Expert Systems With Applications 197 (2022) 116704

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A cascade approach for automatic segmentation of cardiac structures in short-axis cine-MR images using deep neural networks

Italo Francyles Santos da Silva [a,*], Aristófanes Corrêa Silva [a], Anselmo Cardoso de Paiva [a], Marcelo Gattass [b]

[a] Applied Computing Group (NCA - UFMA), Federal University of Maranhão, Av. dos Portugueses, SN, Bacanga, 65085-580, São Luís, MA, Brazil
[b] Pontifical Catholic University of Rio de Janeiro, R. São Vicente, 225, Gávea, 22453-900, Rio de Janeiro, RJ, Brazil

ARTICLE INFO

Keywords:
Cardiac structures segmentation
Fully convolutional networks
Cine-MRI
ACDC dataset

ABSTRACT

Cardiovascular diseases are responsible for millions of deaths every year. In this scenario, non-invasive exams such as cine-magnetic resonance imaging (cine-MRI) have favored a better understanding of these pathologies, helping early diagnosis and previous treatments essential to improve the quality of life of individuals. Through this exam, specialists can obtain more accurate information about cardiac structures, including the myocardium, the left ventricular cavity, and the right ventricle. Given this context, this work presents an automatic method for the segmentation of these cardiac structures in short-axis cine-MRI images. The proposed method uses a cascade approach and is therefore divided into three main steps. The first step consists of extracting a region of interest to reduce the scope of processing. The second applies a fully convolutional network proposed to generate the initial segmentations of the myocardium, left ventricular cavity, and right ventricle. These initial segmentations are passed on to the third step, called refinement, in which a mask reconstruction module based on U-Net is used to restore the generated segmentations. In addition, in this step some specific post-processing techniques are also applied for each structure of interest. The proposed method achieves promising results in tests with the ACDC challenge dataset, both at the local level, and in the evaluation made by the challenge's own platform, in which the proposed method proves to be competitive with the best approaches.

1. Introduction

Cardiovascular diseases (CVDs) correspond to a group of problems related to the heart and blood vessels. They appear according to age and are linked to a set of factors such as eating habits, sedentary lifestyle, or even genetic factors. Worldwide, CVDs are responsible for 17.5 million deaths per year (Hazra, Mandal, Gupta, Mukherjee, & Mukherjee, 2017). Therefore, the sooner they are diagnosed and treated, the greater the chances of improving the quality of life of individuals.

The continuous advancement of technologies has allowed a better understanding of the pathologies that affect people, mainly through non-invasive exams, such as magnetic resonance imaging (MRI). In the cardiac context, this exam is also called cine-magnetic resonance (cine-MR) and it is considered one of the safest methods for cardiovascular diagnosis through the analysis of the anatomy and morphology of the heart, as well as for the identification of congenital CVD (Faridah Abdul Aziz et al., 2013; Sara et al., 2014).

By means of cine-MR, specialists can get more accurate information about cardiac structures, among them the right ventricle (RV), left ventricular cavity (LVC), and myocardium (Myo). This is also important in the postoperative follow-up, with the evaluation of cardiac functions and ventricular anatomy (Sechtem, Pflugfelder, Gould, Cassidy, & Higgins, 1987). However, obtaining this information can be a time-consuming process, in which the specialist needs to analyze many images and manually delimit the regions of interest (ROI) to analyze them. In addition, this procedure tends to fatigue the specialist, reducing the accuracy of his evaluation.

In image processing, the delimitation of a region of interest is called segmentation. In the context of CVDs identification, computational methods can assist specialists by providing an automatic segmentation to speed up the analysis of exams and the evaluation of cardiac structures.

