technologies

MDPI

Article

The U-Net Family for Epicardial Adipose Tissue Segmentation and Quantification in Low-Dose CT

Lu Liu [1], Runlei Ma [2,3], Peter M. A. van Ooijen [2], Matthijs Oudkerk [4], Rozemarijn Vliegenthart [2], Raymond N. J. Veldhuis [1,5] and Christoph Brune [1,*]

[1] Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, 7522 NB Enschede, The Netherlands; l.liu-2@utwente.nl (L.L.); r.n.j.veldhuis@utwente.nl (R.N.J.V.)

[2] Department of Radiology, University Medical Center Groningen, 9713 GZ Groningen, The Netherlands; r.ma@umcg.nl (R.M.); p.m.a.van.ooijen@umcg.nl (P.M.A.v.O.); r.vliegenthart@umcg.nl (R.V.)

[3] Department of Radiology, Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing 210023, China

[4] Faculty of Medical Sciences, University of Groningen, 9712 CP Groningen, The Netherlands; m.oudkerk@rug.nl

[5] Department of Information Security and Communication Technology, Norwegian University of Science and Technology, 7034 Trondheim, Norway

* Correspondence: c.brune@utwente.nl

Abstract: Epicardial adipose tissue (EAT) is located between the visceral pericardium and myocardium, and EAT volume is correlated with cardiovascular risk. Nowadays, many deep learning-based automated EAT segmentation and quantification methods in the U-net family have been developed to reduce the workload for radiologists. The automatic assessment of EAT on non-contrast low-dose CT calcium score images poses a greater challenge compared to the automatic assessment on coronary CT angiography, which requires a higher radiation dose to capture the intricate details of the coronary arteries. This study comprehensively examined and evaluated state-of-the-art segmentation methods while outlining future research directions. Our dataset consisted of 154 non-contrast low-dose CT scans from the ROBINSCA study, with two types of labels: (a) region inside the pericardium and (b) pixel-wise EAT labels. We selected four advanced methods from the U-net family: 3D U-net, 3D attention U-net, an extended 3D attention U-net, and U-net++. For evaluation, we performed both four-fold cross-validation and hold-out tests. Agreement between the automatic segmentation/quantification and the manual quantification was evaluated with the Pearson correlation and the Bland–Altman analysis. Generally, the models trained with label type (a) showed better performance compared to models trained with label type (b). The U-net++ model trained with label type (a) showed the best performance for segmentation and quantification. The U-net++ model trained with label type (a) efficiently provided better EAT segmentation results (hold-out test: DCS = 80.18 ± 0.20%, mIoU = 67.13 ± 0.39%, sensitivity = 81.47 ± 0.43%, specificity = 99.64 ± 0.00%, Pearson correlation = 0.9405) and EAT volume compared to the other U-net-based networks and the recent EAT segmentation method. Interestingly, our findings indicate that 3D convolutional neural networks do not consistently outperform 2D networks in EAT segmentation and quantification. Moreover, utilizing labels representing the region inside the pericardium proved advantageous in training more accurate EAT segmentation models. These insights highlight the potential of deep learning-based methods for achieving robust EAT segmentation and quantification outcomes.

Keywords: epicardial adipose tissue; deep neural networks; segmentation; low-dose computed tomography

check for updates

Citation: Liu, L.; Ma, R.; van Ooijen, P.M.A.; Oudkerk, M.; Vliegenthart, R.; Veldhuis, R.N.J.; Brune, C. The U-Net Family for Epicardial Adipose Tissue Segmentation and Quantification in Low-Dose CT. *Technologies* **2023**, *11*, 104. https://doi.org/10.3390/technologies11040104

Academic Editors: Yudong Zhang and Zhengchao Dong

Received: 20 June 2023
Revised: 31 July 2023
Accepted: 3 August 2023
Published: 5 August 2023

1. Introduction

Epicardial adipose tissue (EAT) is located between the pericardium and the myocardium [1]. It has various distributions and is commonly found on the heart surface, in the atrioventricular and interventricular grooves, in the right ventricle lateral wall, and near the coronary arteries [2]. In many studies, it has been considered a source of inflammatory mediators and cytokines, and EAT volume has been associated with coronary artery disease [3]. EAT volume has been evaluated as an imaging biomarker for the diagnosis of pathological states such as metabolic syndrome and visceral obesity [4–6].

Manual or semi-automatic measurement of EAT volume by radiologists, while feasible, proves to be time-consuming and impractical for extensive clinical practice. To address this limitation, automated EAT volume quantification methods have been proposed, aiming to provide efficient solutions. Most EAT volume quantification is typically carried out using cardiac computed tomography (CT) scans. However, in this study, our focus lies specifically on EAT segmentation and quantification in non-contrast low-dose CT (LDCT) scans. In particular, EAT exhibits an irregular and noncontinuous shape, accompanied by a lack of uniformity in spatial distribution. Additionally, the presence of other fat tissues, such as mediastinal fat, positioned outside the pericardium presents a challenge. In CT images, these structures can be visually similar to EAT and are located nearby. The thin layer of the pericardium serves as a crucial reference for distinguishing EAT from these similar structures. One of the primary motivations for focusing on LDCT is its reduced contrast on coronary structures and the different thickness compared to standard-dose CT scans. Consequently, the segmentation and quantification of EAT in LDCT pose greater challenges. By addressing these complexities and providing a robust and accurate EAT segmentation methodology for LDCT, we aim to enhance the clinical utility of automated EAT volume quantification in various medical settings. Such advancements hold the potential to streamline cardiac assessments and improve patient care, given the widespread usage of LDCT in clinical practice.

As for automatic EAT segmentation methods, some works [7–11] utilized non-contrast standard-dose CT or LDCT, while some works [12–15] utilized coronary CT angiography (CCTA), which requires a higher radiation dose with contrast to ensure adequate image quality while capturing fine details on the coronary arteries. Recently, there has also been work utilizing magnetic resonance images (MRI) [16]. However, many works provided insufficient information on the setting of the protocol for EAT label acquisition. Furthermore, the variation in automated quantification of EAT volume across different CT acquisition protocols remains an unanswered question. The influence of protocol settings on EAT segmentation and the resulting volume of EAT remains unclear. Access to implementation codes or datasets is crucial for the reproducibility and comparison of work with the state-of-the-art. Due to variations in data set-up among different studies, a strict comparison of reported accuracy values can be misleading. Therefore, the main objective of our research is to thoroughly evaluate and compare state-of-the-art segmentation methods using our extensive dataset comprising 154 low-dose CT scans. To assess the segmentation performance, we employ well-established metrics such as the Dice similarity coefficient (DSC), mean Intersection of Union (mIoU), sensitivity, and specificity. Additionally, considering the clinical relevance of EAT volume, we utilize Pearson correlation and Bland–Altman analysis to measure the interscan reproducibility of EAT measurement in low-dose CT scans [17]. Through this rigorous evaluation, we aim to contribute valuable insights into the efficacy of existing segmentation techniques and to improve the understanding of the assessment of EAT in the context of low-dose CT imaging.

1.1. Related Work

The early automatic EAT segmentation methods were based on classic mathematical models such as intensity-based approaches [18], region growing approaches, active contour approaches [19], and atlas-based approaches [20,21]. Later, some machine learning-based methods with handcrafted features [7,8,22–24] or with clustering algorithms [13,25] were proposed for EAT segmentation. However, some are rarely fully automatic. Most recently, deep learning (DL)-based approaches have shown success in many medical image segmentation tasks [26].

The rise of deep learning brings to EAT segmentation methods more automation, as it can learn intricate features directly from images in an end-to-end manner rather than designing and selecting features manually. With the available advanced computer hardware such as graphical processing units (GPUs), the state-of-the-art DL-based segmentation approaches have outperformed many previous traditional methods. The basis of DL-based segmentation is the convolutional neural network (CNN). A standard CNN consists of an input layer, functional hidden layers, and an output layer. Commonly used functional layers include the 2D or 3D convolutional operations, pooling layers (e.g., max-pooling), normalization layers (e.g., batch normalization), activation functions (e.g., rectified linear unit (ReLU)), transposed convolutional operations, and upsampling layers. The input of the network could be CTs in 2D or 3D format and the corresponding labels. The output of the network is a matrix where each element could be a probabilistic score or a set of category indexes. Generally, for medical image segmentation, the U-Net structure [27] and more advanced 3D U-Net [28], the V-Net [29], the attention U-Net [30] with attention gates, and U-Net++ [31] with dense connections and deep supervision [32] have together formed a U-Net family that has gained popularity and success. Many recent works on EAT segmentation were based on or inspired by the U-Net family.

A fully automated EAT segmentation model proposed by Commandeur et al. [33,34] utilized a multi-task fully convolutional neural network with a statistical shape model. Santini et al. [35] applied standard U-Net on EAT segmentation with a dataset of 119 CTs, and Zhang et al. [9] utilized two U-Nets and morphological layers to achieve better EAT segmentation results. One of the most recent works by He et al. [12] proposed a 3D U-Net architecture with attention gates (AG) and deep supervision for EAT segmentation. Attention gates play a crucial role in prioritizing essential regions within an image during the segmentation process, enhancing the model's accuracy. The attention mechanism is a widely utilized approach, especially in tasks such as cardiac image segmentation [36,37]. Many works showed quantitatively promising performance, but some were trained and evaluated with a very small dataset [7,9] or with private datasets [12,33,34]. The labeling protocol varies in various works. For labeling, commonly used labels include the region inside the pericardium, the contour/surface of the pericardium, pixel-wise EAT labels, and the contour/surface of EAT regions, etc.

