A Cascade Approach for Automatic Segmentation of Coronary Arteries Calcification in Computed Tomography Images Using Deep Learning

Alan de C. Araújo[1(✉)], Aristófanes C. Silva[1], João M. Pedrosa[2], Italo F. S. Silva[1], and João O. B. Diniz[3]

[1] Applied Computing Group - Federal University of Maranhão (NCA/UFMA), São Luís, MA 65080040, Brazil
alan.araujo@nca.ufma.br, {ac.silva,italo.francyles}@ufma.br
[2] Institute for Systems and Computer Engineering, Technology and Science (INESC TEC), 4200-465 Porto, Portugal
joao.m.pedrosa@inesctec.pt
[3] Federal Institute of Education, Science and Technology (IFMA), Grajaú, MA 65940-000, Brazil
joao.bandeira@ifma.edu.br

Abstract. One of the indicators of possible occurrences of cardiovascular diseases is the amount of coronary artery calcium. Recently, approaches using new technologies such as deep learning have been used to help identify these indicators. This work proposes a segmentation method for calcification of the coronary arteries that has three steps: (1) extraction of the ROI using U-Net with batch normalization after convolution layers, (2) segmentation of the calcifications and (3) removal of false positives using Modified U-Net with EfficientNet. The method uses histogram matching as preprocessing in order to increase the contrast between tissue and calcification and normalize the different types of exams. Multiple architectures were tested and the best achieved 96.9% F1-Score, 97.1% recall and 98.3% in the OrcaScore Dataset.

Keywords: Coronary artery calcium · Segmentation · U-Net · EfficientNetB0 · OrcaScore Dataset

1 Introduction

Cardiovascular diseases are the leading causes of global mortality [4]. One indicator of the potential occurrence of cardiovascular disease is the amount of coronary artery calcification (CAC) [2]. CAC is commonly computed through computed tomography scans, which can be performed with or without contrast. The contrast-enhanced scan, known as coronary CT angiography (CCTA), involves the injection of an ionizing substance that highlights soft tissues, aiding in the

A. Cunha et al. (Eds.): MobiHealth 2023, LNICST 578, pp. 95–109, 2024.
https://doi.org/10.1007/978-3-031-60665-6_7

visualization of both soft tissues (heart, arteries, and veins) and hard tissues (bone structures). However, due to the use of a radioactive substance, this examination is contraindicated for certain patients. In such cases, a non-contrast CT scan (CSCT) is performed, which, in turn, exhibits a lower contrast between soft and hard tissues, making their visualization more challenging [3].

In the past, human interaction was required to differentiate calcium in the arteries from other calcifications, such as the aorta and bones [5]. Due to technological advancements, new methods of working with images and large amounts of data have emerged, and artificial intelligence has brought new approaches to problems involving imaging. Consequently, methods have been developed to automatically identify CAC, including deep learning-based approaches [7]. These approaches have been successfully applied to medical images, aiding in monitoring, particularly in detecting subtle findings that a physician might overlook and small details invisible to the naked eye [1].

Despite significant efforts to support the development of automatic CAC identification methods primarily based on deep learning, there is still room for new experiments and improvements. Therefore, the paper's main objective is to propose a cascaded method for segmenting calcifications in the coronary arteries. The proposed method utilizes histogram matching to enhance the contrast between calcification and tissue, simulating the contrast-enhanced scan, thus eliminating the need for the patient to undergo the contrast-enhanced examination. The method involves using a U-Net with batch normalization after the convolutional layers to extract the region of interest (ROI) and a U-Net with the encoder replaced by EfficientNetB0 for calcification segmentation. Additionally, a specific step is included to remove false positives, which are prevalent in this segmentation problem.

This paper is organized as follows: In Sect. 2, we discuss related works on calcifications in coronary arteries segmentation, exploring advancements in previously published methods. Section 3 provides detailed information about the proposed method for calcification segmentation. The results are presented in Sect. 4.3, and finally, Sect. 5 concludes the paper by summarizing the findings and discussing future work.

