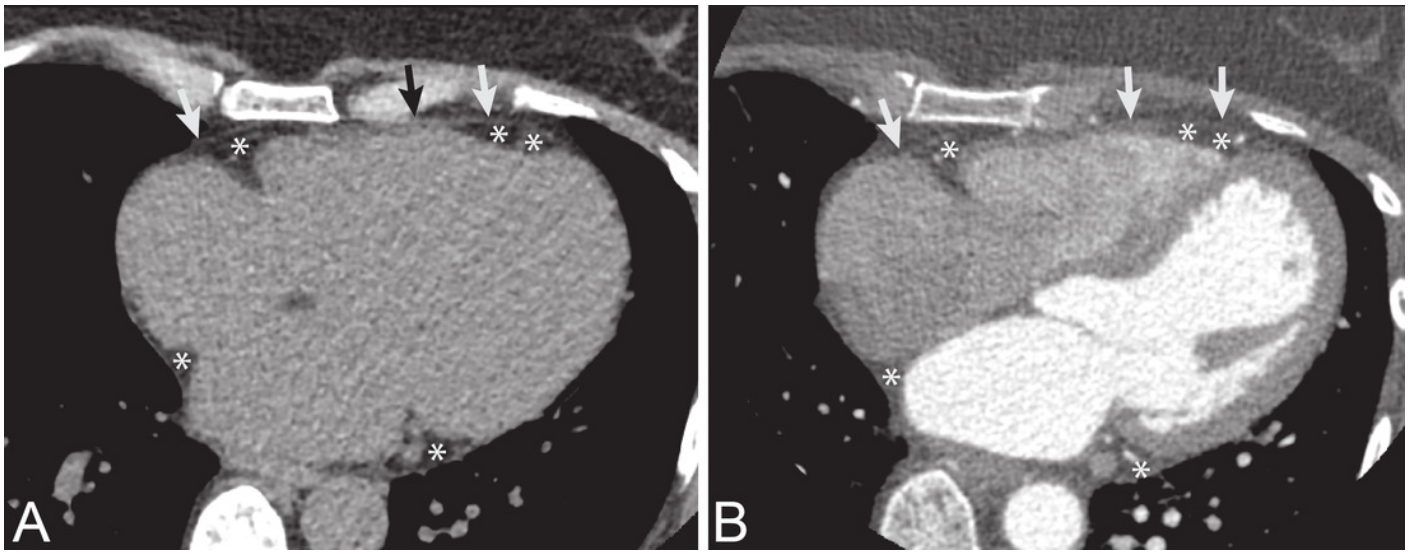

Figure 2

The proposed network ~~of~~ epicardial fat segmentation.

The network contains two main parts of the standard 3D U-Net integrated with the attention gate, and the deep supervision modules.