1.2. Contributions

In this study, we comprehensively evaluated state-of-the-art segmentation methods for EAT using 154 non-contrast LDCT scans. The LDCT scans were part of the coronary artery calcium CT scans from the Risk Or Benefit IN Screening for CArdiovascular diseases (ROBINSCA) trial [38]. We provided two types of labels for the LDCT scans: (a) region inside the pericardium and (b) pixel-wise EAT labels. The LDCT scans were ECG-triggered and generated with low radiation doses, ensuring patient safety. We provide clear and sufficient information about the data and label acquisition protocol in Section 2.1. To achieve representativeness and diversity, we selected four methods from the U-Net family as the EAT segmentation models: 3D U-Net [28], 3D attention U-Net [30], U-Net++ [31], and the recent deep attention U-Net (DAU-Net) [12]. The model structures, along with the attention gates and deep supervision modules in these models, are meticulously explained in Section 2.2. For rigorous evaluation, we utilized four-fold cross-validation and hold-out tests in Section 3. The segmentation results were analyzed using the Dice similarity coefficient, the mean intersection of union, sensitivity, specificity, and visualization. Furthermore, we performed a quantitative analysis of EAT volume using the Pearson correlation and Bland–Altman analysis in Section 3.3. In Section 4, we delve into crucial aspects, including label types, domain knowledge, patch size, training time, deep supervision, and evaluation, to gain deeper insights into EAT segmentation using deep learning. Additionally, we propose potential future directions in the aspect of domain knowledge, data unification, benchmarking, and deep learning techniques, aiming to enrich the contributions of this research. Our comprehensive evaluation and in-depth discussions contribute valuable insights to the field of EAT segmentation, and we believe this work will significantly advance the application of deep learning techniques in cardiac image analysis and clinical practice.

2. Data and Methods

2.1. Data Set-Up

For method benchmarking, we randomly selected 160 samples from the Risk Or Benefit IN Screening for CArdiovascular Diseases (ROBINSCA) dataset [38]. Samples of patients with hernias were excluded from the study due to the significant alteration in the shape and location of the organs. Thus, 154 samples were used for the evaluation. The ROBINSCA trial included subjects from the general population, with men aged between 45 and 74 and women aged between 55 and 74. This is a multi-center dataset with CT screening performed at the Gelre Hospital (Apeldoorn), the Bronovo Hospital (The Hague), and the University Medical Center Groningen (Groningen). A second-generation dual-source computed tomography (DSCT) system (Somatom Flash, Siemens, Erlangen, Germany) was used in all examinations. The CT acquisition protocol was as follows: (1) pitch: 3.4, (2) tube voltage: 120 kVp, (3) tube current: 80 mAs, (4) rotation time: 280 ms, (5) collimation: 120 × 0.6 mm, (6) matrix: 512 × 512, (7) ECG triggering: prospective, 60% of the R-R interval during inspiratory breath-hold, (8) upper limit: below carina, (9) lower limit: apex/bottom edge heart. The CT reconstruction protocol was (1) slice thickness: 3.0 mm, (2) slice increment: 1.5 mm, (3) FOV: 250 mm, (4) kernel: b35f (sharp).

EAT is characterized as fatty-like tissue intensity positioned between the myocardium and the visceral pericardium, with intensity limits ranging from −190 HU to −30 HU [39]. Manual annotation of EAT in CT scans can be accomplished through two primary methods: (a) annotating the region within the pericardium and then applying HU value thresholding to extract adipose tissue [11], or (b) directly annotating the region of EAT and utilizing HU value thresholding to generate final labels [12]. Due to the use of low-dose CT in our study, the second annotation strategy can be particularly challenging due to the reduced visibility of fat tissue and increased noise. As such, to ensure feasibility and efficiency, we chose the first annotation strategy. The labels were annotated by an experienced radiologist using the open-source medical imaging processing software 3D Slicer 4.10.2 (https://www.slicer.org/) (accessed on 19 February 2019) [40]. Two kinds of label maps were obtained, as shown in Figure 1: (1) the region inside the pericardium (in the second column, colored in yellow), and (2) the EAT volumes (in the third column, colored in yellow). To reduce the workload, all the annotations were made in the axial view semi-automatically. The radiologist annotated the region inside the pericardium on some 2D slices, and the annotations in between were generated automatically with the 'fill between slices' effect of the 3D Slicer. Finally, the radiologist checked and corrected the generated annotations. To obtain the EAT, the thresholding of −190 HU to −30 HU was applied, and then the morphological operations of erosion and dilation were used to reduce the noise.

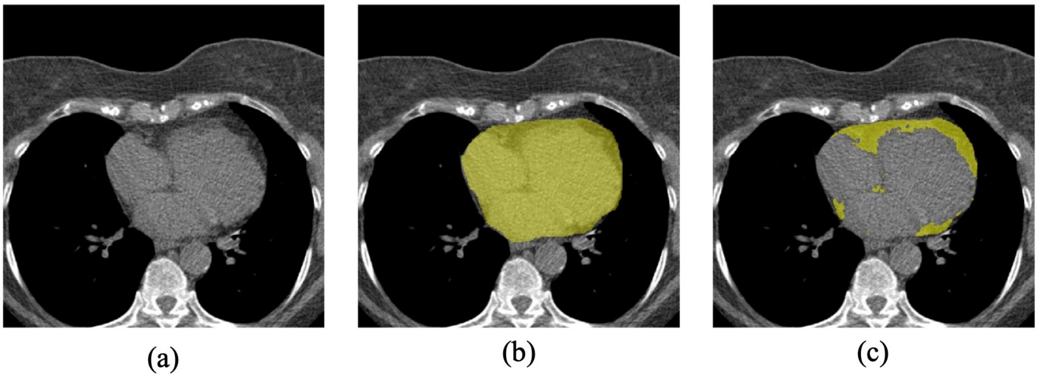


Figure 1. (**a**) An example of a CT slice, (**b**) the binary label of the region inside the pericardium, and (**c**) the pixel-wise EAT label.

2.2. Methods Description

2.2.1. The 3D U-Net

The original U-Net, proposed by Ronneberger et al. [27], is a U-shape CNN based on the fully convolutional network [41] for biomedical image segmentation. It consists of a downsampling path, an upsampling path, and skip connections in between. The downsampling path is a standard CNN with multiple convolution operations followed by a ReLU and max pooling, and the upsampling path combines the feature maps from deconvolution and concatenation. The use of data augmentation and the special architecture lets U-Net outperform the prior best methods in many medical image segmentation tasks. The 3D U-Net [28] extended the standard U-Net by using 3D convolution operations rather than 2D convolutions. Figure 2 illustrates the network architecture. In the downsampling path, each layer consists of two convolutions with a kernel size of 3 × 3 × 3 and a max-pooling layer with strides of two. In the upsampling path, each layer consists of an upsampling layer with a kernel size of 2 × 2 × 2 by strides of two, and two convolutions which are the same as the convolutions in the downsampling path. The skip connection concatenates the feature maps from the downsampling path to the upsampling path. In all layers, the ReLU and batch normalization (BN) are applied.

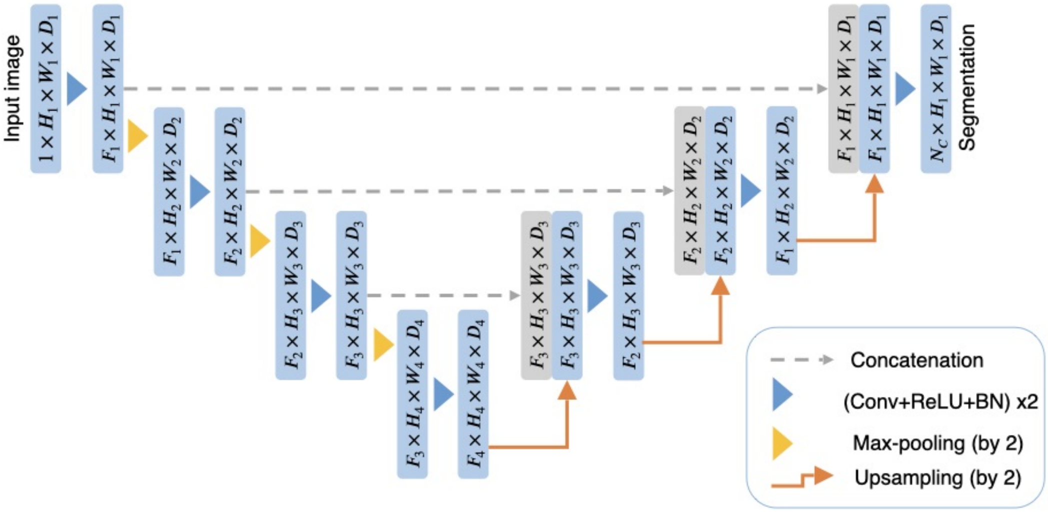

Input image
$1 \times H_1 \times W_1 \times D_1$
$F_1 \times H_1 \times W_1 \times D_1$
$F_1 \times H_2 \times W_2 \times D_2$
$F_2 \times H_2 \times W_2 \times D_2$
$F_2 \times H_3 \times W_3 \times D_3$
$F_3 \times H_3 \times W_3 \times D_3$
$F_3 \times H_4 \times W_4 \times D_4$
$F_4 \times H_4 \times W_4 \times D_4$
$F_3 \times H_3 \times W_3 \times D_3$
$F_3 \times H_3 \times W_3 \times D_3$
$F_2 \times H_3 \times W_3 \times D_3$
$F_2 \times H_2 \times W_2 \times D_2$
$F_2 \times H_2 \times W_2 \times D_2$
$F_1 \times H_2 \times W_2 \times D_2$
$F_1 \times H_1 \times W_1 \times D_1$
$F_1 \times H_1 \times W_1 \times D_1$
$N_C \times H_1 \times W_1 \times D_1$
Segmentation
Concatenation
(Conv+ReLU+BN) x2
Max-pooling (by 2)
Upsampling (by 2)

Figure 2. The 3D U-Net architecture. Blue boxes are feature maps, while gray boxes are concatenation maps. F_l denotes the number of feature maps in layer l; H denotes the height; W denotes the width; D denotes the number of dimensions; N_c denotes the number of classes. "Conv" is an abbreviation for convolution layer, "ReLU" stands for rectified linear unit, and "BN" is an abbreviation for batch normalization. The input image is downsampled by a factor of two at each level in the downsampling path and upsampled by a factor of two in the upsampling path.