2 Related Works

In this section, we present works from the literature that are related to the proposed problem. The criteria for selecting related works were studies that employed fully convolutional networks (FCNs) for cardiac calcification segmentation and did not utilize contrast-enhanced examinations.

In [7], a DenseRAUnet is employed, which combines a Dense U-Net, ResNet, and dilated convolution. The images are preprocessed by resizing them to 512×512 pixels. The loss function is a combination of Bootstrap and IoU to effectively balance the background class and the calcification class. The model is trained on non-contrast thoracic CT scans in a 2.5D approach and evaluated on cardiac CT scans. The results show a remarkable F1-Score of 95.4% and a precision of

99.1%. However, it does not perform preprocessing to minimize the issues caused by the low contrast between tissues and calcifications in non-contrast CT scans.

In contrast, [2] introduces a multi-task model for simultaneous segmentation of both the coronary artery region and calcifications. The model was trained and evaluated on three datasets, namely DISCHARGE, OrcaScore, and CADMAN. During training, the authors utilized the weighted uncertainty loss [9]. The proposed method achieved an F1-Score of 92.8% for the calcification segmentation task. However, this method does not have a way to mitigate the differences in scans performed on different CT scanners, which hinders the learning process of the network.

In [3], the authors employ five 3D U-Net models trained from zero using the ADAM optimizer and Dice as the loss function. Each of them was trained with a different distribution os the dataset. The final result is obtained by majority voting among the five models. As a preprocessing step, the resolution of the scans is reduced to either 2.5 or 3 mm. To recover lost information, all voxels above 130 HU in the original image are intensified in the reduced-resolution image. A total of 783 patients are used for training the five 3D U-Net models. The method achieves an F1-Score of 97.4% on the OrcaScore Dataset. However, this model does not have a mechanism to prevent or remove false positives, which are common in this problem due to calcifications in regions such as the aorta and mitral valve.

Despite the numerous proposed methods, achieving accurate segmentation of calcifications remains challenging due to the low contrast between tissues and calcifications in contrast-free examinations, as well as the variations between scans acquired from different tomography machines. Consequently, developing a segmentation method that uses preprocessing to enhance the contrast between tissue and calcification in non-contrast CT scans, as well as normalize the various scans performed on different CT scanners, can facilitate their identification. Additionally, incorporating a specific step for false positive removal, which is common in this problem of coronary arteries calcification segmentation, would be beneficial.

3 Material and Methods

This section presents information regarding the dataset used, the proposed method, and the applied techniques.

3.1 Datasets

A separate image database was utilized for each step of the method. For the ROI extraction step, the CT Heart database, consisting of 1000 computed tomography scans with heart annotations, was employed. This database contains a total of 2532 slices. The dataset was divided into 60% for training, 20% for validation, and 20% for testing. Three data augmentation techniques were applied during training, including horizontal flipping, vertical flipping, and 45° rotation. Ultimately, the distribution consisted of 6076 images for training, 506 images for validation, and 506 images for testing.

For the calcification segmentation step, the OrcaScore database was used, which consists of computed tomography scans from 32 patients, totaling 1540 slices. This database includes scans acquired from four different tomography machines in four distinct hospitals, providing a greater diversity of examinations. The dataset was divided into 20 patients for training, 4 patients for validation, and 8 patients for testing. Data augmentation techniques were applied to the training set, including horizontal flipping, vertical flipping, and 45° rotation. Ultimately, the distribution consisted of 4248 images for training, 150 images for validation, and 301 images for testing.

The image database for the false positive removal step was formed through preprocessing, which involved extracting a 96×96 pixel ROI from the predictions made in the calcification segmentation step, as well as ROIs from the masks in the OrcaScore database. Thus, the database consists of 198 true-positive images and 32 false positive images. The distribution of true-positive images is 80% for training, 10% for validation, and 10% for testing. As for the false positive images, they are divided as follows: 60% for training, 20% for validation, and 20% for testing. Due to the limited amount of data, data augmentation techniques are applied to generate new images. For the true-positive training set, horizontal flipping, vertical flipping, and 45° rotation techniques are performed, resulting in a total of 540 true-positive images. For the false positive set, a larger number of images is generated to balance the database. The augmentation techniques applied include horizontal flipping, vertical flipping, and rotation from 10 to 70° in 5° increments, resulting in a total of 240 false positive images.

