A Generalization Study of Automatic Pericardial Segmentation in Computed Tomography Images

Rúben Baeza[1,2](✉), Carolina Santos[3], Fábio Nunes[3,4], Jennifer Mancio[4], Ricardo Fontes-Carvalho[3,4], Miguel Coimbra[2,5], Francesco Renna[2,5], and João Pedrosa[1,2]

[1] Faculdade de Engenharia da Universidade do Porto, Porto, Portugal
up202103374@up.pt
[2] Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência, Porto, Portugal
[3] Faculdade de Medicina da Universidade do Porto, Porto, Portugal
[4] Centro Hospitalar de Vila Nova de Gaia e Espinho, Vila Nova de Gaia, Portugal
[5] Faculdade de Ciências da Universidade do Porto, Porto, Portugal

Abstract. The pericardium is a thin membrane sac that covers the heart. As such, the segmentation of the pericardium in computed tomography (CT) can have several clinical applications, namely as a preprocessing step for extraction of different clinical parameters. However, manual segmentation of the pericardium can be challenging, time-consuming and subject to observer variability, which has motivated the development of automatic pericardial segmentation methods.

In this study, a method to automatically segment the pericardium in CT using a U-Net framework is proposed. Two datasets were used in this study: the publicly available Cardiac Fat dataset and a private dataset acquired at the hospital centre of Vila Nova de Gaia e Espinho (CHVNGE).

The Cardiac Fat database was used for training with two different input sizes - 512 × 512 and 256 × 256. A superior performance was obtained with the 256 × 256 image size, with a mean Dice similarity score (DCS) of 0.871 ± 0.01 and 0.807 ± 0.06 on the Cardiac Fat test set and the CHVNGE dataset, respectively.

Results show that reasonable performance can be achieved with a small number of patients for training and an off-the-shelf framework, with only a small decrease in performance in an external dataset. Nevertheless, additional data will increase the robustness of this approach for difficult cases and future approaches must focus on the integration of 3D information for a more accurate segmentation of the lower pericardium.

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within projects DSAIPA/AI/0083/2020 and LA/P/0063/2020.

A. Cunha et al. (Eds.): MobiHealth 2022, LNICST 484, pp. 157–167, 2023.
https://doi.org/10.1007/978-3-031-32029-3_15

Keywords: Pericardium Segmentation · Deep Learning · U-Net · Computed Tomography

1 Introduction

Coronary artery disease (CAD) is the most common type of heart disease and the leading cause of death worldwide [10]. CAD happens when the coronary arteries become hardened and narrowed causing the reduction of blood flow to the heart muscle, leading to ischemia. This is due to the buildup of deposits of calcium, fatty lipids, and inflammatory cells, called plaque, on the inner walls of coronary arteries (atherosclerosis). As a result, the heart muscle is deprived from the blood or oxygen it needs, which can lead to chest pain (angina) and myocardial infarction.

Many risk factors increase the probability of CAD including cholesterol, diabetes, smoking, being overweight, or family history of heart disease. In these cases, it is advisable to test for CAD [6]. Nowadays, quantifying calcium deposits in the coronary arteries through calcium scans is the most common non-invasive technique for screening CAD. A coronary calcium scan uses computerized tomography (CT) to detect calcium deposits in the coronary arteries of the heart. A higher coronary calcium score suggests a higher chance of the presence of a plaque and a higher risk of a future myocardial infarction [2].

However, CAD especially in people below 50 years of age, can be present without calcium (non-calcified plaque) and may not be detected by this technique which leads some patients to be misdiagnosed. Early studies suggest that changes in epicardial adipose tissue (EAT) may play an important role in the pathogenesis of CAD, emerging as a relevant factor for cardiovascular risk stratification [14]. These structures can be visualized in non-contrast cardiac CT performed for assessment of coronary calcium [9]. EAT assessment typically implies a prior segmentation of the pericardium - a thin sac that covers the heart composed by a double layered membrane. However, manual segmentation of the pericardium is user-dependent, due to the difficulty of correctly identifying the area among multiple observers and remains a tedious task, not suitable for clinical practice [4], being essential to enhance the repeatability of the results and to reduce time.