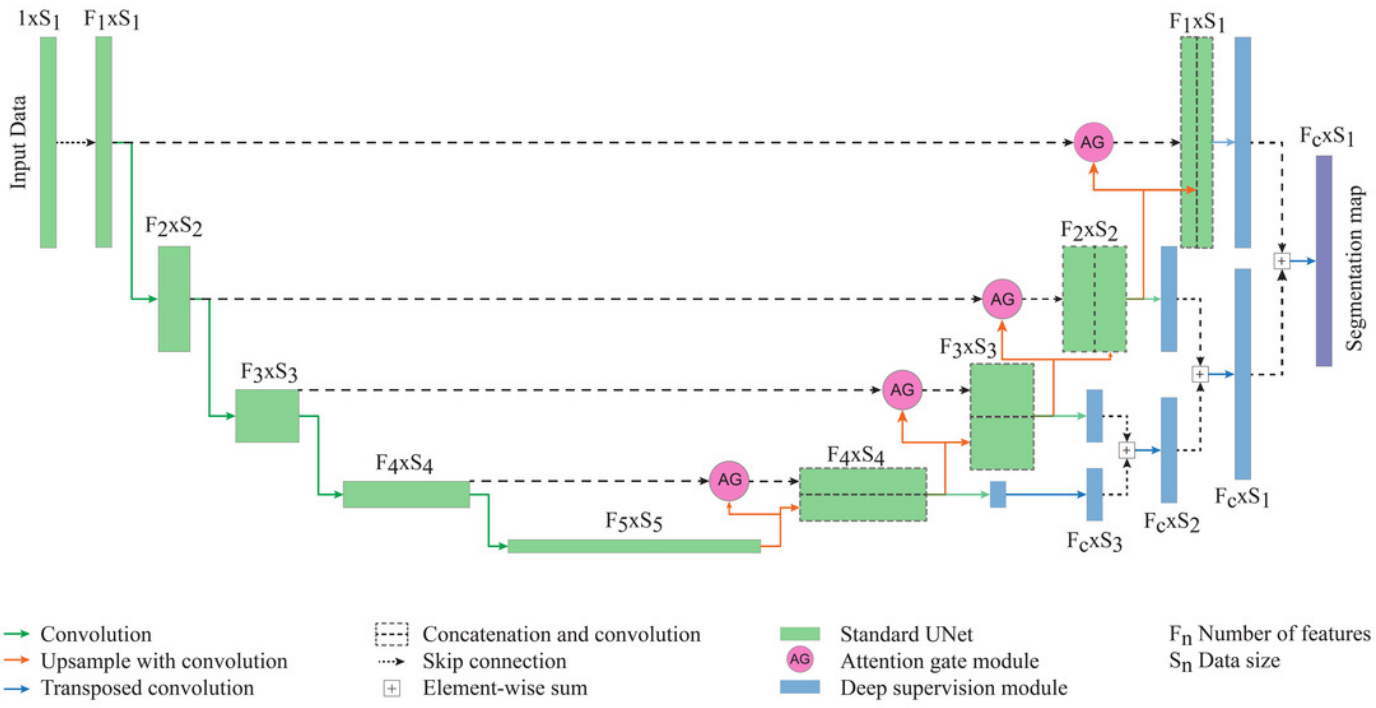

Figure 3

Illustration of framework of the proposed method.

The upper row ~~is the second experiment which~~ performs the network training from scratch with a non-contrast CT dataset. The lower row ~~is the third experiment which~~ performs the network re-training on a contrast-enhanced CT dataset. No post-processing is required in this framework.