3.2 Proposed Method

The proposed method follows a cascaded approach, whereby the output of each step serves as the input for the subsequent step. It is divided into three steps, namely ROI extraction, calcification segmentation, and false positive removal. Each step is designed to address a specific aspect of the task, as illustrated in Fig. 1.

3.2.1 ROI Extraction

Computed tomography scans can vary significantly depending on the tomography machine, vendor, hospital, and other factors. Therefore, histogram matching [12] was applied to make the diverse scans more similar to each other and enhance the contrast between the ROI and other regions. Additionally, this preprocessing step also increases the contrast between calcifications and surrounding tissue, which aids in subsequent stages. To determine the reference image, we performed an initial ROI segmentation. The patient slice with the highest F1-Score in this initial segmentation was selected as the reference image. An example of the application of histogram matching can be seen in Fig. 2.

The architecture of the network used in the first stage was the modified U-Net proposed by [6]. The structure consists of down-sampling and up-sampling

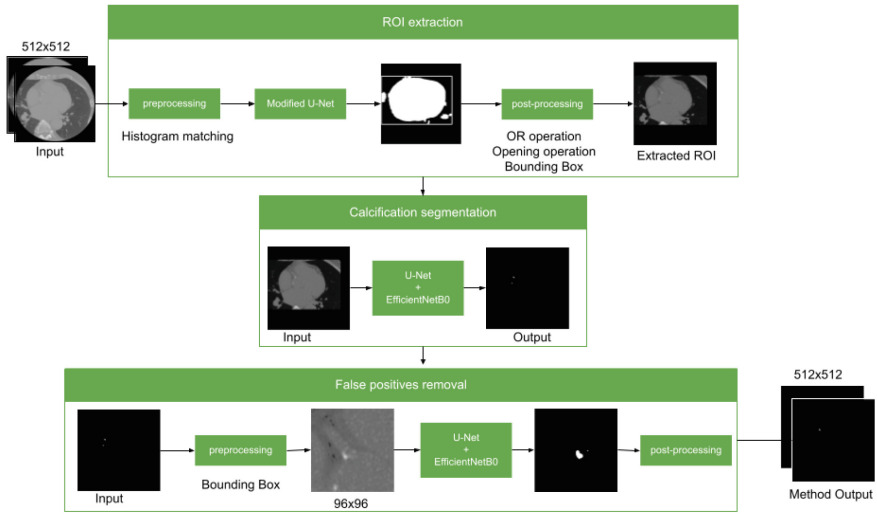


Fig. 1. Overview of the proposed method.

(a) (b) (c)

Fig. 2. Histogram matching application: (a) original image (b) reference image e (c) modified histogram image.

layers, giving it a U-shaped format. The down-sampling blocks consist of successive 3×3 convolution layers and batch normalization layers, followed by rectified linear unit (ReLU) activation functions and 2×2 max-pooling operations. The number of filters is doubled in each block, starting with 64 filters and reaching a total of 1024. On the other hand, the up-sampling blocks consist of 2×2 convolution layers, with each respective block from the down-sampling section having a corresponding 3×3 convolution layer. The last layer includes a transposed convolution that outputs an image with the same dimensions as the input image.

After the initial segmentation of the ROI, a final mask is formed by performing an OR morphological operation [12] on all the segmentations of the patient's slices. This means that for each pixel in the final mask, if at least one segmentation predicts it as part of the ROI, it will be marked as belonging to

the ROI. An example of the OR operation applied to the problem can be seen in Fig. 3.