In recent years, several research studies investigated the automatic segmentation of the pericardium and EAT on CT using machine and deep learning algorithms. One of the first studies assessing a technique for the segmentation of the fat surrounding the heart compartments was reported by Rodrigues et al. [11]. They proposed a feature extraction method comprising an intersubjective registration based on the creation of an atlas and then, classification. This method can obtain an ideal performance in the Dice similarity score (DSC) (97.9 %) for the segmentation of EAT. However, several issues must be considered in the case of atlas-based techniques as the performance depends heavily on the registration step [8]. Commandeur et al. [4] used a deep learning approach consisting of a multitask computational graph CNN network where the pericardium was first segmented and then its inner region. This study reported a high correlation between deep learning quantification and manual segmentations of EAT of two experienced tracers ($r = 0.973$ and $r = 0.979$; $P < 0.001$). Zhang et al. [15] proposed a method based on the U-Net framework, applying dual U-Nets with a morphological layer on cardiac CT scans. The first U-Net is used to detect the pericardium and the second to find and segment the epicardial fat from the pericardium. The proposed method achieves a mean DSC of 91.19%. Although many of these studies show promising results, the effectiveness of these tools has not yet been demonstrated in large populations. Furthermore, there are still challenges to be overcome namely in the integration of 3D and prior information (e.g. pericardial geometry).

The work presented in this report is a preliminary study towards the automatic segmentation of the pericardium in CT imaging. A state of the art architecture is trained for this purpose using publicly available data. Furthermore, a generalization analysis of the considered segmentation solution is performed using a private external dataset.

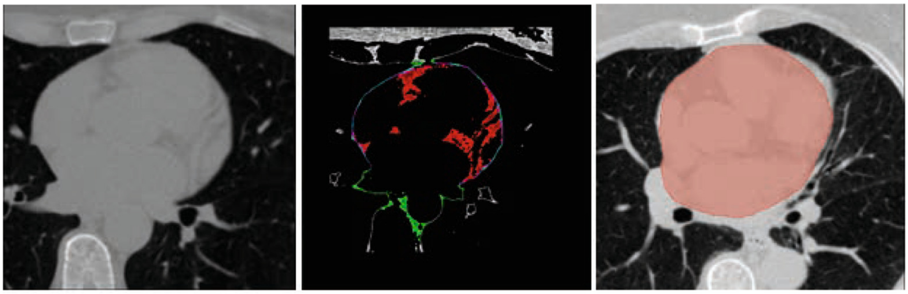
2 Methods

2.1 Datasets

In this work two datasets were employed: the publicly available Cardiac Fat dataset from Rio de Janeiro and a private dataset acquired from the hospital centre of Vila Nova de Gaia e Espinho (CHVNGE).

Cardiac Fat Dataset. The Cardiac Fat database [1] was acquired in Rio de Janeiro and released publicly by Rodrigues et al. [11]. The dataset includes 20 CT scans with 878 slices belonging to 20 patients as DICOM images (512×512 pixels). The CT scans were obtained with two different scanners (Phillips and Siemens) with a slice thickness of 3 mm. The average age of the patients is 55.4. The original ground truth was obtained via manual segmentation by a physician and a computer scientist who labeled the epicardial fat, pericardial fat (mediastinal), and finally, the pericardium layer. It is relevant to point out that CTs were manually centred and cropped to the pericardium and thresholded to the adipose tissue range of $[-200, -30]$ Hounsfield Unit (HU). Manual annotations of different structures are made available for these images. For this reason it is not possible to make a direct match of the labels with the original DICOM images. An example of a CT slices and the corresponding manual segmentations is shown in Fig. 1(a).

CHVNGE Dataset. The CHVNGE dataset is a subset of 20 patients randomly selected from the EPICHEART (The influence of EPICardial adipose tissue in HEART diseases) Study (ClinicalTrials.gov: NCT03280433), collected at the CHVNGE in Vila Nova de Gaia, Portugal. The EPICHEART study is a translational study designed to investigate the association between EAT and CAD, left atrial remodelling and postoperative atrial fibrillation, and frailty syndrome in patients with symptomatic severe aortic stenosis referred to aortic valve



(a) (b)

Fig. 1. Dataset overview. (a) Cardiac Fat dataset. Left: DICOM images; right: DICOM in fat HU range and manual segmentations of cardiac fat. Red indicates epicardial fat, green indicates mediastinal fat. (b) CHVNGE Dataset. DICOM images with pericardium in light red. (Color figure online)

replacement. This is a prospective cohort including 574 patients who underwent pre-operative assessment, intra-operative samples collections, and follow-up in hospital and at 6 months after surgery. All data was anonymized prior to analysis for the purposes of this study. The dataset includes 20 CT scans with 909 slices as DICOM images (512 × 512 pixels). All CT scans were acquired on a Siemens Somatom Sensation 64 with an ECG-triggered scanning protocol (tube voltage 120 kV, tube current 190 mA, gantry rotation 330 ms, collimation 24 × 1.2 mm, pitch 0.2) and a slice thickness of 3 mm. The pericardial segmentation was obtained via manual segmentation by one of the authors (CS). An example of a CT slice and manual pericardial segmentation is shown in Fig. 1(b).