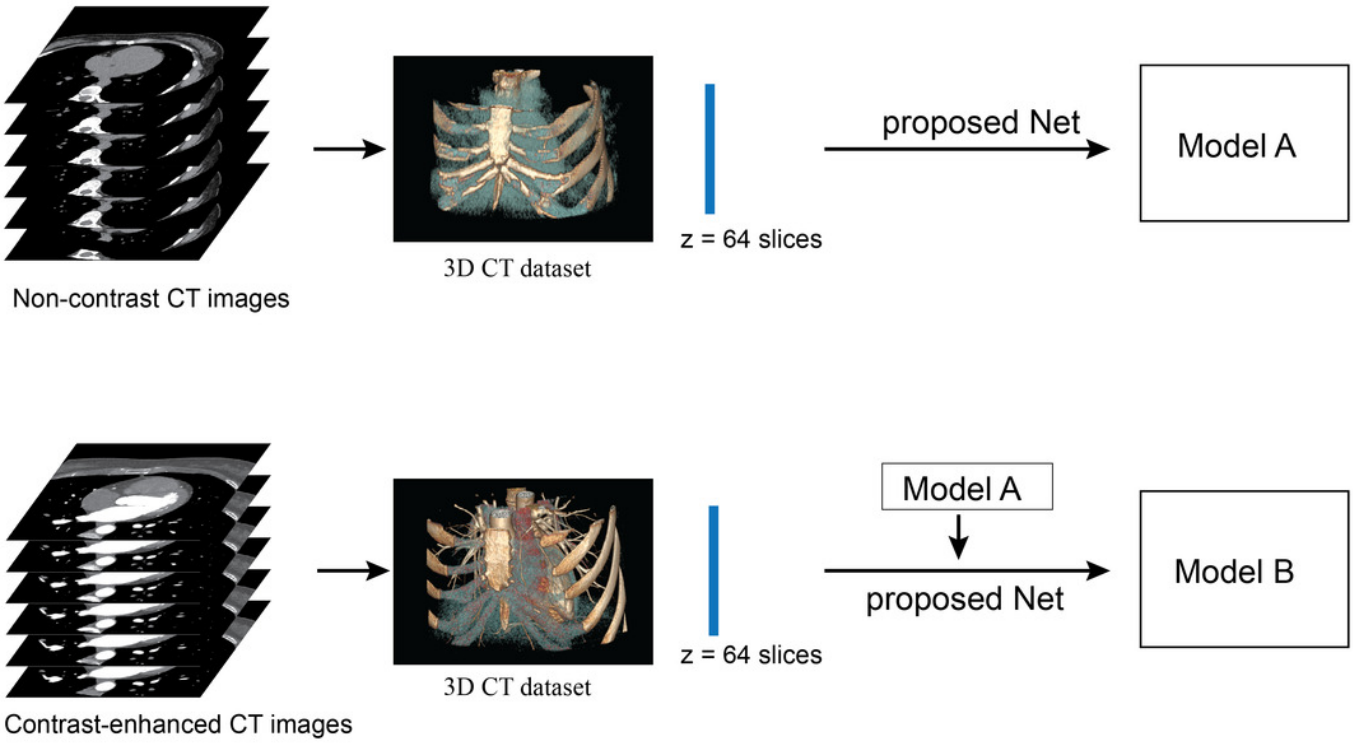

Figure 4

Example of segmentation result of proposed method on non-contrast CT images.

The first, second and third rows contain a source dataset in axial view, segmentation result in axial view and segmentation result in 3D reconstruction. The yellow color represents the segmented ECF using the proposed method and green color represents the ground-truth. The DSC values from left to right ~~volume sets~~ are 90.90%, 88.63%, 95.31% and 87.86%, respectively. The fat volumes from left to right ~~volume sets~~ are 95.69 ml, 116.95 ml, 106.99 ml and 91.18 ml, respectively.