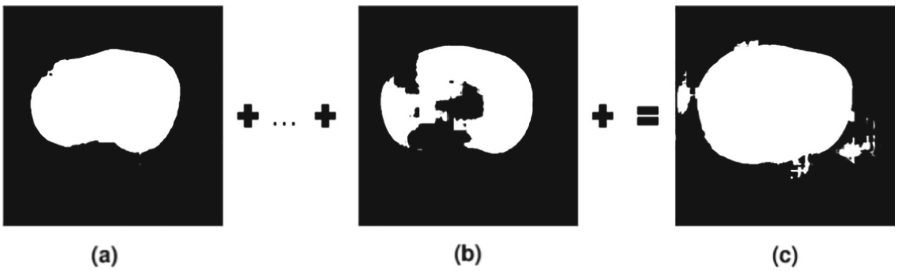


Fig. 3. Example of the OR morthological operation: (a) slice 1, (b) slice N, and (c) final mask.

In addition, an opening operation [12] is applied to eliminate possible artifacts present in the final mask. In this step, a fixed-size rectangular structuring element of 7×7 is used. An example of the opening operation can be seen in Fig. 4.

It can be observed that the operation removes small objects and thin lines from the image while preserving the shape and size of larger objects. Finally, a bounding box is created around the final mask to ensure that the ROI has been extracted, as shown in Fig. 5.

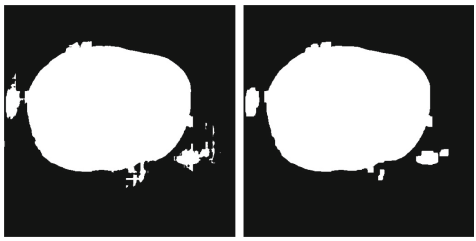
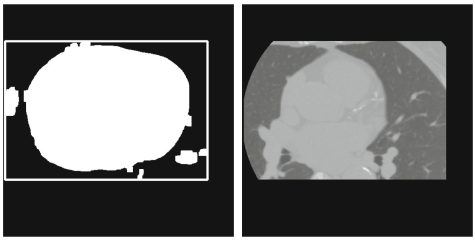


Fig. 4. Example of opening operation: (a) original image and (b) resulting image.

3.2.2 Calcification Segmentation

In the calcification segmentation stage, the initial segmentation of calcifications in the coronary arteries is performed using the chosen architecture. Due to the fact that the input image in this stage is only the ROI extracted in the previous stage, some false positives are avoided in regions such as the sternum and spine. However, there are still false positives within the ROI, mainly calcifications in the aorta.



(a) (b)

Fig. 5. Example of: (a) bounding box after postprocessing and (b) extracted ROI from first step

The neural network architecture used in the second and third stages was the modified U-Net with the encoder path based on EfficientNetB0. The standard U-Net structure is present in the up-sampling layers, while the down-sampling part features the EfficienteNetB0 model, which includes a 3×3 convolutional layer with 64 filters. Following that, there are 5 blocks of deep convolution, each consisting of a 3×3 convolutional layer and a 1×1 convolutional layer. Each decoding block contains a transposed convolutional layer of size 2×2 obtained from the previous layer with a stride of 2, concatenated with a 1×1 convolution for each respective block in the down-sampling section. In the last layer, a transposed convolution takes place, returning an output image with the same dimensions as the input image.

3.2.3 False Positive Removal

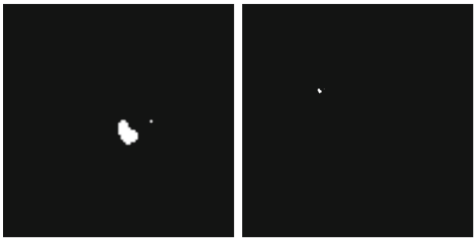
In the false positive removal step, preprocessing and post-processing techniques are applied to remove false positives caused mainly by calcifications in other regions of the ROI that are not part of the coronary arteries, such as the aorta.

During the preprocessing sub-step, a bounding box is created around the predictions made in the calcification segmentation step. These bounding boxes are expanded to 96×96 pixels size and then an ROI is formed from the 512×512 image as shown in Fig. 6. In this way, the ROI provides contextual information about the location of the calcification segmentation, aiding in its identification.

During the post-processing sub-step, the output images of this step are 96×96 pixels. Then, these images are placed in the same location in the original image from which they were extracted (512×512 image), resulting in an output image of the method that is 512×512 pixels, the same size as the input image (Fig. 7).