2.2 Image Registration

Given the mismatch between the labels and original DICOMs due to the manual cropping and centering of CT slices for labelling, an image registration methodology was used to align the labels to the DICOM images. In this way, full DICOM images can be used to train a deep learning segmentation model without the need for centering and cropping on external datasets.

First, the DICOM images were converted to the same range as labeled images of [−200, −30] HU to facilitate the recognition of key points between the two images. Image registration was then performed between pairs of DICOM and labeled CTs using the ORB (Oriented FAST and Rotated BRIEF) algorithm [13]. The process consists first of finding the key points using an algorithm called FAST, which mainly uses an intensity threshold between the central pixel and a circular ring around the center. Then, ORB uses BRIEF (Binary Robust Independent Elementary Features) descriptors to describe each keypoint. BRIEF is a bit string description constructed from a set of binary intensity tests between n pairs (x, y) of pixels which are located inside the patch (a key point is in the center of the patch). For further details, the reader is referred to the original publications of the ORB algorithm [13]. Next, a brute-force matcher was used to match the key points of the two images, selecting the best results while removing the noisy ones. The default parameters and the hamming distance were used for the ORB and the matcher, respectively. Finally, using the RANSAC algorithm it is possible to highlight the inliers (correct matches) and find the transformation matrix [7].

It should be noted that the images are misaligned in exactly the same way within each patient, therefore only one manually chosen transformation matrix is applied per patient. With this, there will be 20 transformation matrices corresponding to the 20 patients in the database.

2.3 Pericardial Segmentation

In this section, the strategy adopted to obtain the pericardial reference segmentations is initially presented. Finally, the models developed will be described, as well as the data preprocessing.

Pericardium Labelling. While the Cardiac Fat dataset has pericardial contours available, label inconsistencies were found in some images. Therefore, to train and test the segmentation model, only the EAT label was used to obtain an approximation of the pericardial mask using a Convex Hull [3]. This algorithm was applied to all binary masks of the EAT and an example is shown in Fig. 2.

Fig. 2. Convex hull pericardium labelling example. Left: EAT mask; middle: generated pericardium label using convex hull; right: EAT (red) and Pericardial (yellow) labels registered and overlapped with DICOM images. (Color figure online)

Data Preprocessing. Before training the network, the CT slices were first clipped to [−1000, 1000] HU and then normalized to a range between 0 and 1. A total of 121 slices from the original dataset that did not include pericardium labels were also removed during training.

Model Architecture. Pericardial segmentation was then performed using a U-Net architecture [12], as show in Fig. 3. The U-Net is a popular framework for deep learning models, it often obtains excellent performance in image segmentation, especially in the area of medical image processing. This model was trained from scratch with a total number of 31,031,685 parameters. Two different input sizes were tested to evaluate the importance of contextual information and network receptive field: 512 × 512 (the original image size) and 256 × 256.

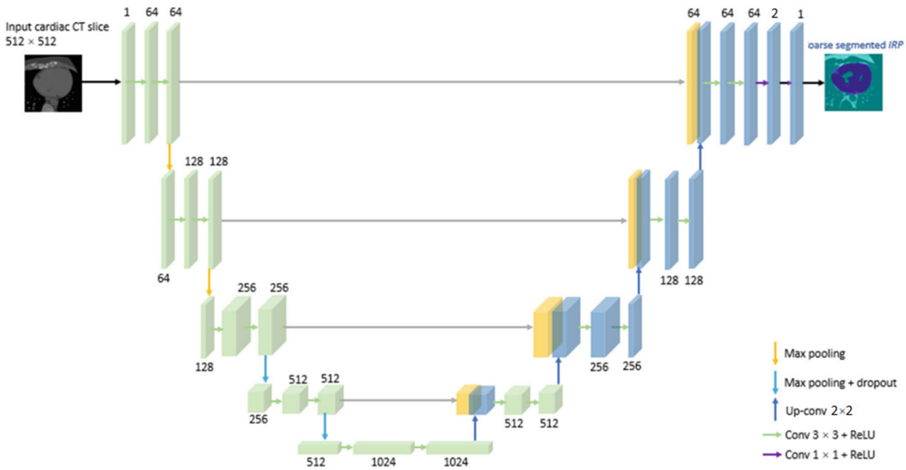


Fig. 3. U-Net architecture used for pericardial segmentation (adapted from [15]).