Among these methods, it is possible to highlight those based on Deep Learning, such as fully convolutional networks (FCN) that have

* Corresponding author.
E-mail addresses: francyles@nca.ufma.br (I.F.S. da Silva), ari@nca.ufma.br (A.C. Silva), paiva@nca.ufma.br (A.C. de Paiva), mgattass@tecgraf.puc-rio.br (M. Gattass).

https://doi.org/10.1016/j.eswa.2022.116704
Received 18 May 2021; Received in revised form 10 February 2022; Accepted 19 February 2022
Available online 4 March 2022

shown expressive results in different applications, for example, image and video object detection (Caelles et al., 2017; Long, Shelhamer, & Darrell, 2015) and also in medical image segmentation (Oktay et al., 2018; Ronneberger, Fischer, & Brox, 2015). FCNs are based on the feature extraction from the most basic to the most specific level, analyzing latent patterns in order to, from that, identify which pixels are part of the regions of interest.

Still in the field of medical images, FCNs can be applied in exams with visualization in two or more dimensions. However, depending on the proposed solution, processing these exams may require high computational costs. Therefore, developing an accurate and computationally viable method becomes a challenging task.

In this scenario, this work presents a method for cardiac cine-MR image segmentation. This method uses a cascade approach and is divided into three main steps: ROI extraction, initial segmentation, and refinement. In the first step, an approach based on U-Net (Ronneberger et al., 2015) is used to locate the ROI. In the second step, the extracted ROI is submitted to a proposed FCN that uses an Efficient-Net (Tan & Le, 2019) and the convolutional blocks Inception (Szegedy et al., 2015), Attention (Oktay et al., 2018), and Squeeze-and-Excitation (Hu, Shen, & Sun, 2018). The proposed FCN generates an initial segmentation that is passed as input to the third step which is responsible for improving the obtained results using post-processing techniques, including a mask reconstruction module based on U-Net. Finally, the final segmentation consists of the generated masks for the myocardium, left ventricular cavity, and right ventricle.

This work has the following contributions: (1) the proposition of a cascade method composed of three steps for the segmentation of cardiac structures, in which each step uses a domain-specific approach; (2) a U-Net-based approach for locating and extracting ROI that aims to reduce scope and processing; (3) the proposed FCN architecture for initial segmentation that combines Efficient-Net with Inception, Attention and SAE blocks and skip connections; (4) the proposed mask reconstruction module based on U-Net to refine the initial segmentation of the cardiac structures.

The present work is organized as follows: in Section 2, the related works are presented; in Section 3, the steps of the proposed method are depicted; Section 4 shows the obtained results; in Section 5, we present a discussion; and, lastly, the conclusion and future works are presented in Section 6.

2. Related works

In the literature, there are many works presenting methods based on deep learning for cardiac cine-MR segmentation. Some of them are directed to the segmentation of LV contours while others take into account the LVC and the RV, performing this task in one or more steps.

Tran (2016) proposed a method that was a pioneer in the use of FCN for the segmentation of the left and right ventricles. The FCN architecture implemented is the same proposed by Long et al. (2015) with no relevant changes. This method assumes a priori that the ventricular cavity is always in the center of the images and uses that information to extract the ROI that involves the ventricles. However, that structure is not centralized in all cases, which is the main justification for the segmentation failures presented by the method. The results achieved Dice coefficients of 0.92 and 0.96, respectively, for the segmentation of the endocardium and epicardium in experiments with the Sunnybrook Cardiac Dataset (Radau et al., 2009).

Tan, Liew, Lim, and McLaughlin (2017) proposed a semi-supervised method that uses a CNN to predict the central point of the LV and, from there, perform the ROI extraction that is passed as an input to the segmentation step. In this process, the ROI is converted from the Cartesian to the polar space and then serves as an input for another CNN responsible for delimiting the LV contours via regression. This method obtained Jaccard index of 0.74 in experiments with the LVSC dataset (Suinesiaputra et al., 2014). Despite the robustness and accurate results, this approach requires human intervention in its initial steps, and therefore is more susceptible to failure.

The polar space is also explored in the method proposed by Hu, Pan, Wang, Yin, and Ye (2019). The authors implemented an approach divided into two steps. The first consists of the ROI extraction from a coarse segmentation generated by SegNet (Badrinarayanan, Kendall, & Cipolla, 2017). Then, the ROI and mask are converted to polar space and used as input to a refinement module in which an optimization process based on edge mapping and dynamic programming is applied. In experiments with the Sunnybrook Cardiac Dataset this method achieved Dice coefficients of 0.90 and 0.93 for segmentation of the endocardium and epicardium, respectively.