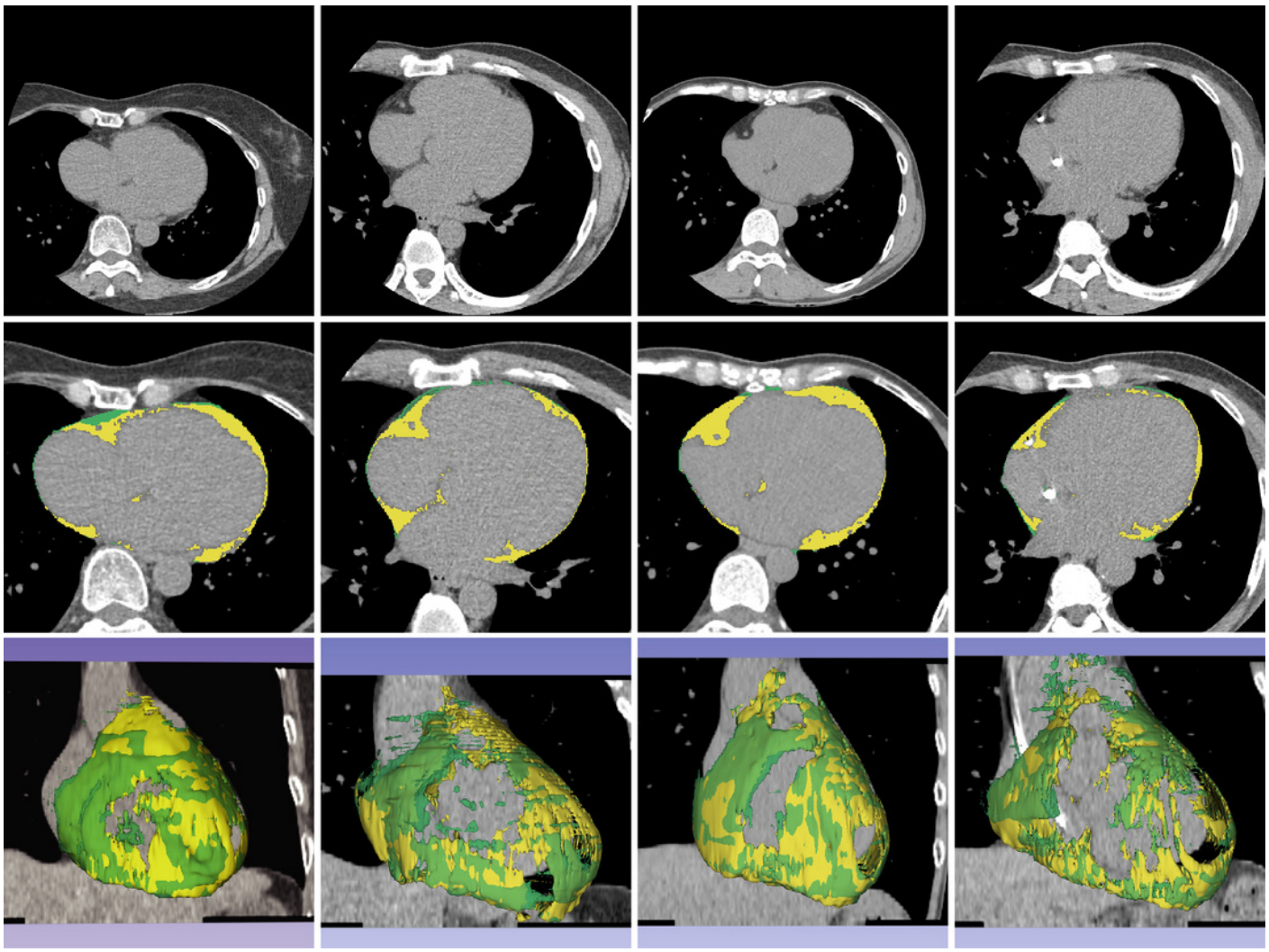

Figure 5

Example of segmentation result of proposed method on contrast-enhanced CT images.

The first, second and third rows contain a source dataset in axial view, segmentation results in axial view and segmentation results in 3D reconstruction. The yellow color represents the segmented ECF and green color represents the ground-truth. The DSC values from left to right ~~volume sets~~ are 80.23 %, 93.38%, 72.26% and 92.72%, respectively. The fat volumes from left to right volume-sets are 201.20 ml, 112.48 ml, 92.28 ml and 112.98 ml, respectively.

