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 myocardium and enclosed by the thin-layer pericardium membrane. 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%.

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 myocardium and enclosed by the thin-layer pericardium membrane. 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 and located inside the thin-layer pericardium membrane. The adipose tissue located outside the pericardium is called paracardial fat [1] and 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 the thin 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-deep learning techniques have been proposed for ECF segmentation, including genetic algorithms, region-of-interest selection with thresholding, and fuzzy c-means clustering [2-4]. Deep learning techniques have been applied to a wide variety of medical image segmentation problems with great success [5-7]. A recent article [8] demonstrates that deep learning algorithms outperform conventional methods for medical image segmentation in terms of accuracy. But most previous

80 studies involved large solid organs or tumor segmentation [9-11]. The segmentation of relatively
81 small and complex structures with high inter-patient variability, such as ECF, has been far less
82 successful. Recently, a few deep learning approaches to ECF segmentation have made progress
83 on this problem [12-14]. In this paper, we build upon the previous work by presenting a novel
84 deep learning model for 3D segmentation of ECF, We propose a solution of automatic
85 segmentation of ECF volume using a deep learning based approach that is evaluated on both
86 non-contrast and contrast-enhanced CT datasets. The NCCT dataset uses coronary calcium
87 scoring and the CECT dataset uses contrast-enhanced coronary CT angiography (CE-CCTA).
88 The model is first learned from scratch on the NCCT dataset with coronary calcium scoring CT.
89 To cover the entire heart, it is scanned in 64 slices with 2.5 mm thickness on each acquisition.
90 Then, the model pre-trained on that NCCT dataset is transferred to the CECT dataset which uses
91 CE-CCTA. The CE-CCTA study is performed in 256 slices with 0.625 mm thickness
92   One of the key contributions of this paper is that we validate the performance of our
93 newly developed 3D CNN-based approach on these difficult tasks. Since segmentation of ECF
94 requires utilization of both voxel intensity and location information, we integrate two attention
95 gate (AG) and deep supervision modules (DSV)into a standard 3D U-Net architecture. Our
96 proposed model has better performance than the recent state-of-the-art approaches we evaluate
97 because of the integration of AG and DSV modules. The AG module is used to focus on the
98 target structures by suppressing irrelevant regions in the input image. The DSV module is used to
99 increase the number of learned features by generating a secondary segmentation map combining
100 the different resolution levels of network layers. The second main contribution is the use of
101 transfer learning, taking a model pre-trained on NCCT data, and applying it to CECT data, using
102 only a small amount of data for the re-training. This approach has benefits in clinical applications
103 for both NCCT and CECT data for ECF segmentation. Furthermore, our proposed solution is
104 3D-based and does not require preprocessing and postprocessing steps, thus it can easily
105 integrate into the clinical workflow of CT acquisition to rapidly generate ECF volume results for
106 the physician in clinical practice.
107
108

109 **Related works**
110   Conventional non-deep learning methods have been proposed for ECF segmentation.
111 Rodrigues et al. [2] proposed a genetic algorithm to recognize the pericardium contour on CT
112 images. Militello et al. [3] proposed a semi-automatic approach using manual region-of-interest
113 selection followed by thresholding segmentation. Zlokolica et al. [4] proposed local adaptive
114 morphology and fuzzy c-means clustering. Rodrigues et al. [15] proposed ECF segmentation in
115 CECT images using the Weka library (an open-source collection of machine learning algorithms)
116 with Random-Forest as the classifier. The experiment, performed on 20 patients, yielded a Dice
117 score of 97.7%. However, these conventional methods required many preprocessing steps before
118 entering the segmentation algorithm. The next evolution of ECF segmentations was performed
119 with a deep learning approach. Commandeur et al. [14] proposed ECF segmentation from non-

120 contrast coronary artery calcium computed tomography using ConvNets. They reported the Dice
121 score of 82.3%.

122   To improve the performance of medical image segmentation, several modifications of U-
123 Net [16] have been proposed. The spatial attention gate has been proposed to focus on the spatial
124 and detailed structure of the important region varying in shape and size [17]. Schlemper et al.
125 [17] demonstrate the performance of the attention U-Net on real-time fetal detection on 2D
126 images and pancreas detection on 3D CT images. He et al. [12] proposed ECF segmentation
127 from CE-CCTA using a modified 3D U-Net approach by adding attention gates (AG). AGs are
128 commonly used in classification tasks [18-21] and have been applied for various medical image
129 problems such as image classification [21, 22], image segmentation [17-22], and image
130 captioning [22]. AG are used to focus on the relevant portion of the image by suppressing
131 irrelevant regions [17]. The integration of AG into the standard U-Net [7, 11, 12, 17, 23] or V-
132 Net [11, 24] has been demonstrated to have benefits for region localization.

133   As mentioned above, the ECF has a complex structure. Some parts contain a thin layer
134 adjacent to the cardiac muscle, which is similar to the microvasculature of the retinal vascular
135 image visualized as small linear structures. In order to improve the performance of segmentation
136 of small structures, several modules have been integrated into the main architecture of U-Net and
137 V-Net, such as dense-layer and deep supervision modules [23-27]. The dense-layer [25, 26] has
138 been used to enhance the segmentation result instead of the traditional convolution in the U-Net
139 model. Deep supervision [23, 24, 27] was used to avoid local minimal traps during the training.
140 The deep supervision helps to improve model convergence and increases the number of learned
141 features [23]. Kearney et al. [23] showed that addition of deep supervision added to the U-Net
142 model can improve the performance of 3D segmentation in CT images of the prostate gland,
143 rectum, and penile bulb.

144   While 2D and 3D deep learning approaches have been used for medical image
145 segmentation, 3D approaches have typically shown better performance than the 2D approaches
146 [9, 10, 28]. For example, Zhou et al. [10] demonstrated the better performance of 3D CNN
147 approaches on multiple organs on 3D CT images, when compared to the 2D based method.
148 Starke et al. [9] also demonstrated that 3D CNN achieved better performance on segmentation of
149 head and neck squamous cell carcinoma on CT images. Woo et al. [28] demonstrated that 3D U-
150 Net provided better performance on brain tissue MRI images, compared with 2D U-Net, on a
151 smaller training dataset. Therefore, in this paper we use a 3D CNN for segmenting epicardial fat
152 in cardiac CT images.