Training Setup. For the training, validation and testing of the model, the Cardiac Fat dataset was randomly divided as follows: 60% of the CT scans for training, 20% of the CT scans for validation, and the remaining 20% for testing. A patient-wise division was applied to avoid bias given that different CT slices of the same patient are highly correlated. The CHVNGE data was used for external testing of the automatic segmentation network and were thus not used for training.

The two models were trained with the binary cross-entropy loss function and using the adaptive moment estimation (Adam) optimizer. The binary cross-entropy is defined by the Eq. (1).

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(p(y_i)) + (1-y_i)\log(1-p(y_i)) \quad (1)$$

where y is the label and $p(y)$ is the predicted probability of the pixel being from the label for all N pixels.

A batch size of 2 and learning rate of 0.0001 was used for the first model. For the model trained with 256 × 256 images, a batch size of 12 and a learning rate of 0.001 was used. These hyper-parameters were selected based on empirical experiments, taking into account the memory constraints of the workstation used. An early stopping callback method was also used to stop training when the validation loss did not improve for 7 consecutive epochs. The two deep learning models were trained using Tensorflow and an NVIDIA GTX 1080 GPU.

Model Evaluation. Evaluation of the segmentation performance was done using the DSC [5], defined as:

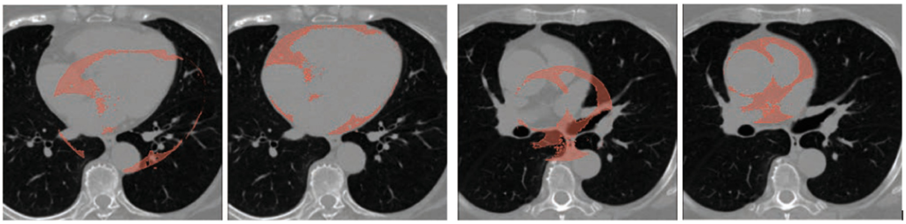
$$DCS(A,B) = \frac{2\,A \cap B|}{|A| + |B|}, \tag{2}$$

where A and B are the pericardial masks of the manual and automatic segmentations. The value of DCS ranges from 0, indicating no spatial overlap between the two segmentations, to 1, indicating complete overlap.

It should be noted that the model returns an array with the dimension of the image in which each pixel has a value from 0 to 1 that corresponds to the probability of it belonging to the pericardium. Thus, to create the predicted binary mask, this array was thresholded at 0.5.

3 Results and Discussion

3.1 Image Registration



(a) (b)

Fig. 4. Two examples ((a) and (b)) before (left) and after (right) image registration.

The result of the image registration for two patients is shown in Fig. 4. It can be seen that before the application of image registration for this patient, the labels were out of alignment with the DICOM image. After registration, a correct alignment of the mask with the DICOM image is verified as seen in Fig. 4. The registration results were evaluated qualitatively and a successful alignment was verified for all patients.

Table 1. DSC of the automatic segmentation for the Cardiac Fat dataset.

Size	Metric
	DCS (mean ± standard deviation)
256 × 256	0.871 ± 0.01
512 × 512	0.831 ± 0.02

3.2 Pericardial Segmentation

Cardiac Fat Dataset. The DCS mean values for the test images with 256 × 256 and 512 × 512 pixels are presented in Table 1. It can be seen that higher DSC was obtained for the 256 × 256 image size, with a DCS mean of 0.871 ± 0.01 for 256 × 256 images and 0.831 ± 0.02 for 512 × 512 images.

Two examples of automatic segmentations predicted by both models are presented in Fig. 5. It can be seen that the images predicted by the 256 × 256 model present slightly better results, being more similar to the corresponding label. In the case of the 512 × 512 images, there are some flaws in the classification of pixels and even inside the pericardium, some pixels are not attributed to it. The bottom row of Fig. 5 highlights the difficulty in segmenting the lower portion of the pericardium due to the presence of other organs in the CT slice.