2.2.2. The 3D Attention U-Net

The attention U-Net proposed by Oktay et al. [30] extended the 3D U-Net by adding grid-based attention gates (AGs) in the upsampling path. In the standard CNN architecture, to capture semantic contextual information, feature maps are progressively downsampled. In this way, the global features are well maintained, while the local features of small structures with large shape variations may be neglected. Thus, the AG was proposed to force the network to focus on local regions. Figure 3 illustrates the 3D attention U-Net architecture. The distinction between this network and the 3D U-Net in Figure 2 is the additional AGs in the upsampling path.

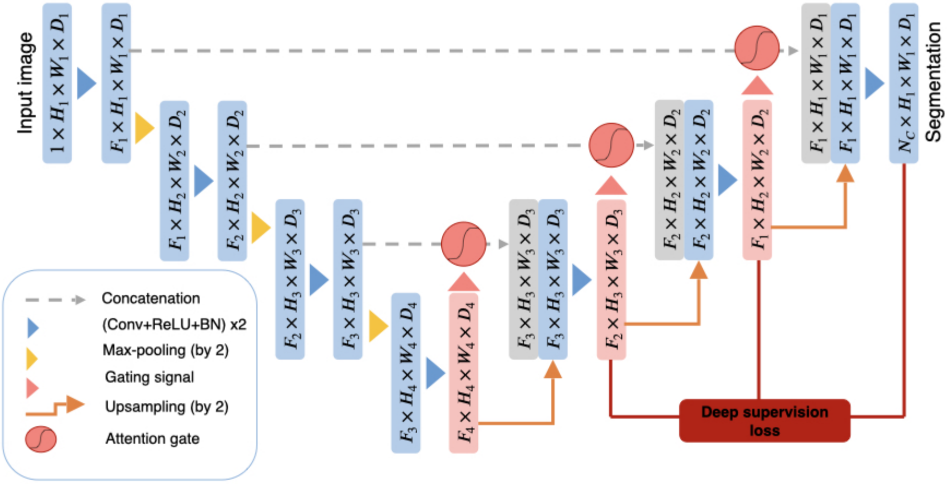

Input image
$1 \times H_1 \times W_1 \times D_1$
$F_1 \times H_1 \times W_1 \times D_1$
$F_1 \times H_2 \times W_2 \times D_2$
$F_2 \times H_2 \times W_2 \times D_2$
$F_2 \times H_3 \times W_3 \times D_3$
$F_3 \times H_3 \times W_3 \times D_3$
$F_3 \times H_4 \times W_4 \times D_4$
$F_4 \times H_4 \times W_4 \times D_4$
$F_3 \times H_3 \times W_3 \times D_3$
$F_3 \times H_3 \times W_3 \times D_3$
$F_2 \times H_3 \times W_3 \times D_3$
$F_2 \times H_2 \times W_2 \times D_2$
$F_2 \times H_2 \times W_2 \times D_2$
$F_1 \times H_2 \times W_2 \times D_2$
$F_1 \times H_1 \times W_1 \times D_1$
$F_1 \times H_1 \times W_1 \times D_1$
$N_C \times H_1 \times W_1 \times D_1$
Segmentation
Deep supervision loss
Concatenation
(Conv+ReLU+BN) x2
Max-pooling (by 2)
Gating signal
Upsampling (by 2)
Attention gate

Figure 3. The 3D attention U-Net architecture. Blue and pink boxes are feature maps, while gray boxes are concatenation maps. F_l denotes the number of feature maps in layer l; H denotes height; W denotes width; D denotes the number of dimensions; N_c denotes the number of classes. The input image is downsampled by a factor of two at each level in the downsampling path and upsampled by a factor of two in the upsampling path. Attention gates scale the concatenated features from the skip connections with the calculated attention coefficients from both the input features and the gating signals. The structure of the AG is shown in Figure 4. The deep supervision part is in red.

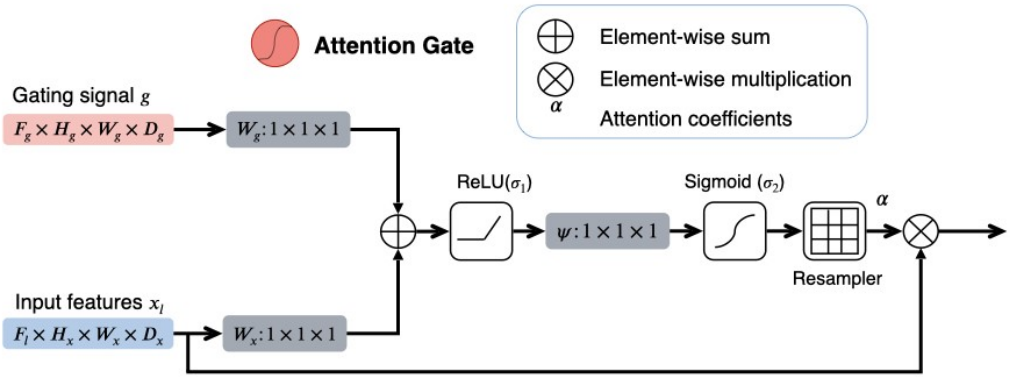


Figure 4. The structure of the AG. The input features x_l are scaled by the computed attention coefficients α in AG. To compute α, the gating signal g and the input features x_l are linearly transformed by W_g and W_x and summed. By analyzing the sum progressively by the ReLU activation, the linear transformation ψ, the sigmoid activation, and a grid resampler of trilinear interpolation, the attention coefficients α are obtained.

Attention gate: The schematic of the AGs is shown in Figure 4. The AGs collect the input features x_l in layer l from the skip connections and the gating signals g from the upsampling path. The gating signal is a grid signal adapted to spatial information. The attention coefficients $\alpha \in [0,1]$ are calculated as follows:

$$\alpha_l = \sigma_2(\psi^T(\sigma_1(W_x^T x_l + W_g^T g + b_g)) + b_\psi) \tag{1}$$

where $\sigma_2(x) = \frac{1}{1+exp(-x)}$ corresponds to the sigmoid activation function, and $\sigma_1(x) = max(0,x)$ corresponds to the ReLU activation function. W_x, W_g, and ψ are linear transformations and

b_g, b_ψ are the corresponding bias terms. The output attention maps are calculated by the element-wise multiplication of input x_l and attention coefficients:

$$\hat{x}_l = x_l \cdot \alpha_l \tag{2}$$

Briefly, AGs highlight salient features in the concatenation maps.

Deep supervision: In the attention U-Net [30], the authors used deep supervision [32] to guarantee that the intermediate feature maps were able to contribute to the final segmentation. As shown in Figure 3, the red parts indicate how to calculate the deep supervision loss. In the standard U-Net architecture, only the feature map at the final layer is used to calculate the loss. With deep supervision, the outputs from hidden layers contribute to the loss too.

2.2.3. DAU-Net

One of the most recent methods, by He et al. [12], applied a network similar to the standard 3D attention U-Net (in Figure 3) for automatic EAT segmentation and quantification of EAT volume from CCTA. They replaced the original ReLU activation function after convolution, max-pooling, and upsampling with the parametric rectified linear unit (PReLU).

2.2.4. U-Net++

Figure 5 illustrates a U-Net++ architecture with a backbone network of 5 layers. Compared to standard U-Net (the blue blocks and black connections), U-Net++ [31] is an architecture that has additional convolution blocks (the green blocks), dense skip connections (the green connections), and deep supervision (the red part) [32].