153
154

155 **Materials & Methods**

156 **CNN architecture**

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

160 original U-Net architecture, we added an attention gate connecting the encoding and decoding
161 paths and deep supervision at the final step of the network. The model is created on a fully 3D
162 structure at each network level. The final layer is an element-wise sum of feature maps of the
163 two last decoding paths. The segmentation map of two classes (epicardial fat and background) is
164 obtained using the output layer with threshold 0.5 to generate the binary classification of the
165 epicardial fat. The architecture of the proposed network is shown in Fig 2.
166   Starting with the standard 3D U-Net architecture, the attention gate module connects each
167 layer of encoding and decoding paths. The gating signal (g) is chosen from the encoding path
168 and the input features (x) are collected from the decoding path. To generate the attention map, g
169 and x go through a 1x1x1 convolution layer and element-wise sum, followed by rectified linear
170 unit (ReLu) activation, a channel-wise 1x1x1 convolutional layer, batch normalization and a
171 sigmoid activation layer. The output of sigmoid activation is concatenated to the input x to get
172 the output of the attention gate module [12, 23].
173   Deep supervision [11, 24] is the module at the final step of the network where it
174 generates the multiple segmentation maps at different resolution levels, which are then combined
175 together. The secondary segmentation maps are created from each level of decoding paths which
176 are then transposed by 1x1x1 convolution. All feature maps are combined by element-wise sum.
177 The lower resolution map is upsampled by 3D transposed convolution to have the same size as
178 the second-lower resolution. Two maps are combined with element-wise sum then upsampled
179 and added to the next level of segmentation map, until reaching the highest resolution level.
180

CT imaging data

182   This experimental study was approved and participant consent was waived by the
183 institutional review board of Siriraj Hospital, Mahidol University (certificate of approval number
184 Si 766/2020). The experimental datasets were acquired from 220 patients with non-contrast
185 enhanced calcium scoring and 40 patients with CE-CCTA. The exclusion criteria were post open
186 surgery of the chest wall. All CT acquisition was performed with the 256-slice multi-detector
187 row CT scanner (Revolution CT; GE Medical Systems, Milwaukee, Wisconsin, United States).
188 The original CT datasets of NCCT and CECT studies were 64 slices in 2.5 mm slice thickness
189 and 256 slices in 0.625 mm slice thickness, respectively. All DICOM images were incorporated
190 into a single 3D CT volume file with preserved original pixel intensity. Due to limitation of GPU
191 memory, the 256 slices of CE-CCTA were pre-processed with rescaling to 64 images in the
192 volume dataset. The final 3D volume dataset in all experiments was 512x512x64. The dataset
193 was raw 12 bits grayscale in each voxel. The area of pericardial fat was defined by fat tissue
194 attenuation inside the pericardium, ranging from -200 HU to -30 HU [15, 29, 30]. The ground-
195 truth segmentation of ECF in all axial slices was performed using the 3D slicer software version
196 4.10.0 [31] by a cardiovascular radiologist with 18 years of experience. No additional feature
197 map or augmentation was performed in the pre-processing step.
198
199

200   **Training framework**

201   The experiments were implemented using the PyTorch (v1.8.0) deep learning library in
202   Python (v3.6.9). The workflow for network training is illustrated in Fig 3. The training and
203   testing processes were performed on a CUDA-enabled GPU (Nvidia DGX-A100) with 40 GB
204   RAM. The experiments were divided into three scenarios: model validity assessment, NCCT,
205   and CECT experiments. The parameters were the same for all three experiments. The networks
206   were trained with RMSprop optimizer and mean squared error loss. The training parameters of
207   learning rate, weight decay, and momentum were le-3, le-8 and 0.9, respectively. The initial
208   random seed was set to be 0. The illustration of the experimental framework is shown in Fig 3.
209   The first experiment was the assessment of the model validity, for which we used 5-fold
210   cross validation. The total dataset consisted of 200 volume-sets (12,800 images), divided into
211   five independent folds. Each fold contained 160 volume-sets (10,240 images) for training and 40
212   volume-sets (2,560 images) for validation, without repeated validation data between folds. The
213   other 20 volume-sets (1,280 images) were left for testing in the second and third experiments.
214   Then the 5-fold cross validation was performed on standard U-Net, AG-U-Net, DSV-U-Net, and
215   the proposed method (AG-DSV-U-Net). For each fold of validation, the model with the best
216   training accuracy after 150 epochs was selected for the validation.
217   The second experiment was to assess segmentation performance by training the network
218   from scratch with the NCCT dataset. The volume matrix of each dataset was 512x512x64 pixels.
219   To compare the performance of segmentation, this experiment was performed with four model
220   architectures: standard U-Net, AG-U-Net, DSV-U-Net, and the proposed method (AG-DSV-U-
221   Net). The network was evaluated with the hold-out method, in which a total of 220 volume-sets
222   (14,080 images) were split into 200 volume-sets (12,800 images) for training (the same used in
223   the first experiment) and 20 volume-sets (1,280 images) for testing. The model output was
224   collected at the maximum number of iterations at the 300th epoch, named model-A.
225   The third experiment was to assess segmentation performance on the CECT dataset and
226   to evaluate the effectiveness of transferring the learning from NCCT to the CECT datasets. The
227   pre-trained 3D model (model-A) was trained on the large calcium scoring NCCT datasets. The
228   key to the success of the transfer learning with 3D U-Net is to fine-tune only the shallow layers
229   (contracting path) [32] instead of the whole network. This contracting path represents the more
230   low-level features in the network [32]. The retraining dataset requires only a small amount of
231   data - in our case only 20 volume-sets of CECT data. Note that these retraining cases are not
232   from the same cases as used in the pre-trained model. To compare the performance of
233   segmentation, this experiment was performed with four model architectures: standard U-Net,
234   AG-U-Net, DSV-U-Net and proposed method (AG-DSV-U-Net). The network was evaluated
235   with the hold-out method, in which   the total 40 volume-sets (2,560 images) were split into 20
236   volume-sets (1,280 images) for training and 20 volume-sets (1,280 images) for testing. The
237   output model is collected at the maximum number of iterations at of 300thh epoch, named
238   model-B.

239

240

Performance evaluation

241

242   The performance of our proposed CNN segmentation is compared with the performance
243   of the existing methods. The evaluation was quantitatively performed by comparison with the
244   reference standard using the Dice similarity coefficient (DSC), Jaccard similarity coefficient
245   (JSC), and Hausdorff distance (HD). An average HD value was calculated using the insight
246   toolkit library of 3D slicer. Differences in the comparison coefficient among the four groups of
247   experiments (standard U-Net, AG-U-Net, DSV-U-Net, and AG-DSV-U-Net) were assessed with
248   a paired Student's t-test. P values <0.05 indicated a statistically significant difference.
249   Differences in the comparison between DSC of segmentation result and ECF volume were
250   assessed with Pearson's correlation coefficient. The Pearson's values of < 0.3 indicated poor
251   correlation, 0.3 to 0.7 indicated moderate correlation, and > 0.7 indicated good correlation. To
252   assess the consistency of the ground-truth, the inter-rater agreement was obtained by pixel-based
253   correlation of DSC using two experienced cardiovascular radiologists with 18 and 14 years of
254   experience. Ten volume sets (640 images) were randomly selected from NCCT dataset for the
255   process of inter-rater correlation.