Abdeltawab et al. (2020) also proposed a method divided into two steps, first locating the LVC to extract an ROI that involves this structure and the myocardium (Myo). This region is then used as an input to the segmentation step based on FCN. The experiments were performed with a locally-acquired dataset. The segmentation process was evaluated considering two cardiac phases: end systole (ES) and end diastole (ED). This method obtained Dice coefficients of 0.96 and 0.92 for the segmentation of the LVC in ES and ED. In the case of Myo, the Dice results were 0.88 (ES) and 0.89 (ED).

The aforementioned works present methods aimed only at the context of the left ventricle. The approaches presented below are related to the segmentation of other cardiac structures. These methods were validated with the ACDC dataset (Bernard et al., 2018). The metrics also were calculated considering the cardiac phases ES and ED.

Isensee et al. (2017) proposed an ensemble of 2D and 3D U-Nets for segmenting RV, LVC, and Myo. The 2D and 3D models receive as input samples with high dimensions, what requires a high computational cost that is not always feasible. As a result, this method obtained respectively for ED and ES Dice coefficients of 0.967 and 0.928 in LVC segmentation; 0.904 and 0.923 for the Myo; 0.951 and 0.904 for the RV.

Baumgartner, Koch, Pollefeys, and Konukoglu (2017) also carried out experiments with the U-Net (2D and 3D) and FCN-8s architectures each with specific dimensions for the input, assuming that the input size reduction can result in loss of information. The segmentation results reached the following Dice coefficients in the ED and ES phases: 0.963 and 0.911 for the LVC; 0.892 and 0.901 for Myo; 0.932 and 0.883 for the RV.

The method proposed by Calisto and Lai-Yuen (2020) consists in an adaptive ensemble of 2D and 3D FCNs, called AdaEn-Net, applied to medical image segmentation, including cardiac cine-MRI. In the adaptation process, a multi-objective optimization algorithm is performed, which determines kernel sizes, number of filters and blocks, thereby defining the width and depth of the network. As a result, this approach obtained Dice coefficients in the ED and ES phases, respectively, 0.958 and 0.903 (LVC); 0.873 and 0.895 (Myo); 0.936 and 0.884 (RV).

The approach proposed by Zotti, Luo, Humbert, Lalande, and Jodoin (2017) presents the Grid-Net architecture which is based on U-Net. However, the skip connections are replaced by convolutional layers in order to increase the number of relevant feature maps. In the ED and ES phases, this approach obtained Dice coefficients of 0.964 and 0.912 (LVC); 0.886 and 0.902 (Myo); 0.941 and 0.882 (RV).

Khened, Kollerathu, and Krishnamurthi (2019) addressed the cardiac cine-MRI segmentation with a method based on Multi-scale Residual Dense-Nets. The proposal uses a densely connected FCN to produce more feature maps while avoiding the problem of gradient explosion. The results obtained by this method for the ED and ES phases in relation to the Dice coefficient were 0.964 and 0.912 (LVC); 0.886 and 0.902 (Myo); 0.941 and 0.882 (RV).

Simantiris and Tziritas (2020) presented a method composed of two steps. First, an ROI extraction process inspired by Grinias and Tziritas (2017) is performed. Next, the ROI extracted is submitted to the segmentation step that uses a CNN with dilated convolutions, called DCNN. According to the authors, dilated convolutions are used

Table 1
An overview of the related works.