(a) (b) (c)

Fig. 6. Example of: (a) 512×512 image with bounding box represented by red square (b) bounding box 96×96 and (c) 96×96 ROI extracted from original image



(a) (b)

Fig. 7. Example of: (a) output image 96×96 and (b) method output image 512×512

4 Experiment Setup

This section presents information regarding the configurations of the experiments such as dataset division, the metrics used to evaluate the method and the results of differents architectures and between other methods.

4.1 Experiments Setup

For the training of the ROI extraction step, the Dice loss [11] was chosen as the loss function due to its superior performance compared to the Focal loss and cross-entropy. The CT Heart dataset was used, consisting of 6076 images for training, 506 images for validation, and 506 images for testing. Training was conducted for 70 epochs, with a batch size of 2 and a learning rate of 1e−4. Additionally, the Early Stopping technique was employed with a patience of 7 epochs. These experiments, along with all others, were conducted on hardware with the following configuration: an NVIDIA GTX 1660 Super graphics card with 6 GB of VRAM.

For the second step training, the focal loss [10] was used as the loss function because it provides better results when there is class imbalance. The OrcaScore

dataset was used, consisting of 4248 images for training, 150 images for validation, and 300 images for testing. Training was conducted for 70 epochs, with a batch size of 4 and a learning rate of 1e−4. Was also used the Early Stopping technique employed with a patience of 5 epochs.

For the training of the third step, Focal Loss was used. The dataset was divided into 1266 images for training, 36 images for validation, and 36 images for testing. Training was conducted for 50 epochs, with a batch size of 24 and a learning rate of 1e−4. Additionally, the Early Stopping technique was employed with a tolerance of 7 epochs.

4.2 Results Evaluation

For the results evaluation, the metrics of F1-Score, precision, and recall were used [8]. Precision measures the proportion of correct positive class predictions compared to the total number of samples classified as positive (Eq. 1). Recall measures the proportion of true positive samples that were correctly classified by the model compared to the total number of positive samples (Eq. 2). F1-Score is the harmonic mean between precision and recall (Eq. 3).

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{VP + FN} \tag{2}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3}$$

where TP stands for true positives, FP, in turn, stands for false positives, and FN represents false negatives.

4.3 Results and Discussion

In Table 1, an improvement can be observed in the modified U-Net compared to the original U-Net for the ROI extraction task. It is important to note that these metrics were obtained by testing only on the CT Heart dataset, as it contains heart annotations. The ROI extraction was performed on the OrcaScore dataset, and it can be visualized in Fig. 8.

Table 1. Comparison between architectures for ROI extraction

Image	F1-Score (%)	Precision (%)	Recall (%)
Original U-Net	63,0	83,2	93,3
Modified U-Net	**85,0**	**85,4**	**96,8**

(a) (b) (c)

Fig. 8. ROI extraction examples in OrcaScore dataset: (a) patient 1 (b) patient 2 and (c) patient 3

Tests were conducted with images with and without histogram matching in order to validate the improvement of metrics when different types of images are more similar to each other, as well as to enhance the contrast between calcifications and tissues. The comparison between these tests can be visualized in Table 2.

Table 2. Comparison between images with and without histogram matching

Image	F1-Score (%)	Precision (%)	Recall (%)
Without histogram matching	12,3	11,3	13,7
With histogram matching	**81,7**	**84,2**	**95,0**

After comparing the preprocessing, experiments were conducted using other architectures, specifically the U-Net with other modifications in the encoder path, such as ResNet-50 and EfficientNetB3. This allowed us to evaluate architectures with more parameters and deeper structures. Table 3 shows the metrics obtained by each network in the task of segmenting coronary arteries calcifications.