256

257

Results

258

259   The patient demographics are shown in Table 1. The training dataset consisting of 200
260   NCCT scan had an average age of 61.43 years and an average volume of 135.75 ml. The NCCT
261   testing dataset had a similar distribution, with an average age of 67.80 years and an average
262   volume of 127.59 ml. For the contrast-enhanced dataset, the average ages of the training and
263   testing datasets were 65.85 and 60.85 years, respectively. The average volumes of ECF of the
264   training and testing datasets were 117.13 and 121.43 ml, respectively. The inter-rater agreement
265   of ECF ground-truth by two cardiovascular radiologists is about 91.73.±1.27%,which indicates
266   excellent correlation.
267   Five-fold cross validation experiments on our NCCT dataset were used to evaluate the
268   validity and repeatability performance of the proposed method. The dataset was split into
269   training (80%) and validation (20%) for each fold. On each model architecture, the validation
270   data exhibits good results across each fold. The proposed method also demonstrates the best
271   average performance (DSC = 89.02), when compared with the other methods (at $p<0.05$ when
272   using validation data obtained in the five-fold cross-validation in the t-test), see Table 2.
273   The experimental result on the hold-out NCCT test dataset is shown in Table 3. The
274   proposed method demonstrates excellent results, achieving average DSC, JSC, HD values of
275   90.06±4.60, 82.42±6.91 and 0.25±0.14, respectively. The baseline of the experiment is the
276   standard 3D U-Net, which demonstrates good results with DSC, JSC and HD values of
277   84.87±5.73, 74.12±8.0.8 and 0.34±0.18, respectively. The segmentation results of the modified
278   U-Net models (AG-U-Net, DSV-U-Net, and the proposed method) demonstrate statistically
279   significant improvement compared with the standard U-Net based on DSC ($p<0.05$). The

280 difference between segmentation results of AG-U-Net and DSV-U-Net is not statistically
281 significant (p>0.05). The DSC, JSC, HD values of AG-U-Net are 89.59+4.45, 81.41+6.77 and
282 0.27+0.12, respectively. The proposed method statistically improved the segmentation result
283 (p<0.05) compared with AG-U-Net as well. While the proposed method is better than DSV-U-
284 Net, the difference is not statistically significant (p>0.05). Examples of segmentation results of
285 the proposed method are shown in Fig 4.
286   The experimental result of the proposed method on the CECT dataset is shown in Table
287 4. This transfer learning approach achieved average DSC, JSC and HD values of 88.16±4.57,
288 79.10±6.75 and 0.28±0.20, respectively. The segmentation result of the proposed method
289 demonstrates statistically significant improvement, when compared with the other methods based
290 on DSC (p<0.0.5). The segmentation results of the standard 3D U-Net and DSV-U-Net
291 demonstrate good similar performance (p>0.05). The segmentation results of the standard 3D U-
292 Net and DSV-U-Net are statistically significantly better than those of AG-U-Net (p<0.05).
293 Examples of segmentation results of transfer learning with the proposed methods are shown in
294 Fig 5.
295
296

Discussion

298   Segmentation of ECF is a difficult image segmentation task because of the thin layer and
299 complex structures at the outer surface of the heart. The ECF is also variable in distribution,
300 depending on body habitus. In general, obese patients have larger amounts of ECF than do thin
301 patients. Segmentation of ECF is more challenging than the segmentation of other cardiac
302 structures.
303   Most CNN approaches work on 2D images whereas in clinical practice, 3D volume
304 segmentation is used [33]. The 2D-based CNN approaches such as ResNet and VGG are not
305 applicable for 3D datasets. The model architectures for 2D CNN and 3D CNN are different [9,
306 10, 28]. 3D CNN has an advantage over 2D-CNN by extracting both spectral and spatial
307 features simultaneously, while 2D CNN can extract only spatial features from the input data [5].
308 For this reason in general, the 3D CNN is more accurate than a 2D one [5, 34]. 2.5D CNN has
309 been developed to solve the memory consumption problem of 3D models [35]. 2.5D CNN has at
310 least three approaches [35, 36]. The first is a combination of outputs of 2D CNNs in three
311 orthogonal planes (axial, coronal, and sagittal) with majority voting. The second is to use 2D
312 CNNs with 3 or 5 channels from adjacent 3 or 5 slices. Third is to apply 2D CNNs with
313 randomly oriented 2D cross sections. In the final step, 2.5D segmentation requires an additional
314 post-processing step to generate 3D output [37]. Although, the 3D CNN requires more resources
315 and time for the model training, for the best performance, we use the 3D CNN in our
316 implementation. The best performing methods for 3D volume segmentation of medical data are
317 U-Net and V-Net. V-Net has more trainable parameters in its network architecture. Recent
318 experimental comparisons of U-Net and V-Net on medical data have not shown statistically
319 significant differences in performance [12, 38]. However, U-Net is less complex and easier to

320 modify so that additional modules can be integrated into the standard U-Net in order to improve
321 the performance.
322   Several state-of-the-art approaches for CNN-based segmentation of ECF have recently
323 been proposed. Commandeur et al. [14] proposed the first CNN-based method for ECF
324 segmentation in a non-contrast 2D CT dataset using a multi-task convolutional neural network
325 called ConvNets. They reported a Dice score of 82.3% for the segmentation result. He et al. [12]
326 proposed another CNN-based method on a 3D CECT dataset using AG integrated into 3D U-
327 Net. The segmentation result was reported to have a DSC of 88.7% [12]. We repeated the
328 experiment by implementing the AG in 3D U-Net on our NCCT dataset by hold-out testing. The
329 amount of training data (200 volume-sets) is more than the one used in the previous article (150
330 volume-sets) [12]. Our 4 layer AG-U-Net method demonstrates significantly improved
331 performance with DSC of 89.59%, as compared with 3 layer AG-U-Net of 86.54% ($p < 0.05$).
332 That should be due to more layers of our network. In our implementation, the AG-U-Net has
333 more convolutional layers (4 layers) and removes the sigmoid at the end of the network.
334 However, the AG integration provides significantly better performance ($p < 0.05$) as compared
335 with standard 3D U-Net (DSC of 84.87%). To the best of our knowledge, our experiment uses
336 the largest volume size of any 3D CT dataset (512x512x64).
337   We introduce a novel approach to 3D segmentation of ECF by integrating both AG and
338 DSV modules into all layers of the 3D U-Net deep learning architecture. The AGs are commonly
339 used in natural image analysis and natural language processing [39, 40], which can generate
340 attention-awareness features. The AG module is beneficial for organ localization, which can
341 improve organ segmentation [17]. During CNN training, AG is automatically learned to focus on
342 the target without additional supervision [41]. The AG module can improve model accuracy by
343 suppressing feature activation in irrelevant regions of an input image [17]. The AG module is
344 used to make connections between encoding and decoding paths on the standard U-Net. The
345 DSV module is used to deal with the vanishing gradient problem in the deeper layers of a CNN
346 [11, 27]. The standard approach provides the supervision only at the output layer. But the DSV
347 module propagates the supervision back to the earlier layer by generating a secondary
348 segmentation map combining information from different resolution levels. The losses of this
349 segmentation map are weighted and added to the final loss function and this can effectively
350 increase the performance [42]. The DSV module is used by adding it into the decoding path of
351 the 3D U-Net. The AG-DSV modules had been implemented in previous work [11, 24] for
352 kidney [24] tumor segmentation (Kidney Tumor Segmentation Challenge 2019), as well as for
353 liver [11] and pancreas [11] tumor segmentation (Medical Decathlon Challenge 2018).
354   The experiment demonstrated that our proposed method (AG-DSV-U-Net) achieves
355 excellent performance with average and max DSC values of 90.06% and 95.32%, respectively.
356 Our proposed method also shows a significant improvement of performance (90.06%), when
357 compared with the previous state-of-the-art network (86.54%) on the same dataset ($p < 0.05$). The
358 example of the results was shown in Fig 4. While one might expect the segmentation
359 performance to improve with fat volume of the dataset, unexpectedly, the statistical analysis