Related works	Methods
Tran (2016)	FCN
Tan et al. (2017)	LV segmentation in polar space using CNN
Hu et al. (2019)	SegNet + refinement based on optimization
Abdeltawab et al. (2020)	FCN and modified U-Net for segmentation
Isensee et al. (2017)	Ensemble of 2D and 3D U-Nets
Baumgartner et al. (2017)	2D and 3D U-Nets
Calisto and Lai-Yuen (2020)	AdaEn-Net
Zotti et al. (2017)	Grid-Net
Khened et al. (2019)	Multi-Scale Residual Dense-Net
Simantiris and Tziritas (2020)	CNN with dilated convolutions
Proposed method	U-Net for ROI extraction + initial segmentation via proposed FCN + Refinement with reconstruction module

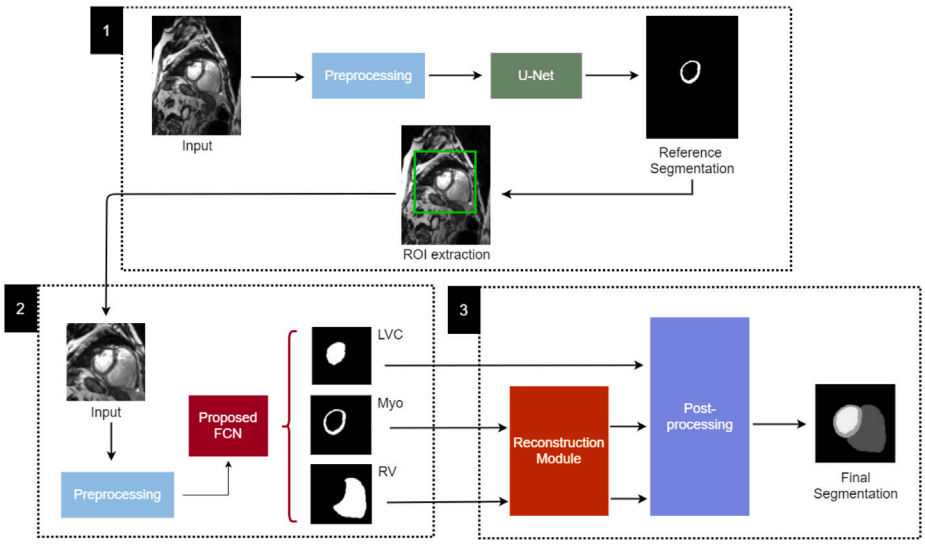

Fig. 1. An overview of the proposed method and its steps: (1) ROI extraction, (2) initial segmentation, and (3) refinement.

to preserve localization accuracy and reduce the number of trainable parameters. As a result, for the ED and ES phases, respectively, this method obtained Dice coefficients of 0.964 and 0.912 (LVC); 0.886 and 0.902 (Myo); 0.941 and 0.882 (RV).

The method proposed by the present work is inspired by some of the mentioned approaches and also divides the segmentation process into steps. However, different from the related works, each step was developed considering specific features of the exam slices and the structures of interest, and using techniques capable of providing promising results without significantly increasing computational cost. For the ROI extraction, an approach based on U-Net is used. For the initial segmentation process, it is proposed an FCN architecture that combines the Efficient-Net with Inception, Attention, and Squeeze-And-Excitation blocks. In addition, as a refinement, it is proposed a mask reconstruction module based on U-Net. An overview of the related works is shown in Table 1.

3. Proposed method

The proposed method for the segmentation of the cardiac structures in cine-MRI can be seen more generally in Fig. 1. This method is composed of three main steps: The first focuses on reducing the scope of the image, extracting a small region that encompasses the structures of interest which are the LVC, Myo, and RV. The second step comprises the generation of the initial LVC, Myo, and RV segmentations that will be passed on to the third step responsible for the refinement process that uses the reconstruction module and postprocessing techniques.

3.1. ROI extraction

The first step of the proposed method consists in the extraction of a small region (ROI) that encompasses LVC, Myo, and RV. This process intends to reduce the scope of each volume, eliminating background regions that may interfere with the learning of the segmentation model.

This step also aims to reduce the computational cost of the proposed solution, as it avoids dealing with the slices in their original size. Another advantage is the mitigation of pixel class imbalance, a common problem in medical image processing (Gao, Zhang, Liu, & Wu, 2020), by the elimination of more background regions.

In the ROI extraction step, each slice is submitted to U-Net (Ronneberger et al., 2015). This network is utilized to produce masks that will serve as a reference to locate the most appropriated bounding box for the structures of interest. These processes will be detailed below.