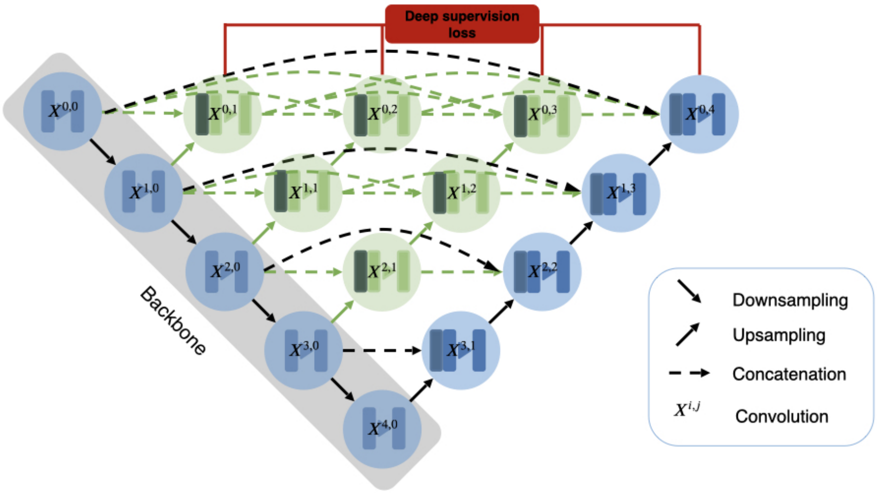


Figure 5. The U-Net++ architecture with a backbone network of 5 layers. Every circle block indicates a convolution operation with an activation function and BN. The blue blocks and the black connections indicate the original U-Net architecture. The additional convolution blocks and dense connections are in green. The red parts indicate deep supervision.

Dense skip connection: The number of additional convolution blocks depends on the pyramid level. In the standard U-Net, the feature maps from the downsampling path are concatenated to the feature maps in the upsampling path directly through the skip connections. Conversely, in the U-Net++ structure, due to the additional convolution blocks, more feature maps from convolution blocks at the same level are concatenated to the corresponding upsampling layers. In Figure 5, the feature map produced by the corresponding convolution block is represented by $x^{i,j}$, where i indicates the level of the block and j indicates the index of the convolution block. The feature maps represented by $x^{i,j}$ are calculated as follows:

$$x^{i,j} = \begin{cases} C(x^{i-1,j}), & j = 0 \\ C([[x^{i,k}]_{k=0}^{j-1}, U(x^{i+1,j-1})]), & j > 0 \end{cases} \tag{3}$$

where $C(\cdot)$ denotes a convolution operation followed by the ReLU function and batch normalization, $U(\cdot)$ denotes the upsampling operation, and [ ] denotes the concatenation operation. For each block, it fuses all the feature maps from the prior blocks at the same level and an upsampled feature map from the lower level. For example, $x^{1,2} = C([x^{1,0}, x^{1,1}, U(x^{2,1})])$. In this U-Net architecture, with a backbone network of 5 layers, the feature maps from convolution block $x^{0,j}, j \in 1,2,3,4$ are responsible for the deep supervision.

3. Experiments and Results

3.1. Experiment Set-Up

Data: For the experiments, the dataset was divided into five groups of 31, 31, 31, 31, and 30 CT scans. Groups 1-4 were used for the four-fold cross-validation, and the last group of 30 CTs was reserved for the hold-out evaluation. The initial size in 2D was 512 × 512 pixels. Due to hardware memory limitations, all images were downsampled to 256 × 256 pixels in the axial view.

Implementation details: To bring our experiments into correspondence with the original papers, in the experiments with 3D U-Net, 3D attention U-Net, and DAU-Net, the models were trained using the Dice loss [29]:

$$L_{DSC}(Y, \hat{Y}) = 1 - \frac{2 \cdot P(Y \bigcap \hat{Y})}{P(Y) + P(\hat{Y})} \tag{4}$$

where Y denotes the ground truth, $\hat{Y}$ denotes the predictions, and $P(\cdot)$ denotes the number of voxels. In the experiments with U-Net++, a combination of the binary cross-entropy and the Dice loss, which is the same as the loss function in the U-Net++ paper, was used as the loss function:

$$L = -(\frac{1}{2} \cdot \sum_i Y_i \cdot log\hat{Y}_i + \frac{2 \cdot P(Y \bigcap \hat{Y})}{P(Y) + P(\hat{Y})}) \tag{5}$$

where Y_i and $\hat{Y}_i$ denote the ith pixel of the ground truth or prediction. The backbone model for the U-Net++ was the VGG16 model with 2D convolutions [42]. The models of 3D U-Net and 3D attention U-Net were trained with the publicly available PyTorch implementation in the attention U-Net work https://github.com/ozan-oktay/Attention-Gated-Networks (accessed on 2 March 2021) [30]. The models of U-Net++ were trained with the Keras implementation in the original paper https://github.com/MrGiovanni/UNetPlusPlus/tree/master/keras (accessed on 3 March 2021). As the implementation of the DAU-Net was not publicly available, we implemented it according to the attention U-Net implementation [30]. The code for our experiments is shared on Github at https://github.com/Nirvanall/U-net_family_EAT_segmentationin_LDCT (accessed on 28 July 2023). For all experiments, we used a patch size of 256 × 256 × 96, except for U-Net++. As the U-Net++ was implemented with 2D convolution, we used a patch size of 256 × 256 and a batch size of 64. The initial learning rate was set as 0.0001, and all models were trained with the data augmentation setting, which was the same as the standard U-Net. More details can be found in Table 1. All models were trained with a maximum training epoch number of 500, with early stopping. All experiments were carried out using an NVIDIA Quadio RTX 6000 with 24 GB of memory.

Table 1. Implementation details.

Model	3D U-Net	3D Attention U-Net	DAU-Net	U-Net++
convolution dimension	3	3	3	2
deep supervision	False	True	True	True
decoder block type	Upsample	Upsample	Upsample	Transpose
optimizer	SGD	SGD	SGD	Adam

Evaluation metrics: For evaluation, we selected four metrics—the Dice similarity coefficient (DSC), the mIoU, the sensitivity, and the specificity. DSC computes the overlap of the region between the ground truth and the predicted segmentation:

$$DSC = \frac{1}{N}\frac{2 \cdot P(Y \bigcap \hat{Y})}{P(Y) + P(\hat{Y})}$$

where N indicates the number of CTs, Y denotes the ground truth, $\hat{Y}$ denotes the predictions, and $P(\cdot)$ denotes the number of pixels or voxels. A higher DSC indicates better prediction results. mIoU computes the similarity between the predicted segmentation and the ground truth:

$$mIoU = \frac{1}{N}\frac{P(Y \bigcap \hat{Y})}{P(Y \bigcup \hat{Y})}$$

mIoU is a more strict metric compared to DSC, and a higher mIoU generally indicates better segmentation. Sensitivity measures the rate of corrected predictions divided by the ground truth of the positive regions:

$$sensitivity = \frac{1}{N}\frac{P(Y \bigcap \hat{Y})}{P(Y)}$$

Higher sensitivity indicates the prediction is better at predicting positive regions, which are the target regions. Specificity measures the rate of corrected predictions over the ground truth of the negative regions:

$$specificity = \frac{1}{N}\frac{P(\bar{Y} \bigcap \bar{\hat{Y}})}{P(\bar{Y})}$$

where $\bar{Y}$ and $\bar{\hat{Y}}$ indicate the ground truth and prediction of the negative regions, which are the background. Higher specificity indicates a better ability to predict background.

3.2. Results

Herein, we demonstrate both cross-validation and hold-out evaluation for the selected methods. We trained all four networks using two types of labels separately, which included the region inside the pericardium (Figure 1b) and EAT labels (Figure 1c). The corresponding results are summarized in Tables 2 and 3.

3.2.1. Four-Fold Cross-Validation

For cross-validation, all models were trained with 93 CTs. From the cross-validation results shown in Table 2, the models trained with pericardium masks showed better performance compared to models trained with EAT labels. The U-Net++ models showed better results on almost all metrics compared to the 3D models.

Table 2. Cross-validation results of the selected models. The highest Pearson correlation is marked in bold.

Model	Label Type	DSC (%)	mIoU (%)	Sensitivity (%)	Specificity (%)	Correlation
3D U-Net	Pericardium EAT	74.90 ± 1.30 59.04 ± 0.92	60.92 ± 1.47 42.45 ± 0.74	77.67 ± 2.15 70.99 ± 2.42	99.53 ± 0.00 98.96 ± 0.00	0.7588 0.5648
3D attention U-Net	Pericardium EAT	68.52 ± 6.38 54.94 ± 6.43	56.62 ± 5.97 41.56 ± 4.78	70.50 ± 10.22 57.34 ± 9.92	99.69 ± 0.00 99.47 ± 0.00	0.2085 0.3883
DAU-Net	Pericardium EAT	80.06 ± 0.50 71.91 ± 0.49	67.30 ± 0.91 56.58 ± 0.69	91.80 ± 0.46 83.01 ± 0.28	99.34 ± 0.00 99.17 ± 0.00	0.8448 0.8596
U-Net++	Pericardium EAT	86.16 ± 0.23 77.42 ± 0.71	75.97 ± 0.51 63.78 ± 0.90	88.31 ± 0.55 84.61 ± 1.61	99.72 ± 0.00 99.47 ± 0.00	**0.9123** 0.7303