360 demonstrates that there is poor correlation between segmentation performance and fat volume
361 (Pearson's correlation 0.2).
362   The 3D volume size of our dataset is larger (512x512x64) compared with the previous
363 work (512x512x32) [12]. For comparative analysis on different numbers of slices and image
364 resolution of datasets, the previous work demonstrates that a 40-slice of volume dataset achieves
365 1% higher DSC than 32-slice and 24-slice [12]. However, the training time is also increased. We
366 set up additional experiments to test the effect of different numbers of slices and image
367 resolution on segmentation performance with our proposed model (AG-DSC-U-Net). The
368 training, testing datasets, and hyperparameters are the same as defined in the NCCT experiment.
369 The experiment of effect of the number of slices was performed by rescaling the slices with 64,
370 32 and 16 slices and using a fixed image resolution with 512x512 pixels. The segmentation
371 results are DSC 90.06%, 81.76%, 78.93%, respectively. The experiment to determine the effect
372 of different image resolution scales was performed by rescaling the resolution with 512x512,
373 256x256 and 128x128 pixels, with the number of slices fixed at 64. The segmentation results are
374 DSC 90.06%, 86.19% and 83.73% respectively. The 512x512 image resolution and 64 slices
375 still give the best performance, with significant improvement over lower resolution (p<0.05).
376 More slices and higher image resolution of the dataset let the network extract more spatial
377 information that can help to improve segmentation accuracy. Furthermore, because the ECF is a
378 thin layer along the heart contour, more spatial resolution will improve segmentation accuracy.
379 To give the best performance, we chose 64 slices for our implementation which is a perfect fit
380 with the original NCCT dataset, having 64 images in each dataset. In the CECT, the original CT
381 dataset had 256 slices and needed to be rescaled to 64 slices. Due to the limitation of the
382 proposed model and current GPU architecture, the voxel size of train and test datasets cannot be
383 extended beyond 64 slices. The other limitation of this study is the size of the dataset: 220
384 volume-sets for NCCT experiment and 40 volume-sets for CECT experiment.
385   In clinical practice, the cardiac CT scan can be performed in NCCT or CECT or using
386 both methods. For this reason, the ECF can also be either segmentation from NCCT or CECT
387 dataset. To the best of our knowledge, ours is the first implementation of ECF segmentation on
388 NCCT and CECT datasets. In our experiment, we started to train with the NCCT dataset (200
389 volume-sets). We used the concept of transfer learning to re-train with a similar dataset by taking
390 a small amount of the dataset (Fig 3). We re-trained the pre-trained NCCT model with a small
391 amount of CECT data (20 volume-sets). Weed test the model with additional testing of 20
392 volume-sets. The experimental result shows good performance with a DSC value of 88.16%. The
393 performance is also significantly better than that obtained with the standard U-Net and AG-U-
394 Net (Table 4). Training from scratch is more generally applicable but requires a large amount of
395 training samples. Both NCCT and CECT datasets are similar in appearance but differ in color.
396 The training from scratch approach of both datasets required a large amount of data samples. The
397 transfer learning method is more practical because it uses smaller amounts of training data and
398 still yields good performance. Additionally, our proposed re-trained model demonstrates good
399 performance as compared with the previous training from scratch (88.7%) [12].

400   Future studies could include investigations into more data diversity from multiple CT
401 vendors, larger patient variation, and testing the model across different healthcare centers.
402 Further investigation of clinical correlation between CNN segmentation of ECF volume and
403 occurrence of cardiovascular disease is another interesting research questions.
404
405

Conclusions

407   We have introduced a CNN-based approach for ECF segmentation using integration of
408 AG and DSV modules into the standard 3D U-Net. ECF segmentation is one of the most difficult
409 medical image segmentation tasks. We trained the NCCT dataset from scratch and also re-trained
410 on a CE CT dataset from the pre-trained NCCT model. We successfully improved the
411 performance of ECF segmentation on both NCCT and CECT datasets when compared with the
412 previous state-of-the-art methods. It is expected that this proposed method has the potential to
413 improve the performance in other difficult segmentation tasks. The concept of training and
414 retraining models can be also applied to other medical image segmentation problems.
415
416

Acknowledgements

418   Thank you to Dr.Jitladda Wasinrat who is the second cardiovascular radiologist for
419 analysis of inter-rater agreement. This work was partially supported by a grant from the Mahidol
420 University Office of International Relations to Haddawy in support of the Mahidol-Bremen
421 Medical Informatics Research Unit.

422
423
424

References

426 [1]. Bertaso AG, Bertol D, Duncan BB, Foppa M. Epicardial fat: definition, measurements and
427 systematic review of main outcomes. Arq Bras Cardiol. 2013;101(1):e18-e28.
428 [2]. Rodrigues ÉO, Rodrigues LO, Oliveira LSN, Conci A, Liatsis P. Automated recognition of
429 the pericardium contour on processed CT images using genetic algorithms. Computers in
430 Biology and Medicine. 2017;87:38-45.
431 [3]. Militello C, Rundo L, Toia P, Conti V, Russo G, Filorizzo C, Maffei E, Cademartiri F, La
432 Grutta L, Midiri M, Vitabile S. A semi-automatic approach for epicardial adipose tissue
433 segmentation and quantification on cardiac CT scans. Computers in Biology and Medicine.
434 2019;114.
435 [4]. Zlokolica V, Krstanović L, Velicki L, Popović B, Janev M, Obradović R, Ralević NM,
436 Jovanov L, Babin D. Semiautomatic Epicardial Fat Segmentation Based on Fuzzy c-Means
437 Clustering and Geometric Ellipse Fitting. Journal of Healthcare Engineering. 2017;2017.
438 [5]. Singh SP, Wang L, Gupta S, Goli H, Padmanabhan P, Gulyás B. 3D Deep Learning on
439 Medical Images: A Review. Sensors. 2020;20(18).
440 [6]. Kim M, Yun J, Cho Y, Shin K, Jang R, Bae H-J, Kim N. Deep Learning in Medical
441 Imaging. Neurospine. 2019;16(4):657-68.