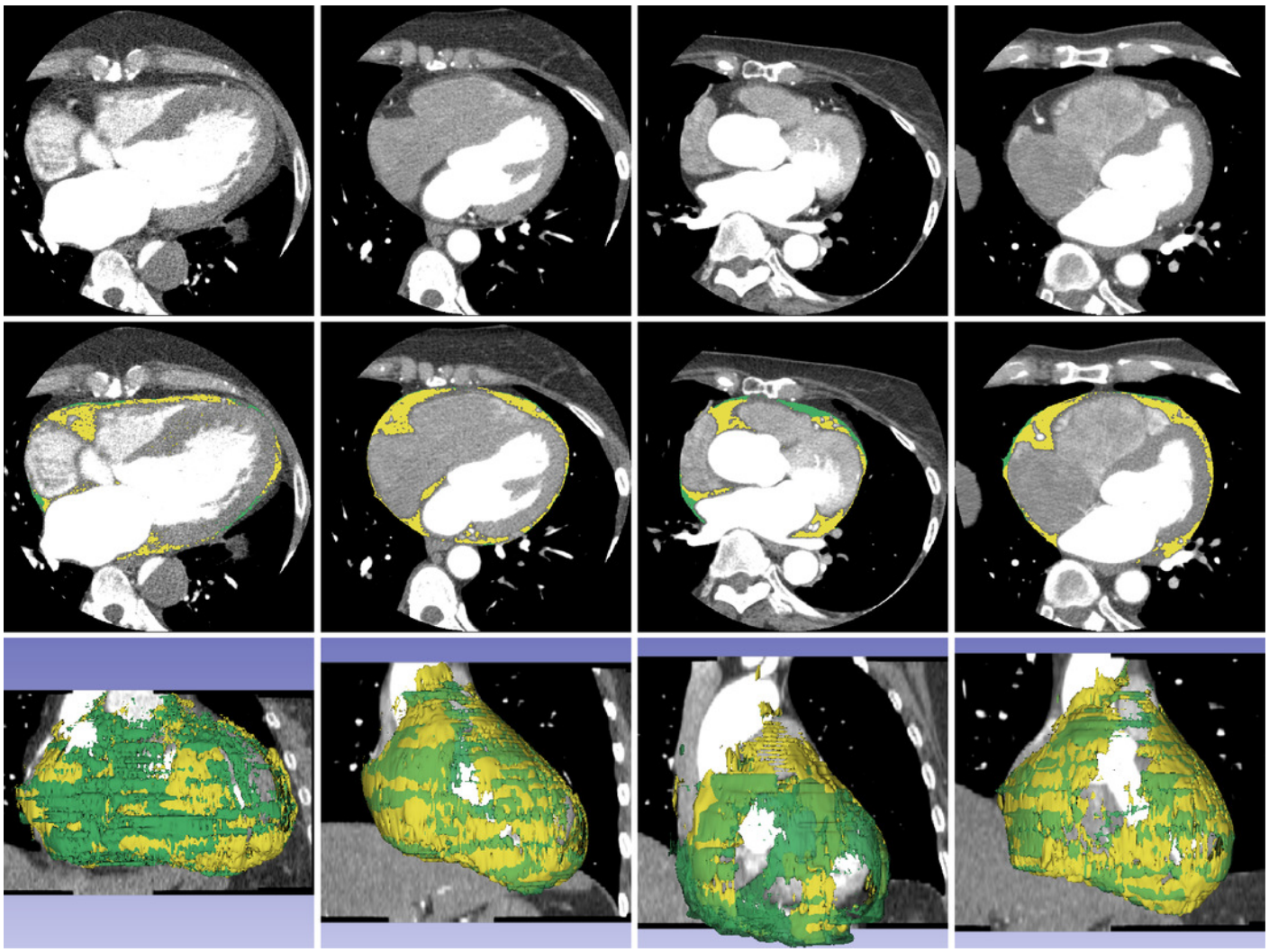

Table 1(on next page)

Patient characteristics of the CNN non-contrast and contrast-enhanced CT datasets

Patient characteristics of the CNN non-contrast and contrast-enhanced CT datasets

Table1: Patient characteristics of the CNN non-contrast and contrast-enhanced CT datasets

	Non-contrast CT		Contrast-enhanced CT	
	Training dataset	Testing dataset	Training dataset	Testing
No. of records	200	20	20	20
Average age (years)	61.41±12.27	67.80±11.66	65.85±8.36	60.75±10.31
Average volume (ml)	135.75±60.09	127.59±35.51	117.13±69.29	121.43±40.21
Min volume (ml)	6.39	71.86	47.34	66.03
Max volume (ml)	327.44	208.20	374.82	201.20

Table 2(on next page)

Results of 5-fold cross-validation. Mean Dice score coefficient and standard deviation were used to assess model validity and repeatability on non-contrast CT dataset.

Results of 5-fold cross-validation. Mean Dice score coefficient and standard deviation were used to assess model validity and repeatability on non-contrast CT dataset.

Table 2. Results of 5-fold cross-validation. Mean Dice score coefficient and standard deviation were used to assess model validity and repeatability on non-contrast CT dataset.

Fold	U-Net	AG-U-Net	DSV-U-Net	AG-DSV-U-Net (Proposed method)
1	87.48	87.93	87.70	89.55
2	89.97	89.65	89.93	90.73
3	89.22	89.76	88.30	88.76
4	89.50	90.08	85.56	89.61
5	86.44	84.77	82.84	86.46
mean	88.52	88.44	86.91	89.02

Table 3(on next page)

Experimental results with standard 3D U-Net, AG-U-Net and proposed method (AG-DSV-U-Net) on non-contrast CT dataset

Experimental results with standard 3D U-Net, AG-U-Net and proposed method (AG-DSV-U-Net) on non-contrast CT dataset

Table 3. Experimental results with standard 3D U-Net, AG-U-Net and proposed method (AG-DSV-U-Net) on non-contrast CT dataset

Non-contrast CT	U-Net	AG-U-Net	DSV-U-Net	AG-DSV-U-Net (Proposed method)
DSC	84.87±5.73	89.59±4.45	89.70±4.81	90.06+4.60
JSC	74.12±8.08	81.41±6.77	81.64±7.33	82.21+6.91
HD	0.34+0.18	0.27+0.12	0.28+0.14	0.25+0.14

Table 4(on next page)

Table 4. Transferred learning from pre-trained model to contrast-enhanced CT dataset

Table 4. Transferred learning from pre-trained model to contrast-enhanced CT dataset

Table 4. Transferred learning from pre-trained model to contrast-enhanced CT dataset

	U-Net	AG-U-Net	DSV-U-Net	AG-DSV-U-Net (Proposed method)
DSC	85.58+4.99	82.47+4.33	85.07+4.96	88.16+4.57
JSC	75.11+7.19	70.39+6.03	74.32+7.07	79.10+6.75
HD	0.41+0.36	0.34+0.23	0.35+0.30	0.28+0.20