The observed worse performance for the 512 × 512 images is likely due to the fact that the larger image size results in a decrease in the size of the receptive field thus giving less contextual information to the network, which can be important for segmentation. However, the use of different hyperparameters for the two models (batch size and learning rate) can also play a role. Even so, in both models a satisfactory DCS was obtained, taking into account that a standard U-Net architecture was used. Furthermore, the pericardium labels used for training, validation and testing on the Cardiac Fat dataset were obtained based on the convex hull of the EAT masks, which necessarily means that they will have missing sections, particularly on the lower and upper slices of the pericardium.

Table 2. DSC of the proposed automatic segmentation for the CHVNGE dataset.

Size	Metric
	DCS (mean ± standard deviation)
256 × 256	0.807 ± 0.06
512 × 512	0.769 ± 0.06

CHVNGE Dataset. The models trained on the Cardiac Fat dataset were then evaluated on the CHVNGE dataset to evaluate performance in an external dataset. The mean DCS for the 20 patients and the respective standard deviation are presented in the Table 2. It can be seen that, again, images with 256 × 256

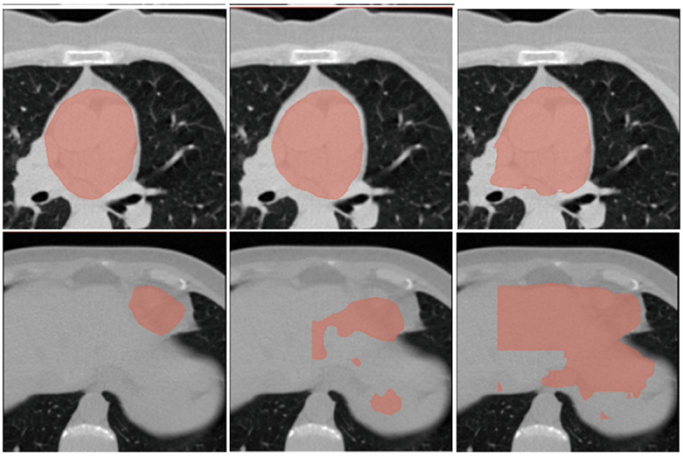


Fig. 5. Examples of pericardium segmentation for the Cardiac Fat dataset. Left: manual segmentation; middle: 256 × 256 automatic segmentation; right: 512 × 512 automatic segmentation).

showed better results with a DCS mean of 0.807 ± 0.06 for images with 512 × 512 pixels and 0.769 ± 0.06 for 512 × 512 images.

Two examples of automatic segmentations by both models are presented in Fig. 6. On the top row, it can again be seen that the results are better using 256 × 256 images and that the prediction is more similar to the corresponding label. However, examining the bottom row of the Fig. 6, it can be seen that the significant calcification of the aortic valve in this patient led to a failure of the segmentation network with a large portion of the center of the pericardium excluded from the segmentation. This example shows the need to train the model with a larger amount of data and especially with as much variability as possible with the aim of making the model more generalizable.

Still it is noteworthy that the results were only moderately worse compared to those obtained in the Cardiac Fat dataset (Table 1) with a difference of about 0.06 relative to the DCS metric. This decrease in performance was however expected since the model was trained for the Cardiac Fat dataset. In addition, the data from the hospital has more patients tested, increasing the variability and consequently the difficulty for the model to predict correctly.

3.3 Limitations and Future Work

In spite of the promising results obtained in this work, significant limitations remain as mentioned throughout the report.

First, the quantity and quality of the data and annotations that were used for training is clearly insufficient. While the 878 CT slices that compose the Cardiac Fat dataset are a significant number, it should be noted that these come from only 20 patients. As such, they are highly correlated and do not adequately