442 [7]. Hesamian MH, Jia W, He X, Kennedy P. Deep Learning Techniques for Medical Image
443 Segmentation: Achievements and Challenges. Journal of Digital Imaging. 2019;32(4):582-96.
444 [8]. Renard F, Guedria S, Palma ND, Vuillerme N. Variability and reproducibility in deep
445 learning for medical image segmentation. Scientific Reports. 2020;10(1):13724.
446 [9]. Starke S, Leger S, Zwanenburg A, Leger K, Lohaus F, Linge A, Schreiber A, Kalinauskaite
447 G, Tinhofer I, Guberina N, Guberina M, Balermpas P, von der Grün J, Ganswindt U, Belka C,
448 Peeken JC, Combs SE, Boeke S, Zips D, Richter C, Troost EGC, Krause M, Baumann M, Löck
449 S. 2D and 3D convolutional neural networks for outcome modelling of locally advanced head
450 and neck squamous cell carcinoma. Sci Rep. 2020;10(1):15625.
451 [10]. Zhou X. Automatic Segmentation of Multiple Organs on 3D CT Images by Using Deep
452 Learning Approaches. In: Lee G, Fujita H, editors. Deep Learning in Medical Image Analysis :
453 Challenges and Applications. Cham: Springer International Publishing; 2020. p. 135-47.
454 [11]. Turečková A, Tureček T, Komínková Oplatková Z, Rodríguez-Sánchez A. Improving CT
455 Image Tumor Segmentation Through Deep Supervision and Attentional Gates. Frontiers in
456 Robotics and AI. 2020;7(106).
457 [12]. He X, Guo BJ, Lei Y, Wang T, Fu Y, Curran WJ, Zhang LJ, Liu T, Yang X. Automatic
458 segmentation and quantification of epicardial adipose tissue from coronary computed
459 tomography angiography. Physics in Medicine and Biology. 2020;65(9).
460 [13]. Rodrigues ÉO, Pinheiro VHA, Liatsis P, Conci A. Machine learning in the prediction of
461 cardiac epicardial and mediastinal fat volumes. Computers in Biology and Medicine.
462 2017;89:520-9.
463 [14]. Commandeur F, Goeller M, Betancur J, Cadet S, Doris M, Chen X, Berman DS, Slomka
464 PJ, Tamarappoo BK, Dey D. Deep Learning for Quantification of Epicardial and Thoracic
465 Adipose Tissue from Non-Contrast CT. IEEE Transactions on Medical Imaging.
466 2018;37(8):1835-46.
467 [15]. Rodrigues ÉO, Morais FFC, Morais NAOS, Conci LS, Neto LV, Conci A. A novel
468 approach for the automated segmentation and volume quantification of cardiac fats on computed
469 tomography. Computer Methods and Programs in Biomedicine. 2016;123:109-28.
470 [16]. Ronneberger O, Fischer P, Brox T, editors. U-Net: Convolutional Networks for
471 Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted
472 Intervention – MICCAI 2015; 2015 2015//; Cham: Springer International Publishing.
473 [17]. Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, Rueckert D. Attention
474 gated networks: Learning to leverage salient regions in medical images. Medical Image Analysis.
475 2019;53:197-207.
476 [18]. Wu Y, Zhang X, Xiao Y, Feng J. Attention Neural Network for Water Image
477 Classification under IoT Environment. Applied Sciences. 2020;10(3).
478 [19]. Sharmin S, Chakma D. Attention-based convolutional neural network for Bangla sentiment
479 analysis. AI & SOCIETY. 2021;36(1):381-96.
480 [20]. Fei H, Zhang Y, Ren Y, Ji D. Optimizing Attention for Sequence Modeling via
481 Reinforcement Learning. IEEE Transactions on Neural Networks and Learning Systems. 2021:1-
482 10.
483 [21]. Kelvin X, Jimmy B, Ryan K, Kyunghyun C, Aaron C, Ruslan S, Rich Z, Yoshua B. Show,
484 Attend and Tell: Neural Image Caption Generation with Visual Attention. 2015/06/01: PMLR;
485 2015. p. 2048-57.

486 [22]. Zhao B, Feng J, Wu X, Yan S. A survey on deep learning-based fine-grained object
487 classification and semantic segmentation. International Journal of Automation and Computing.
488 2017;14(2):119-35.
489 [23]. Kearney V, Chan JW, Wang T, Perry A, Yom SS, Solberg TD. Attention-enabled 3D
490 boosted convolutional neural networks for semantic CT segmentation using deep supervision.
491 Physics in Medicine & Biology. 2019;64(13):135001.
492 [24]. Tureckova A, Turecek T, Komínková Z, Rodríguez-Sánchez A, editors. KiTS challenge:
493 VNet with attention gates and deep supervision2019.
494 [25]. Liang Z, Liu H, Zhao X, Yu L. Segmentation of Retinal Vessels Based on DenseNet-
495 Attention-Unet Model Network. Proceedings of the 2nd International Conference on Industrial
496 Control Network And System Engineering Research; Kuala Lumpur, Malaysia: Association for
497 Computing Machinery; 2020. p. 111–7.
498 [26]. Wu C, Zou Y, Zhan J. DA-U-Net: Densely Connected Convolutional Networks and
499 Decoder with Attention Gate for Retinal Vessel Segmentation. IOP Conference Series: Materials
500 Science and Engineering. 2019;533:012053.
501 [27]. Chen-Yu L, Saining X, Patrick G, Zhengyou Z, Zhuowen T. Deeply-Supervised Nets.
502 2015/02/21: PMLR; 2015. p. 562-70.
503 [28]. Woo B, Lee M, editors. Comparison of tissue segmentation performance between 2D U-
504 Net and 3D U-Net on brain MR Images. 2021 International Conference on Electronics,
505 Information, and Communication (ICEIC); 2021 31 Jan.-3 Feb. 2021.
506 [29]. Shahzad R, Bos D, Metz C, Rossi A, Kirişli H, van der Lugt A, Klein S, Witteman J, de
507 Feyter P, Niessen W, van Vliet L, van Walsum T. Automatic quantification of epicardial fat
508 volume on non-enhanced cardiac CT scans using a multi-atlas segmentation approach. Medical
509 Physics. 2013;40(9):091910.
510 [30]. Kazemi A, Keshtkar A, Rashidi S, Aslanabadi N, Khodadad B, Esmaeili M, editors.
511 Automated Segmentation of Cardiac Fats Based on Extraction of Textural Features from Non-
512 Contrast CT Images. 2020 25th International Computer Conference, Computer Society of Iran,
513 CSICC 2020; 2020.
514 [31]. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, Bauer C,
515 Jennings D, Fennessy F, Sonka M, Buatti J, Aylward S, Miller JV, Pieper S, Kikinis R. 3D Slicer
516 as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging.
517 2012;30(9):1323-41.
518 [32]. Amiri M, Brooks R, Rivaz H. Fine-Tuning U-Net for Ultrasound Image Segmentation:
519 Different Layers, Different Outcomes. IEEE Transactions on Ultrasonics, Ferroelectrics, and
520 Frequency Control. 2020;67(12):2510-8.
521 [33]. Milletari F, Navab N, Ahmadi S, editors. V-Net: Fully Convolutional Neural Networks for
522 Volumetric Medical Image Segmentation. 2016 Fourth International Conference on 3D Vision
523 (3DV); 2016 25-28 Oct. 2016.
524 [34]. Han L, Chen Y, Li J, Zhong B, Lei Y, Sun M. Liver segmentation with 2.5D perpendicular
525 UNets. Computers & Electrical Engineering. 2021;91:107118.
526 [35]. Zhang C, Hua Q, Chu Y, Wang P. Liver tumor segmentation using 2.5D UV-Net with
527 multi-scale convolution. Computers in Biology and Medicine. 2021;133:104424.
528 [36]. Minnema J, Wolff J, Koivisto J, Lucka F, Batenburg KJ, Forouzanfar T, van Eijnatten M.
529 Comparison of convolutional neural network training strategies for cone-beam CT image
530 segmentation. Computer Methods and Programs in Biomedicine. 2021;207:106192.