Table 3. Comparison between architectures for calcification segmentation

Architecture	F1-Score (%)	Precision (%)	Recall (%)
U-Net + Resnet-34	95,8	97,9	96,6
U-Net + Resnet-50	95,6	**98,2**	95,4
U-Net + Resnet-101	94,6	97,5	96,1
U-Net + EfficientNetB0	**96,6**	98,0	97,2
U-Net + EfficientNetB3	96,0	97,5	**97,5**
U-Net + EfficientNetB5	90,7	93,4	90,5

The modified U-Net with EfficientNetB0 was chosen because it yielded a higher F1-Score, thus achieving a higher harmonic mean between Precision and Recall. After confirming the effectiveness of the architecture for segmenting calcifications in coronary arteries, the method was validated using cross-validation. Table 4 displays the metric values, mean, and standard deviation for each validation fold. Since the train-test split was 75%/25%, cross-validation was performed with 4 folds.

Table 4. Cross validation results

Fold number	F1-Score (%)	Precision (%)	Recall (%)
1	94,8	95,6	97,2
2	93,1	96,2	95,7
3	**96,6**	**98,0**	**97,2**
4	90,8	94,4	93,3
Mean	93,8	95,9	96,2
Standard Deviation	2,4	1,2	2,1

The model analyzed so far had some false positive predictions, mainly in the central area of the heart, where there were calcifications in the aorta, as observed by the red arrows in Fig. 9. These calcifications are not part of the coronary arteries, so even though they are calcifications, they are false positives that should be avoided. In this context, the third step of the model (false positive removal) is necessary.

(a) (b) (b)

Fig. 9. Calcification segmentation example: (a) original image (b) ground-truth and (c) prediction

A comparison between the model with and without the false positive removal step can be observed in the Table 5.

Table 5. Comparison between model with and without false positive removal step

Model	F1-Score (%)	Precision (%)	Recall (%)
Without false positive removal step	96,6	98,0	**97,2**
With false positive removal step	**96,9**	**98,3**	97,1

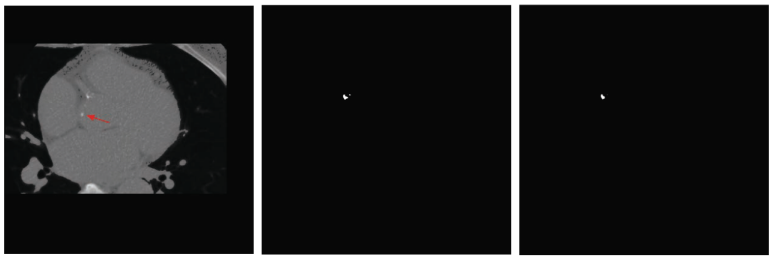
As can be seen, there was an improvement in F1-Score and Precision due to the fact that some false positives were removed. However, Recall decreased due to the loss of some true positive pixels. Despite the loss of some pixels, the calcifications continue to be identified, thus maintaining the objective of segmenting them.

Similarly as made in the second step, experiments with the multiple architectures were conducted, such as U-Net with modifications in the encoder path (ResNet-34 and EfficientNetB5). Table 6 shows the metrics obtained by each network in the task of false positive removal.

Table 6. Comparison between architectures for false positive removal step

Architecture	F1-Score (%)	Precision (%)	Recall (%)
U-Net + Resnet-34	95,9	97,5	96,8
U-Net + Resnet-50	95,8	97,2	96,4
U-Net + Resnet-101	96,1	97,6	97,0
U-Net + EfficientNetB0	96,8	98,2	97,0
U-Net + EfficientNetB3	96,2	97,3	96,9
U-Net + EfficientNetB5	**96,9**	**98,3**	**97,1**

The modified U-Net with EfficientNetB5 was chosen because it yielded a higher F1-Score, Recall and Precision, thus removing false positives with the less loss of pixels into the true positives. An example of an image after the false positive removal step can be seen in the Fig. 10. In this figure, the presence of a calcification in the aorta, indicated by the red arrow, is not predicted as a calcification by the model, demonstrating the effectiveness of the false positive removal step.



(a) (b) (c)

Fig. 10. Calcification segmentation example: (a) original image (b) ground-truth and (c) prediction

4.4 Related Works Comparison

In the calcifications segmentation, the proposed method achieves positive results close to other previous works that were tested in the same dataset (OrcaScore), as can be seen in the Table 7.