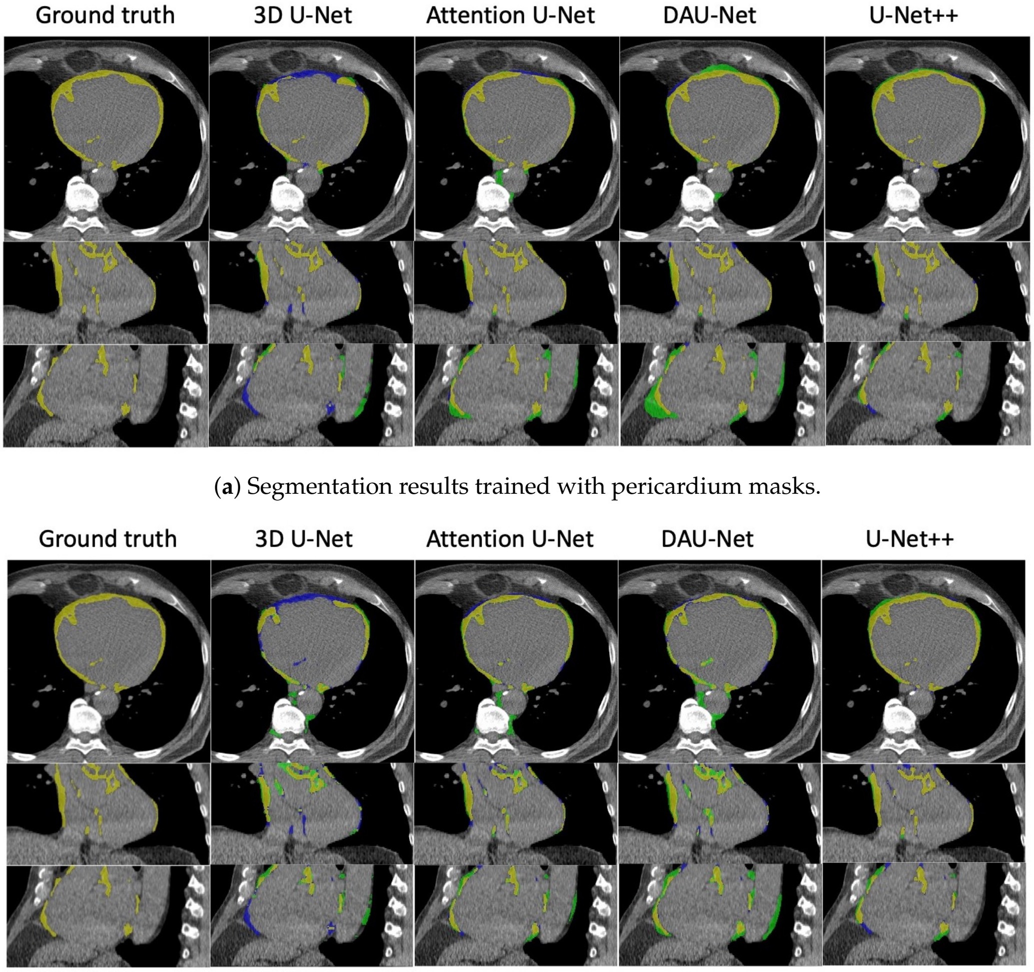
3.2.2. Hold-Out Test

Similarly, for the hold-out test, all models were trained with both types of labels separately. The hold-out test was performed using 30 reserved test samples, and the models were trained with 124 samples. The numerical results are shown in Table 3. The results of the hold-out test show that: (1) most models showed better performance compared to the cross-validation as the result of more training data, (2) generally, the models trained with pericardium masks performed better than the corresponding models trained with EAT labels, and (3) U-Net++ models showed the best performance compared to the other models, both for hold-out test and cross-validation. We noticed that the performance of 3D attention U-Net was unexpectedly similar to or even lower than the performance of 3D U-Net. By checking the evaluation results on every test sample, we found that the variance in 3D attention U-Net segmentation results was much higher. Most of the segmentation results were better, but some cases had extremely low performance. The employment of attention mechanisms might account for the observed lower performance and increased variance. A detailed discussion on this matter is presented in Section 4. Especially for the attention U-Net model trained with EAT labels, though most of the predictions had a Dice higher than 65%, 6 out of 30 predicted segmentation results showed very low performance (some even lower than 1% for DSC).

Table 3. Hold-out test results of the selected models. The highest Pearson correlation is marked in bold.

Model	Label Type	DSC (%)	mIoU (%)	Sensitivity (%)	Specificity (%)	Correlation
3D U-Net	Pericardium EAT	74.42 ± 1.37 63.94 ± 1.20	60.37 ± 1.57 47.84 ± 1.20	72.95 ± 2.67 69.83 ± 2.17	99.64 ± 0.00 99.22 ± 0.00	0.6661 0.6293
3D attention U-Net	Pericardium EAT	74.99 ± 4.79 55.39 ± 5.26	63.48 ± 4.36 41.32 ± 3.96	80.29 ± 7.27 60.11 ± 9.39	99.57 ± 0.00 99.35 ± 0.00	0.5120 0.1386
DAU-Net	Pericardium EAT	82.33 ± 0.39 72.13 ± 0.40	70.43 ± 0.80 56.78 ± 0.61	90.11 ± 0.45 82.69 ± 0.20	99.49 ± 0.00 99.21 ± 0.00	0.8445 0.8047
U-Net++	Pericardium EAT	87.99 ± 0.12 80.18 ± 0.20	78.71 ± 0.30 67.13 ± 0.39	91.45 ± 0.17 81.47 ± 0.43	99.72 ± 0.00 99.64 ± 0.00	**0.9606** 0.9405

The comparisons between all models—3D U-Net, 3D attention U-Net, DAU-Net, and U-Net++—are demonstrated using 2D visualization in axial, coronal, and sagittal views in Figure 6a (models trained with pericardium masks) and Figure 6b (models trained with EAT labels). The corresponding errors are green for false positives and blue for false negatives. From the errors, we see that most of the mistakes were the false positive pixels (in green). In CT, most of the green pixels were the mediastinal fat located outside of the pericardium, but were similar to the EAT visually. Compared to the models trained with pericardium masks, models trained with EAT had more false negative pixels inside the pericardium.



(**a**) Segmentation results trained with pericardium masks.

(**b**) Segmentation results trained with EAT labels.

Figure 6. Visualization of segmentation results. The first column shows the ground truth, and the other columns show predictions from 3D U-Net, attention U-Net, DAU-Net, and U-Net++. The first row shows the segmentation in the axial view, the second row is for the coronal view, and the third row is for the sagittal view. In the segmentation results, the true positive is highlighted in yellow, the false positive is highlighted in green, and the false negative is highlighted in blue.

3.3. Quantitative Analysis of EAT Volume

The Pearson correlation coefficients for every model are shown in Tables 2 and 3. Based on the results of the hold-out test, we show the Bland–Altman plots for all models trained with 124 samples, in Figure 7 for models trained with pericardium masks and Figure 8 for models trained with EAT labels. Quantitatively, U-Net++ showed the best performance in terms of Pearson correlation coefficient (0.9606) and outperformed the other models in the Bland–Altman analysis. Compared to the correlation coefficient ($r = 0.98$) of the EAT volume from two observers reported in [17], all the trained models were below this value.

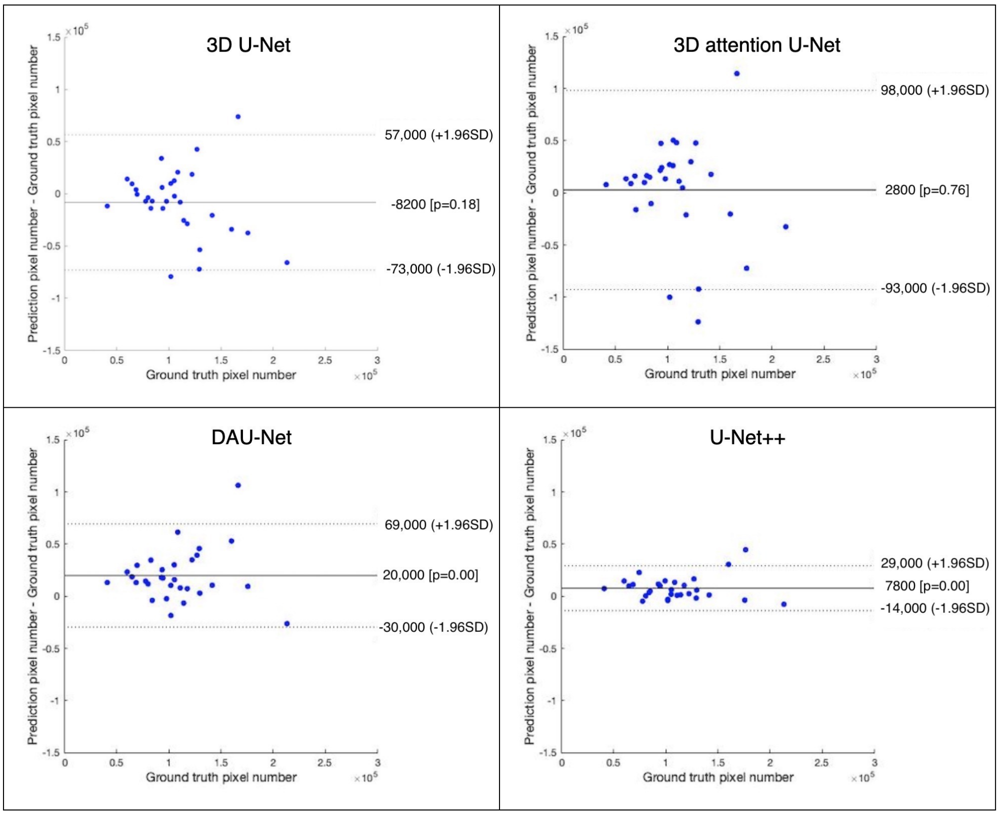


Figure 7. Bland−Altman plots for EAT volume for models trained with the pericardium masks. Comparison between the predicted pixel number of EAT and the ground truth pixel number.