531   [37]. Han Y, Li X, Wang B, Wang L. Boundary Loss-Based 2.5D Fully Convolutional Neural
532   Networks Approach for Segmentation: A Case Study of the Liver and Tumor on Computed
533   Tomography. Algorithms. 2021;14(5).
534   [38]. Ghavami N, Hu Y, Gibson E, Bonmati E, Emberton M, Moore CM, Barratt DC. Automatic
535   segmentation of prostate MRI using convolutional neural networks: Investigating the impact of
536   network architecture on the accuracy of volume measurement and MRI-ultrasound registration.
537   Medical Image Analysis. 2019;58:101558.
538   [39]. Wang F, Jiang M, Qian C, Yang S, Li C, Zhang H, Wang X, Tang X, editors. Residual
539   Attention Network for Image Classification. 2017 IEEE Conference on Computer Vision and
540   Pattern Recognition (CVPR); 2017 21-26 July 2017.
541   [40]. Anderson P, He X, Buehler C, Teney D, Johnson M, Gould S, Zhang L. Bottom-Up and
542   Top-Down Attention for Image Captioning and VQA. ArXiv. 2017;abs/1707.07998.
543   [41]. Oktay O, Schlemper J, Folgoc LL, Lee MJ, Heinrich M, Misawa K, Mori K, McDonagh
544   SG, Hammerla N, Kainz B, Glocker B, Rueckert D. Attention U-Net: Learning Where to Look
545   for the Pancreas. ArXiv. 2018;abs/1804.03999.
546   [42]. Kayalibay B, Jensen G, Smagt PVD. CNN-based Segmentation of Medical Imaging Data.
547   ArXiv. 2017;abs/1701.03056.
548

Figure 1

The example CT dataset of epicardial fat

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

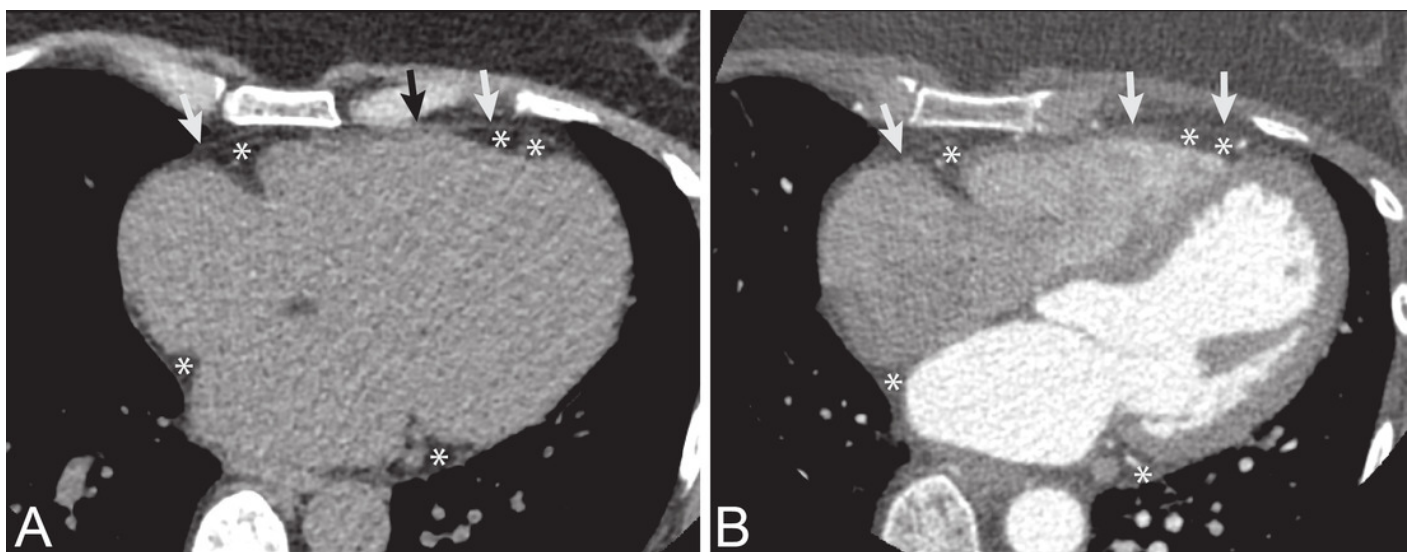



Figure 2

The proposed network for 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.