Table 7. Methods on the segmentation of coronary arteries calcifications

Method	F1-Score (%)	Precision (%)	Recall (%)
[2]	92,8	98,4	96,1
[3]	**97,5**	**99,5**	96,8
[7]	95,4	99,1	91,1
Proposed method	96,9	98,3	**97,1**

The proposed method exhibited metrics close to those of [3]. His method employs 5 3D U-Nets for segmentation voting, which requires high computational cost and powerful hardware configuration, as well as utilizing a more complex architecture compared to the proposed method. Based on this, the proposed method outperformed [3] recall metric, which is used to evaluate true positives. This is due to the use of histogram matching and the proposed architecture, and the specific step for false positive removal made the precision increased. It is worth noting that the proposed method also surpassed the F1-Score of [2] and [7].

5 Conclusion

The study aimed to develop a method for segmenting coronary artery calcifications in the OrcaScore Dataset. Several experiments were conducted to improve the results, including preprocessing, data augmentation, post-processing, and

exploring different architectures. The results presented are based on the best-performing metrics obtained thus far.

The experiments using the U-Net and modified U-Net architectures, along with ROI extraction, histogram matching, and false positive removal, yielded positive outcomes. It is worth noting that only CSCT exam images were used in this work, excluding the less common CCTA exams that involve contrast. Histogram matching, a preprocessing step not employed in previous studies, showed noticeable improvements in the results. Additionally, a dedicated false positive removal step contributed to higher sensitivity metrics.

As future work, it is suggested to explore new architectures, such as attention blocks or hybrid networks with DeepLab [13], since it uses dilated convolutions and captures information from broader contexts without increasing computational complexity, in addition to showing promising results in segmentation tasks.

Acknowledgements. The authors acknowledge the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil, and Fundação de Amparo à Pesquisa Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA) (Brazil), Empresa Brasileira de Serviços Hospitalares (Ebserh) Brazil (Grant number 409593/2021-4) for the financial support.

References

1. de Vos, B.D., Wolterink, J.M., Leiner, T., de Jong, P.A., Lessmann, N., Išgum, I.: Direct automatic coronary calcium scoring in cardiac and chest CT. IEEE Trans. Med. Imaging **38**(9), 2127–2138 (2019)
2. Follmer, B., et al.: Active multitask learning with uncertainty-weighted loss for coronary calcium scoring. Med. Phys. **49**(11), 7262–7277 (2022)
3. Gogin, N., et al.: Automatic coronary artery calcium scoring from unenhanced-ECG-gated CT using deep learning. Diagn. Intervent. Imaging **102**(11), 683–690 (2021)
4. Wang, H., et al.: systematic analysis for the global burden of disease study 2015. Lancet **388**(10053), 2127–2138 (2019)
5. Wang, W., et al.: Coronary artery calcium score quantification using a deep-learning algorithm. Clin. Radiol. **75**(3), 237-e11 (2020)
6. Yoshida, A., Lee, Y., Yoshimura, N., Kuramoto, T., Hasegawa, A., Kanazawa, T.: Automated heart segmentation using U-net in pediatric cardiac CT. Sensors **18**(3), 100127 (2021)
7. Zhang, W., Zhang, J., Du, X., Zhang, Y., Li, S.: An end-to-end joint learning framework of artery-specific coronary calcium scoring in non-contrast cardiac CT. Computing **101**(3), 667–678 (2020)
8. Chinchor, N., Sundheim, B.M.: MUC-5 evaluation metrics. In: Fifth Message Understanding Conference (MUC-5): Proceedings of a Conference Held in Baltimore, Maryland, pp. 25–27 (2020)
9. Kendall, A., Gal, Y., Cipolla, R.: Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7482–7491 (2018)

10. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
11. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Cardoso, M.J., et al. (eds.) DLMIA/ML-CDS -2017. LNCS, vol. 10553, pp. 240–248. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_28
12. Gonzalez, R.C.: Digital Image Processing, 2nd edn. Pearson Education, India (2009)
13. Wang, C., et al.: A three-stage self supervised deep learning network for automatic calcium scoring of cardiac computed tomography images. In: 2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA). IEEE (2022)