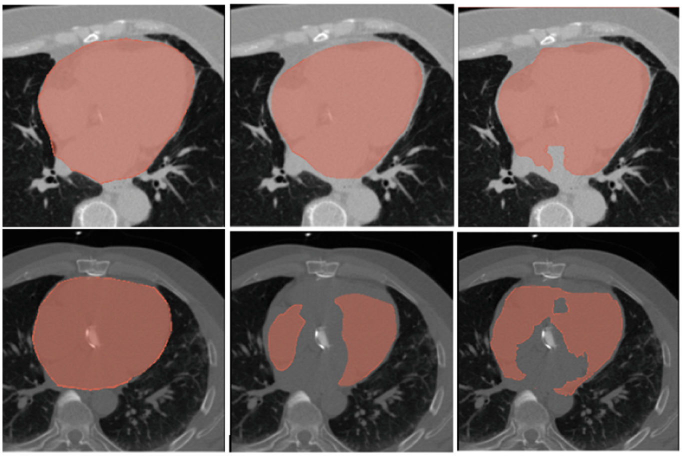


Fig. 6. Examples of pericardium segmentation for the CHVNGE dataset. Left: manual segmentation; middle: 256 × 256 automatic segmentation; right: 512 × 512 automatic segmentation).

represent the variability existing in the population. This was the observed on the CHVNGE dataset, with lower mean DSC and particularly for the example shown in Fig. 6. Furthermore, the labelling used for training is also not ideal as it was obtained from EAT labels and can be incomplete. As such it would be crucial to train the model with more patients and a more consistent pericardium labelling.

Second, in terms of segmentation performance, while the model can detect and locate the pericardium reasonably well throughout most axial slices, it can often fail, particularly for the lower pericardium and for slices where there is no pericardium (below or above). The use of 3D or 2.5D networks to provide further context during segmentation would thus be an extremely important strategy to improve performance in these slices which will be explored in future work.

4 Conclusions

In conclusion, a U-Net segmentation network for automatic pericardial segmentation was proposed, trained on the public Cardiac Fat dataset. A mean DCS of 0.831 ± 0.02 and 0.871 ± 0.01 was obtained on the public dataset using input sizes of 512 × 512 and 256 × 256, respectively. A private CHVNGE dataset was then tested and an expected worse performance with a DCS mean of 0.807 ± 0.06 for the 256 × 256 images and 0.769 ± 0.06 for 512 × 512 images. Overall, these results indicate that reasonable performance can be obtained with a small number of patients and only a small decrease in performance in an external dataset. Future work to improve performance and robustness must focus on increasing the quantity and variability of training and testing data and integrating contextual information to increase performance in the lower pericardium.

References

1. Cardiac fat database - computed tomography. http://visual.ic.uff.br/en/cardio/ctfat/index.php. Accessed 20 Feb 2022
2. Alexopoulos, N., McLean, D.S., Janik, M., Arepalli, C.D., Stillman, A.E., Raggi, P.: Epicardial adipose tissue and coronary artery plaque characteristics. Atherosclerosis **210**(1), 150–154 (2010)
3. Barber, C.B., Dobkin, D.P., Huhdanpaa, H.: The quickhull algorithm for convex hulls. ACM Trans. Mathem. Softw. (TOMS) **22**(4), 469–483 (1996)
4. Commandeur, F., et al.: Fully automated CT quantification of epicardial adipose tissue by deep learning: a multicenter study. Radiol. Artif. Intell. **1**(6), e190045 (2019)
5. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945)
6. Hajar, R.: Risk factors for coronary artery disease: historical perspectives. Heart Views Official J. Gulf Heart Assoc. **18**(3), 109 (2017)
7. Li, X., Liu, Y., Wang, Y., Yan, D.: Computing homography with RANSAC algorithm: a novel method of registration. In: Electronic Imaging and Multimedia Technology IV, vol. 5637, pp. 109–112. SPIE (2005)
8. Militello, C., et al.: A semi-automatic approach for epicardial adipose tissue segmentation and quantification on cardiac CT scans. Comput. Biol. Med. **114**, 103424 (2019)
9. Morin, R.L., Gerber, T.C., McCollough, C.H.: Radiation dose in computed tomography of the heart. Circulation **107**(6), 917–922 (2003)
10. Okrainec, K., Banerjee, D.K., Eisenberg, M.J.: Coronary artery disease in the developing world. Am. Heart J. **148**(1), 7–15 (2004)
11. Rodrigues, É.O., 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. **123**, 109–128 (2016)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
13. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: 2011 International Conference on Computer Vision, pp. 2564–2571. IEEE (2011)
14. Talman, A.H., Psaltis, P.J., Cameron, J.D., Meredith, I.T., Seneviratne, S.K., Wong, D.T.: Epicardial adipose tissue: far more than a fat depot. Cardiovas. Diagn. Ther. **4**(6), 416 (2014)
15. 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 **8**, 128032–128041 (2020)