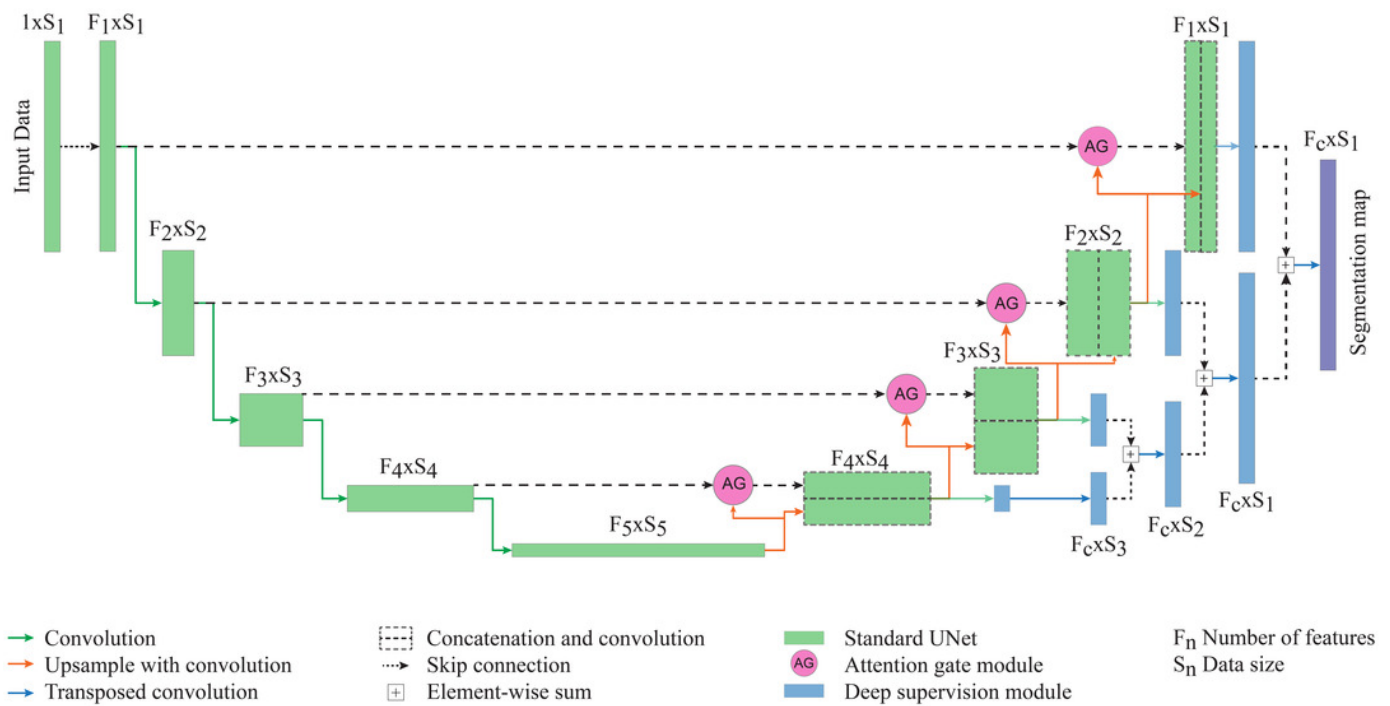



Figure 3

Illustration of framework of the proposed method.

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

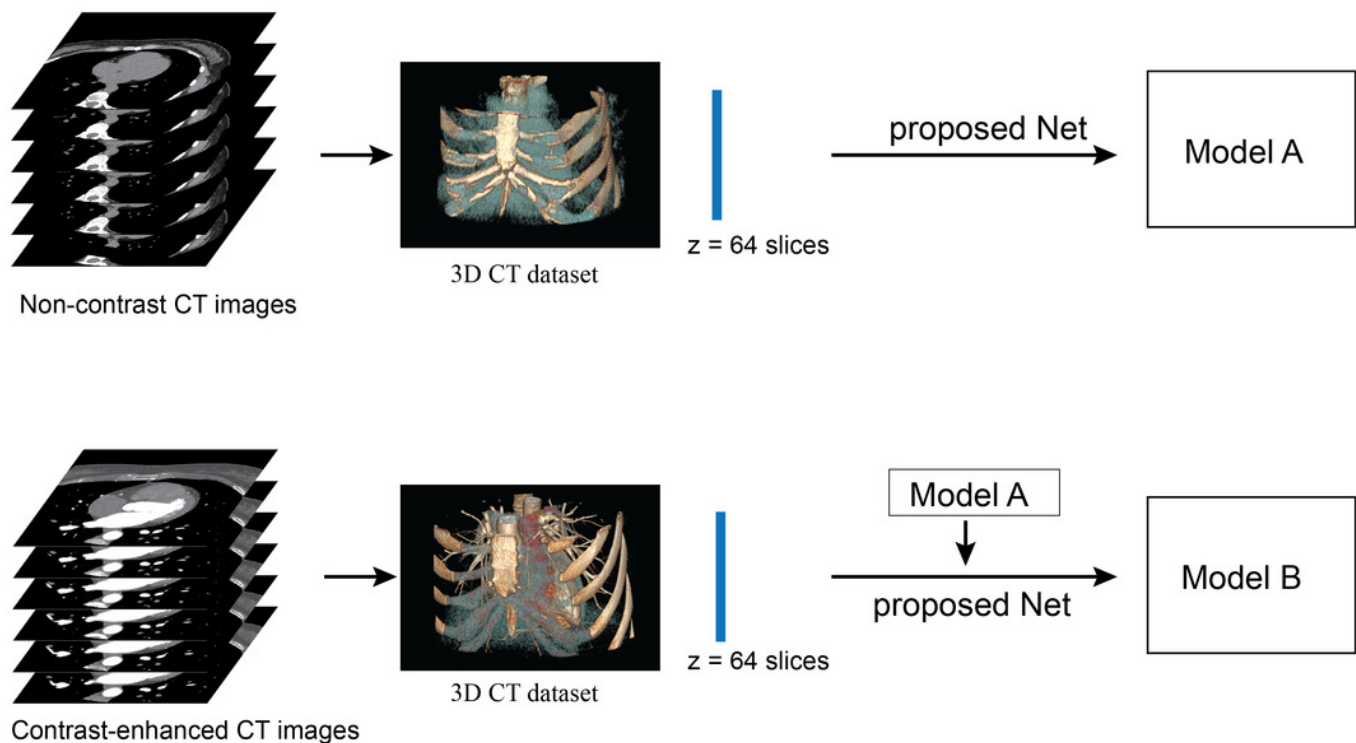



Figure 4

Fig 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, respectively. The yellow color represents the segmented ECF using the proposed method and green color represents the ground-truth. The DSC values for the volume sets from left to right are 90.90%, 88.63%, 95.31% and 87.86%, respectively. The fat volumes from left to right are 95.69 ml, 116.95 ml, 106.99 ml and 91.18 ml, respectively.