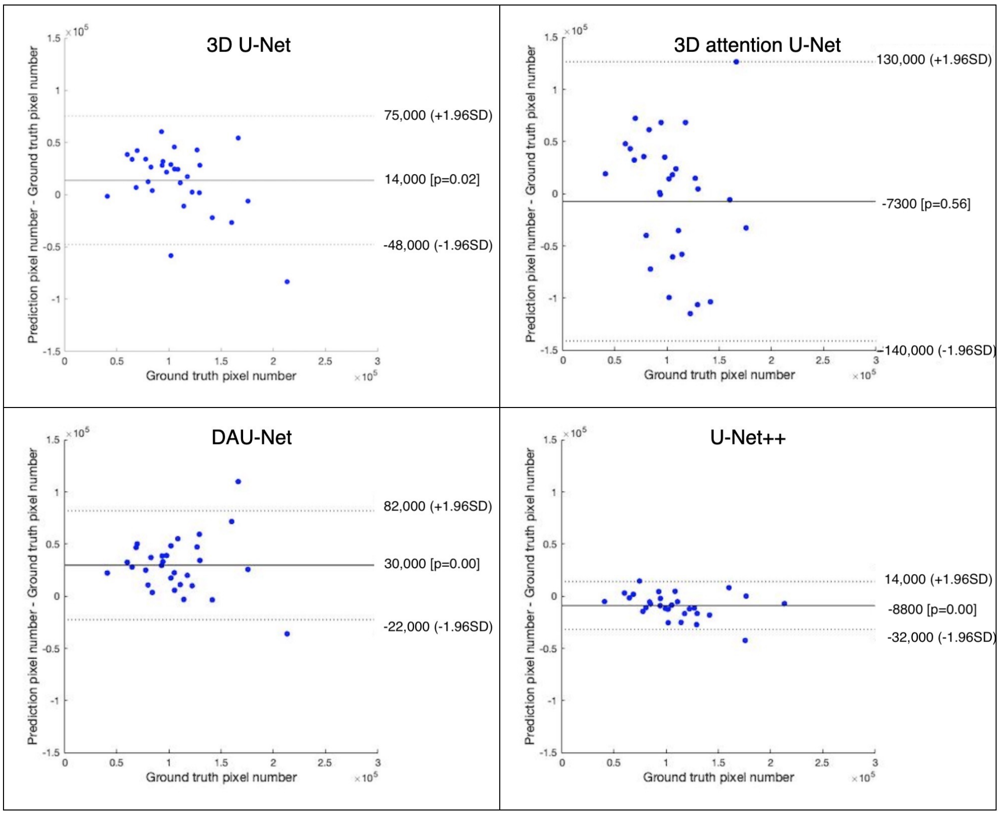


Figure 8. Bland−Altman plots for EAT volume for models trained with the EAT labels. Comparison between the predicted pixel number of EAT and the ground truth pixel number.

4. Discussion

The performance comparison of the state-of-the-art methods for EAT segmentation showed some interesting information: (1) Generally, the neural networks trained with the pericardium masks showed better segmentation and quantification results; and (2) U-Net++ trained with the pericardium masks outperformed the other models on segmentation and quantification. Furthermore, the process, encompassing data collection to evaluation results, involves numerous intricate details. Here, we discuss some points that we believe are crucial to EAT segmentation and quantification.

Label types: To the best of our knowledge, there are two types of labels in the works for EAT segmentation: (1) the pixel-wise or voxel-wise label maps (e.g., in [7,33]), and (2) the contours or outlines of the EAT (e.g., in [12,20]). There are various ways to obtain these labels, but, as most works are based on unpublished data, we cannot know the labeling protocols in detail. Based on our observations of example figures with labels and our extensive experience in labeling, it is evident that there is no ideal label type for EAT segmentation. Regarding pixel-wise or voxel-wise label maps, it is important to consider the physical characteristics of CT scans. Specific noise points can be erroneously classified as EAT within the pericardium. Though we could remove most of the noise with some denoising techniques, it is hard to remove all of them. The contours and outlines may better focus on the region with fat tissue. Because, geometrically, the contours occupy pixels or voxels, some small regions with very thin layers of fat may be neglected during labeling. In addition, the delineation of complex structures could be extremely time-consuming in 3D data such as CTs.

From the evaluation perspective, different label types could lead to different errors. For pixel-wise and voxel-wise labels, as there is noise, these may lead to more false negatives in the prediction. For contours and outlines labels, as some fat tissue may be missing, these may lead to more false positives in the prediction.

Attention mechanism: Compared to the results obtained from other methods, the performance of 3D attention U-Net demonstrates relatively lower accuracy and exhibits larger variance. While attention mechanisms can be advantageous in focusing on informative regions and enhancing performance in certain scenarios, they may also introduce additional complexity and elevate the risk of overfitting. Moreover, the effectiveness of attention mechanisms heavily relies on the dataset characteristics and the specific target structures being segmented. In the case of EAT segmentation in LDCT, where the target object has low visibility, the attention mechanism may lead to more errors. In Figure 6, it is evident that most mistakes made by the attention-based model are false positives below the pericardium, which aligns with the results obtained from the DAU-Net. Hence, the attention mechanisms might pose challenges in excluding structures below the pericardium accurately.

Domain knowledge: As the goal of EAT segmentation network is to find the exact locations of EAT voxel-wisely, ideally, it should mimic the manual segmentation progress as performed by radiologists. Thus, anatomical knowledge about EAT is a crucial reference for segmentation. By using the labels of pericardium masks, the network is forced to learn knowledge about pericardium, which is the foundation to distinguish EAT and adjacent similar structures such as mediastinal fat. Apart from our dataset, some works on EAT segmentation used labels that included pericardium knowledge too [7,20]. However, most works did not incorporate this knowledge deeply in deep neural networks. From a recent review of anatomy-aided deep learning for medical image segmentation [43], we noticed that there are many possible ways to incorporate anatomy information into deep learning. A well-integrated network could take advantage of both anatomy knowledge and data-driven deep learning methods.

Patch size: For models using 3D convolution operations, the patch size is a key hyperparameter for training. Due to the large image size of medical images and the memory limitation of GPUs, usually, it is not feasible to process a 3D image as one input for the input layer. Thus, the common way to solve this is to set a patch size for the input layer. By setting the patch size, the large 3D images are divided into multiple patches and processed separately for training. As the patch size influences the size of feature maps and the contextual information in the training process directly, it influences the segmentation results significantly. From our experiments and related papers, we noticed that, generally, a larger patch size could lead to better results. However, a larger patch size may lead to a much higher computational cost. Thus, there is a trade-off to choose the proper patch size for training performance and efficiency. In this paper, we selected 256 × 256 × 96, which is a relatively large patch size.

Training time and inference time: For all experiments, we set the maximum training epoch number as 500, with early stopping. With the exception of U-Net++, which stopped earlier, the other models were trained until the 500th epoch. The training time for U-Net++ models (usually stopped at around the 80th epoch) was between 12 and 17 h, while the training time for 3D U-Net, attention U-Net, and DAU-Net was between 5 and 8 days, due to the different amounts of training data and different label types. The inference times for all models ranged from 4.33 s to 6 s per sample and were relatively similar. Thus, to verify whether training for a longer time would increase the performance, we trained models of 3D U-Net in the hold-out test set-up for 1000 epochs (in Table 4) with both label types. All models showed improving performance. Particularly, the model trained with EAT showed obvious improvement. However, the computation cost was very high, as the training time doubled.

Table 4. Hold-out test results of 3D U-Net at the 1000 th epoch.

Label Type	DSC (%)	mIoU (%)	Sensitivity (%)	Specificity (%)
Pericardium	76.04 ± 2.08 (+1.62)	63.03 ± 2.27 (+2.66)	73.10 ± 3.69 (+0.15)	99.74 ± 0.00 (+0.10)
EAT	70.97 ± 0.71 (+7.03)	55.58 ± 0.83 (+7.74)	76.45 ± 1.53 (+6.62)	99.38 ± 0.00 (+0.16)

Deep supervision: During experiments, we noticed that deep supervision influenced the decrease in training loss and models' performance sharply. To verify the effects of deep supervision, we trained 3D attention U-Net models with and without deep supervision (in Table 5).

Table 5. Hold-out test results of 3D attention U-Net without deep supervision.

Label Type	DSC (%)	mIoU (%)	Sensitivity (%)	Specificity (%)
Pericardium	65.60 ± 9.26 (−9.39)	54.90 ± 7.94 (−8.58)	61.25 ± 10.75 (−19.04)	99.86 ± 0.00 (+0.29)
EAT	45.85 ± 9.89 (−9.54)	34.93 ± 7.01 (−6.39)	39.26 ± 9.59 (−20.85)	99.83 ± 0.00 (+0.48)

To visualize the prediction of 3D attention U-Net, we show the region inside pericardium segmentation results of 3D attention U-Net with and without deep supervision in Figure 9. Without deep supervision, the segmentation focused too much on the inside region and misses some pixels around the pericardium. This reduced the performance of EAT segmentation sharply, as most EAT locates near the pericardium.