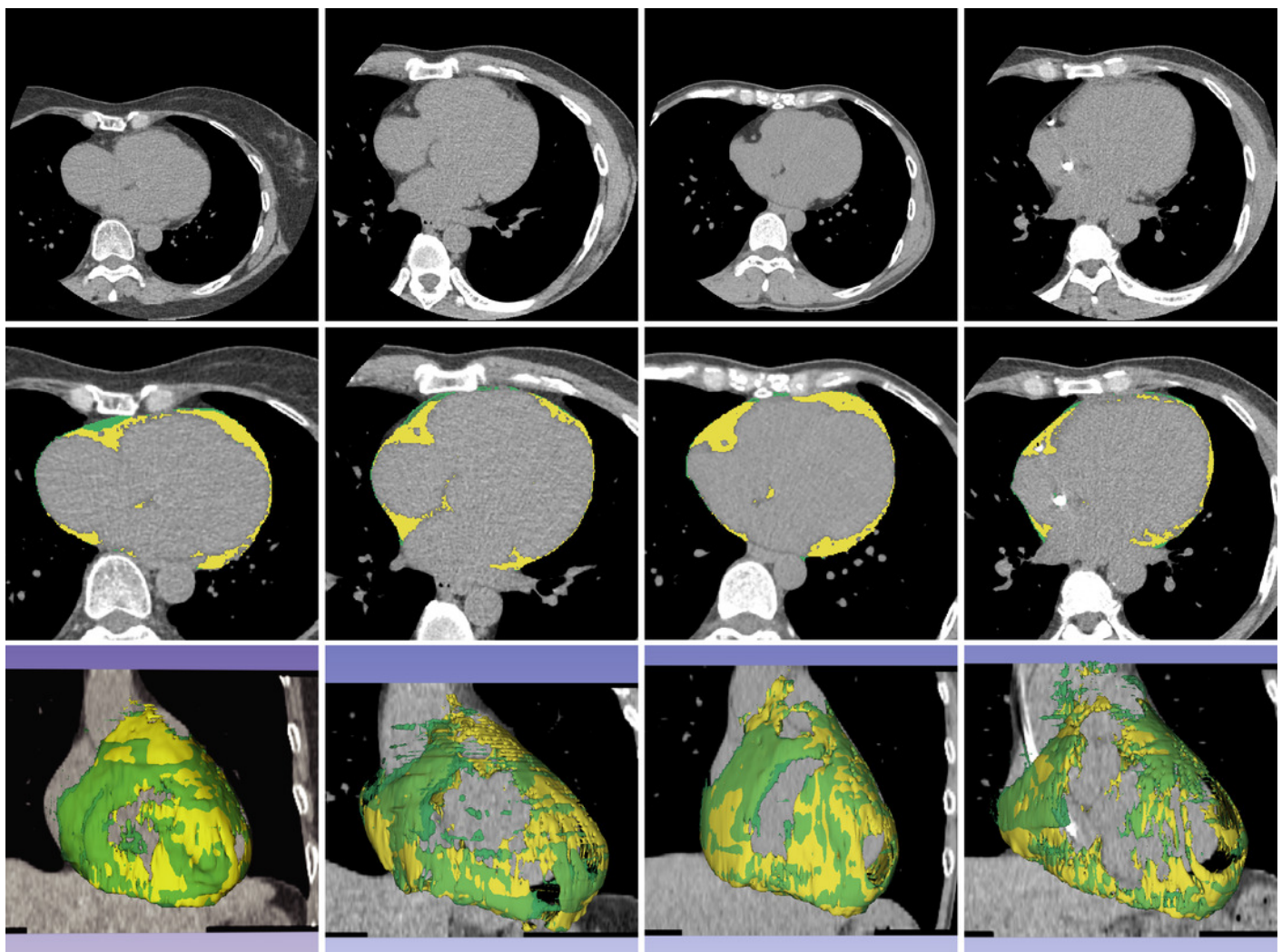
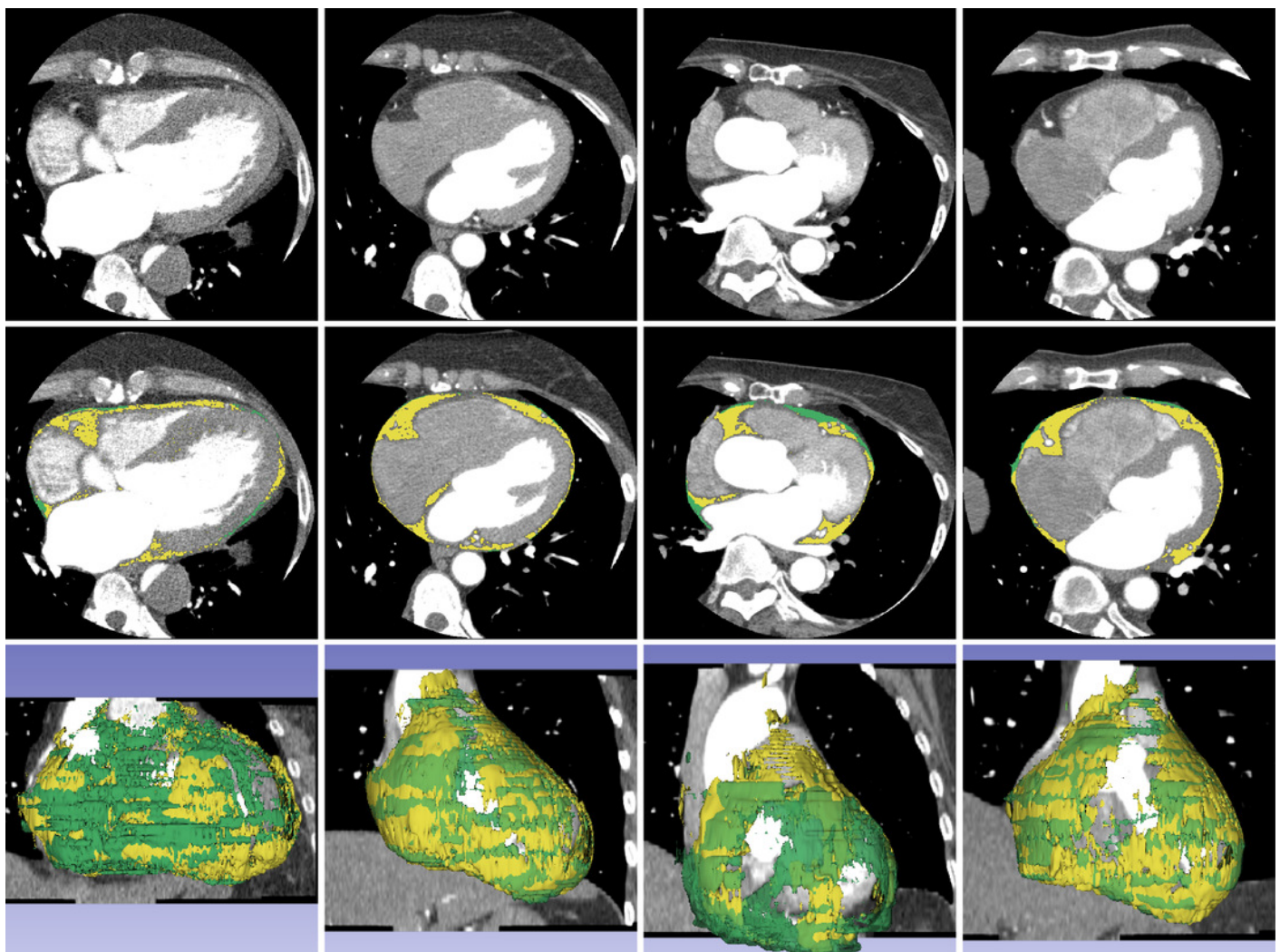



Figure 5

Fig 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 for the volume sets 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.

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

1   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

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

2   Table 2. Results of 5-fold cross-validation. Mean Dice score coefficient and standard deviation
3   were used to assess model validity and repeatability on the 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. Experimental results with standard 3D U-Net, AG-U-Net and the proposed method (AG-DSV-U-Net) on the holdout non-contrast CT dataset

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

1   Table 3. Experimental results with standard 3D U-Net, AG-U-Net and the proposed method (AG-
2   DSV-U-Net) on the holdout 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. Transferred learning from pre-trained model to contrast-enhanced CT dataset

2   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