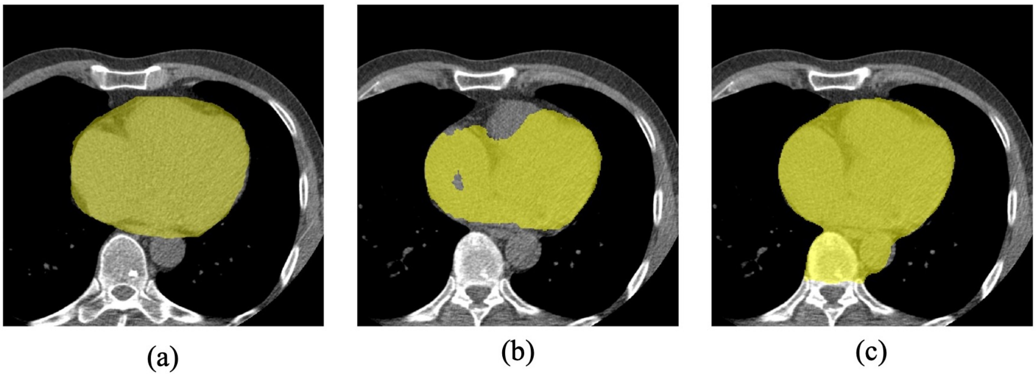


Figure 9. Segmentation results of 3D attention U-Net. (**a**) ground truth. (**b**) segmentation results of 3D attention U-Net without deep supervision. (**c**) segmentation results of 3D attention U-Net with deep supervision.

Evaluation: As some models are trained with 3D inputs and some are trained with 2D inputs, the evaluation could sometimes be tricky. From the previous works on EAT segmentation, we noticed that most of the early works were based on 2D CT slices [7,13,19–21], while some recent works were based on 3D CT images [9,12]. At the evaluation stage, when computing the mean values for evaluation metrics, early works treated one 2D slice as a sample [7], while some recent works treated a 3D image as a sample [12]. Therefore, there is a numerical difference due to the different computation settings. We tested both computation settings of mean values for our trained U-Net++ models. The mean values computed based on 2D samples were slightly higher than those computed based on 3D samples. However, considering that one 3D sample is from one patient, all the evaluation metrics in our paper were calculated based on 3D samples.

Future Work

From the related literature and our experiments, we believe that there is potential for automatic EAT segmentation. Here, we list some future work directions that we hope could be helpful for further research.

Domain knowledge: From the previous works, we noticed the lack of domain knowledge in the research into EAT segmentation and quantification. Many models for EAT segmentation only applied an existing segmentation network to the EAT data directly. Thus, the domain knowledge from the radiology aspect and the anatomical uniqueness of EAT is ignored. We believe that there are more possibilities if we could incorporate deep learning techniques and domain knowledge deeply.

Data unification: Due to the large variety of data acquisition protocols, the comparison between research papers on this topic is hard. Thus, a unified labeling and data acquisition protocol could reduce this barrier and increase the reproducibility of future works.

Benchmark: Compared to some popular medical image segmentation tasks such as brain tumors or lung nodules, there are very limited publicly available data or benchmarks with EAT labels. Thus, the comparison between methods is hard and the reproducibility is very low. With a benchmark, this problem could be solved, and the visibility and popularity of EAT segmentation could be improved.

Deep learning techniques: Recently, the deep learning techniques for segmentation have improved rapidly. Many techniques such as generative adversarial networks, physics-informed deep learning, and graph neural networks have shown success in many other segmentation tasks [44], but have not yet been applied to EAT segmentation.

5. Conclusions

In this work, we compared four state-of-the-art models (3D U-Net, 3D attention U-Net, DAU-Net, and U-Net++) from the U-Net family with regard to their performance for EAT segmentation and quantification. All models were evaluated on a dataset of 154 LDCT from the ROBINSCA trial with two different types of labels. The U-net++ model trained with pericardium masks showed better EAT segmentation and quantification results as well as higher efficiency compared to the other models. Based on the experimental results and existing literature, we examined crucial aspects of EAT segmentation and identified potential areas for future research. While state-of-the-art methods from the U-net family have displayed promising results, they also exhibit certain limitations when it comes to EAT segmentation and quantification. We are optimistic about the potential of deep learning techniques in alleviating the workload of radiologists and potentially replacing labor-intensive manual evaluations. However, there is still a considerable amount of work ahead of us before achieving the necessary standards for clinical application.

Author Contributions: Conceptualization, L.L.; methodology, L.L.; software, L.L.; validation, L.L.; formal analysis, L.L.; investigation, L.L.; resources, P.M.A.v.O., M.O.; data curation, R.M.; writing—original draft preparation, L.L.; writing—review and editing, R.M., R.N.J.V., C.B., R.V.; visualization, L.L.; supervision, R.N.J.V., C.B., R.V.; project administration, R.N.J.V.; funding acquisition, C.B., R.N.J.V., R.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by ZonMw under project B3CARE (project number 104006003). This project has received funding from the EU Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 777826. CB acknowledges support from the Dutch 4TU HTSF program Precision Medicine. Financial support provided by the China Scholarship Council (CSC) for the Ph.D. project of RM is gratefully acknowledged.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used during and/or analyzed in the current study are not publicly available due to patient data privacy concerns. The code will be available upon acceptance.

Conflicts of Interest: The authors declare that they have no competing interests.

Abbreviations

The following abbreviations are used in this manuscript:

EAT	Epicardial adipose tissue
CT	Computed tomography
NCCT	Non-contrast CT
CCTA	CT angiography
MRI	Magnetic resonance images
DSC	Dice similarity coefficient
mIoU	Mean intersection of union
GPU	Graphical processing unit
CNN	Convolutional neural network
ReLU	Rectified linear unit
AG	Attention gate
BN	Batch normalization

References

1. Sacher, F.; Roberts-Thomson, K.; Maury, P.; Tedrow, U.; Nault, I.; Steven, D.; Hocini, M.; Koplan, B.; Leroux, L.; Derval, N. Epicardial ventricular tachycardia ablation: A multicenter safety study. *J. Am. Coll. Cardiol.* **2010**, *55*, 2366–2372. [CrossRef] [PubMed]
2. Ouwens, D.; Sell, H.; Greulich, S.; Eckel, J. The role of epicardial and perivascular adipose tissue in the pathophysiology of cardiovascular disease. *J. Cell. Mol. Med.* **2010**, *14*, 2223–2234. [CrossRef]
3. Mancio, J.; Azevedo, D.; Saraiva, F.; Azevedo, A.; Pires-Morais, G.; Leite-Moreira, A.; Falcao-Pires, I.; Lunet, N.; Bettencourt, N. Epicardial adipose tissue volume assessed by computed tomography and coronary artery disease: A systematic review and meta-analysis. *Eur. Heart J. Cardiovasc. Imaging* **2018**, *19*, 490–497. [CrossRef]
4. Willens, H.; Byers, P.; Chirinos, J.; Labrador, E.; Hare, J.; Marchena, E. Effects of weight loss after bariatric surgery on epicardial fat measured using echocardiography. *Am. J. Cardiol.* **2007**, *99*, 1242–1245. [CrossRef] [PubMed]

5. Natale, F.; Tedesco, M.; Mocerino, R.; Simone, V.; Di Marco, G.; Aronne, L.; Credendino, M.; Siniscalchi, C.; Calabro, P.; Cotrufo, M. Others Visceral adiposity and arterial stiffness: Echocardiographic epicardial fat thickness reflects, better than waist circumference, carotid arterial stiffness in a large population of hypertensives. *Eur. J. Echocardiogr.* **2009**, *10*, 549–555. [CrossRef] [PubMed]
6. Nagayama, Y.; Nakamura, N.; Itatani, R.; Oda, S.; Kusunoki, S.; Takahashi, H.; Nakaura, T.; Utsunomiya, D.; Yamashita, Y. Epicardial fat volume measured on nongated chest CT is a predictor of coronary artery disease. *Eur. Radiol.* **2019**, *29*, 3638–3646. [CrossRef]
7. Rodrigues, É; Morais, F.; Morais, N.; Conci, L.; Neto, L.; Conci, A. A novel approach for the automated segmentation and volume quantification of cardiac fats on computed tomography. *Comput. Methods Programs Biomed.* **2016**, *123*, 109–128. [CrossRef]
8. Kazemi, A.; Keshtkar, A.; Rashidi, S.; Aslanabadi, N.; Khodadad, B.; Esmaeili, M. Segmentation of cardiac fats based on Gabor filters and relationship of adipose volume with coronary artery disease using FP-Growth algorithm in CT scans. *Biomed. Phys. Eng. Express* **2020**, *6*, 055009. [CrossRef]
9. Zhang, Q.; Zhou, J.; Zhang, B.; Jia, W.; Wu, E. Automatic Epicardial Fat Segmentation and Quantification of CT Scans Using Dual U-Nets With a Morphological Processing Layer. *IEEE Access* **2020**, *8*, 128032–128041. [CrossRef]
10. Hoori, A.; Hu, T.; Lee, J.; Al-Kindi, S.; Rajagopalan, S.; Wilson, D. Deep learning segmentation and quantification method for assessing epicardial adipose tissue in CT calcium score scans. *Sci. Rep.* **2022**, *12*, 2276. [CrossRef]
11. Qu, J.; Chang, Y.; Sun, L.; Li, Y.; Si, Q.; Yang, M.; Li, C.; Zhang, X. Deep Learning-Based Approach for the Automatic Quantification of Epicardial Adipose Tissue from Non-Contrast CT. *Cogn. Comput.* **2022**, *14*, 1392–1404. [CrossRef]
12. He, X.; Guo, B.; Lei, Y.; Wang, T.; Fu, Y.; Curran, W.; Zhang, L.; Liu, T.; Yang, X. Automatic segmentation and quantification of epicardial adipose tissue from coronary computed tomography angiography. *Phys. Med. Biol.* **2020**, *65*, 095012. [CrossRef] [PubMed]
13. Zlokolica, V.; Krstanović, L.; Velicki, L.; Popović, B.; Janev, M.; Obradović, R.; Ralević, N.; Jovanov, L.; Babin, D. Semiautomatic Epicardial fat segmentation based on fuzzy c-means clustering and geometric ellipse fitting. *J. Healthc. Eng.* **2017**, *2017*, 5817970. [CrossRef] [PubMed]
14. Li, X.; Sun, Y.; Xu, L.; Greenwald, S.; Zhang, L.; Zhang, R.; You, H.; Yang, B. Automatic quantification of epicardial adipose tissue volume. *Med. Phys.* **2021**, *48*, 4279–4290. [CrossRef]
15. West, H.; Siddique, M.; Williams, M.; Volpe, L.; Desai, R.; Lyasheva, M.; Thomas, S.; Dangas, K.; Kotanidis, C.; Tomlins, P. Deep-learning for epicardial adipose tissue assessment with computed tomography: Implications for cardiovascular risk prediction. *JACC Cardiovasc. Imaging* **2023**, *16*, 800–816. [CrossRef]
16. Bard, A.; Raisi-Estabragh, Z.; Ardissino, M.; Lee, A.; Pugliese, F.; Dey, D.; Sarkar, S.; Munroe, P.; Neubauer, S.; Harvey, N. Others Automated quality-controlled cardiovascular magnetic resonance pericardial fat quantification using a convolutional neural network in the UK Biobank. *Front. Cardiovasc. Med.* **2021**, *8*, 567. [CrossRef]
17. Nakazato, R.; Shmilovich, H.; Tamarappoo, B.; Cheng, V.; Slomka, P.; Berman, D.; Dey, D. Interscan reproducibility of computer-aided epicardial and thoracic fat measurement from noncontrast cardiac CT. *J. Cardiovasc. Comput. Tomogr.* **2011**, *5*, 172–179. [CrossRef]
18. Yalamanchili, R.; Dey, D.; Kukure, U.; Nakazato, R.; Berman, D.; Kakadiaris, I. Knowledge-based quantification of pericardial fat in non-contrast CT data. *Med. Imaging Image Process.* **2010**, *7623*, 76231X.
19. Ding, X.; Terzopoulos, D.; Diaz-Zamudio, M.; Berman, D.; Slomka, P.; Dey, D. Automated epicardial fat volume quantification from non-contrast CT. *Med. Imaging Image Process.* **2014**, *9034*, 90340I.
20. Ding, X.; Terzopoulos, D.; Diaz-Zamudio, M.; Berman, D.; Slomka, P.; Dey, D. Automated pericardium delineation and epicardial fat volume quantification from noncontrast CT. *Med. Phys.* **2015**, *42*, 5015–5026. [CrossRef]
21. Shahzad, R.; Bos, D.; Metz, C.; Rossi, A.; Kirişli, H.; Lugt, A.; Klein, S.; Witteman, J.; Feyter, P.; Niessen, W. Automatic quantification of epicardial fat volume on non-enhanced cardiac CT scans using a multi-atlas segmentation approach. *Med. Phys.* **2013**, *40*, 091910. [CrossRef]
22. Kazemi, A.; Keshtkar, A.; Rashidi, S.; Aslanabadi, N.; Khodadad, B.; Esmaeili, M. Segmentation of Cardiac Epicardial and Pericardial Fats by Using Gabor Filter Bank Based GLCM. In Proceedings of the 2019 26th National And 4th International Iranian Conference On Biomedical Engineering (ICBME), Tehran, Iran, 27–28 November 2019; pp. 177–182.
23. Norlén, A.; Alvén, J.; Molnar, D.; Enqvist, O.; Norrlund, R.; Br Berg, J.; Bergström, G.; Kahl, F. Automatic pericardium segmentation and quantification of epicardial fat from computed tomography angiography. *J. Med. Imaging* **2016**, *3*, 034003. [CrossRef] [PubMed]
24. Kazemi, A.; Keshtkar, A.; Rashidi, S.; Aslanabadi, N.; Khodadad, B.; Esmaeili, M. Automated Segmentation of Cardiac Fats Based on Extraction of Textural Features from Non-Contrast CT Images. In Proceedings of the 2020 25th International Computer Conference, Computer Society Of Iran (CSICC), Tehran, Iran, 25–26 January 2020; pp. 1–7.
25. Albuquerque, V.; Rodrigues, D.; Ivo, R.; Peixoto, S.; Han, T.; Wu, W.; Rebouças Filho, P. Fast fully automatic heart fat segmentation in computed tomography datasets. *Comput. Med. Imaging Graph.* **2020**, *48*, 101674. [CrossRef] [PubMed]
26. Chen, C.; Qin, C.; Qiu, H.; Tarroni, G.; Duan, J.; Bai, W.; Rueckert, D. Deep learning for cardiac image segmentation: A review. *Front. Cardiovasc. Med.* **2020**, *7*, 25. [CrossRef]
27. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.

28. Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In Proceedings of the International Conference On Medical Image Computing And Computer-assisted Intervention, Athens, Greece, 17–21 October 2016; pp. 424–432.
29. Milletari, F.; Navab, N.; Ahmadi, S. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference On 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.
30. Oktay, O.; Schlemper, J.; Folgoc, L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.; Kainz, B. Attention u-net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
31. Zhou, Z.; Siddiquee, M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In Proceedings of the Deep Learnin, Granada, Spain, 20 September 2018; pp. 3–11.
32. Lee, C.; Xie, S.; Gallagher, P.; Zhang, Z.; Tu, Z. Deeply-supervised nets. In Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, Cadiz, Spain, 9–11 May 2015; Volume 38, pp. 562–570. Available online: http://proceedings.mlr.press/v38/lee15a.html (accessed on 19 February 2019).
33. Commandeur, F.; Goeller, M.; Betancur, J.; Cadet, S.; Doris, M.; Chen, X.; Berman, D.; Slomka, P.; Tamarappoo, B.; Dey, D. Deep learning for quantification of epicardial and thoracic adipose tissue from non-contrast CT. *IEEE Trans. Med. Imaging* **2018**, *37*, 1835–1846. [CrossRef]
34. Commandeur, F.; Goeller, M.; Razipour, A.; Cadet, S.; Hell, M.; Kwiecinski, J.; Chen, X.; Chang, H.; Marwan, M.; Achenbach, S. Others Fully automated CT quantification of epicardial adipose tissue by deep learning: A multicenter study. *Radiol. Artif. Intell.* **2019**, *1*, e190045. [CrossRef]
35. Santini, G.; Latta, D.; Vatti, A.; Ripoli, A.; Chiappino, S.; Piagneri, V.; Chiappino, D.; Martini, N. Deep Learning for pericardial fat extraction and evaluation on a population study. *MedRxiv* **2020**. [CrossRef]
36. Guo, S.; Liu, X.; Zhang, H.; Lin, Q.; Xu, L.; Shi, C.; Gao, Z.; Guzzo, A.; Fortino, G. Causal knowledge fusion for 3D cross-modality cardiac image segmentation. *Inf. Fusion* **2023**, *99*, 101864. [CrossRef]
37. Guo, S.; Xu, L.; Feng, C.; Xiong, H.; Gao, Z.; Zhang, H. Multi-level semantic adaptation for few-shot segmentation on cardiac image sequences. *Med. Image Anal.* **2021**, *73*, 102170. [CrossRef] [PubMed]
38. Vonder, M.; Aalst, C.; Vliegenthart, R.; Ooijen, P.; Kuijpers, D.; Gratama, J.; Koning, H.; Oudkerk, M. Coronary artery calcium imaging in the ROBINSCA trial: Rationale, design, and technical background. *Acad. Radiol.* **2018**, *25*, 118–128. [CrossRef] [PubMed]
39. Mihl, C.; Loeffen, D.; Versteylen, M.; Takx, R.; Nelemans, P.; Nijssen, E.; Vega-Higuera, F.; Wildberger, J.; Das, M. Automated quantification of epicardial adipose tissue (EAT) in coronary CT angiography; comparison with manual assessment and correlation with coronary artery disease. *J. Cardiovasc. Comput. Tomogr.* **2014**, *8*, 215–221. [CrossRef] [PubMed]
40. Fedorov, A.; Beichel, R.; Kalpathy-Cramer, J.; Finet, J.; Fillion-Robin, J.; Pujol, S.; Bauer, C.; Jennings, D.; Fennessy, F.; Sonka, M. Others 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn. Reson. Imaging* **2012**, *30*, 1323–1341. [CrossRef] [PubMed]
41. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
42. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2015**, arXiv:1409.1556.
43. Liu, L.; Wolterink, J.; Brune, C.; Veldhuis, R. Anatomy-aided deep learning for medical image segmentation: A review. *Phys. Med. Biol.* **2021**, *66*, 11TR01. [CrossRef]
44. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3523–3542.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.