Before being submitted to U-Net, the slices go through a preprocessing that consists of three procedures: resizing, outlier removal, and normalization. First, slices are resized to 160×160 due to hardware limitations. Next, outliers, which are peaks of intensity inherent in cine-RM, are removed. For this, a technique similar to that proposed by Nasr-Esfahani et al. (2018) is applied, in which, for each slice, a range of pixel values between 0 and 70% of the highest intensity is stipulated. Then, all pixels out of this range are included receiving the maximum or minimum values established. Finally, normalization is performed to change the range of pixel intensity values to be between 0 and 1 according to Eq. (1).

$$N(x,y) = \frac{M(x,y) - \min(M)}{\max(M) - \min(M)} \tag{1}$$

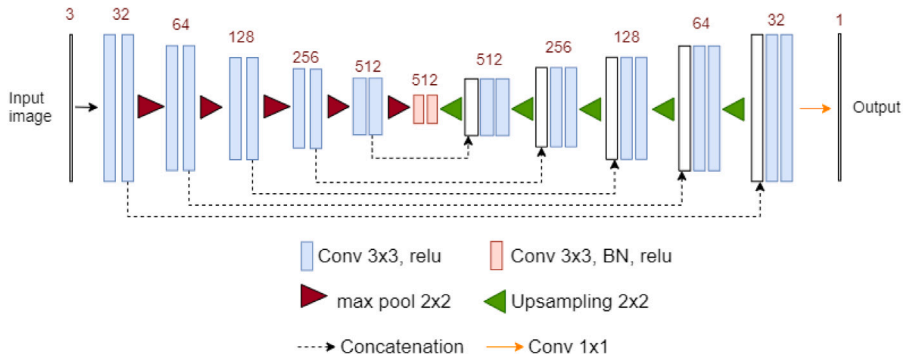

Fig. 2. The U-Net architecture used to generate reference segmentations.

where M (x, y) is the pixel value on the M slice and N (x, y) is the corresponding value on the normalized slice.

The U-Net used to generate the reference segments has a simple architecture (Fig. 2), having VGG16 (Simonyan & Zisserman, 2014) as its backbone, with 5 levels of convolutional blocks followed by Max Pooling 2D operations with 2×2 kernel. The convolutional blocks are composed of two layers of 2D convolution followed by the Relu (Nair & Hinton, 2010) activation function. Batch Normalization is used in the deepest blocks as it is important to avoid overfitting and accelerate the convergence of the model (Santurkar, Tsipras, Ilyas, & Madry, 2018).

The expansion path also has 5 levels of convolutional blocks, of which the first 4 are composed of Upsampling with kernel 2×2, followed by two layers of 3×3 convolution and Relu activation. The feature maps generated by each block in the expansion path are concatenated to those produced in the corresponding level in the contraction path. Thus, it is possible to retrieve features possibly lost due to Max Pooling (e.g., spatial information) and increase the number of maps in the last layers. The final layer is a 1×1 convolution and uses the sigmoid activation function to classify the pixels between 0 and 1, the latter representing the structure of interest.

For the training of U-Net, only the medial slices of MRI volume were selected. This choice is due to the fact that the structures of interest are always visible in the medial slices, which do not always occur in the case of the apical and basal types. In addition to visibility, the structures are presented in a more defined way in relation to the other regions of the image, allowing the segmentation process to produce more accurate results in these cases. Moreover, the network was trained to generate masks only for the myocardium instead of the three structures, because, in the medial slices, Myo and the LVC are more centralized in comparison to the right ventricle. It is worth mentioning that the learning process becomes less complex since only one region will be segmented.

Therefore, for a given volume, all slices will be submitted to U-Net. As a result, this network will generate masks only for the medial slices. These masks are called reference segmentations, and, from them, the ROI to be extracted will be defined.

As these masks may have some noise, it is necessary to remove them. For this, a morphological erosion is applied using a rectangular structuring element of size 3×3. After this procedure, the reference segmentation that has the largest area will be found. From its center, a bounding box will be defined whose dimensions vary according to the original size of the slice before preprocessing. Through tests, the following bounding box sizes were established: 120×120 for slices with an original size less than 200×200; 140×140 for those whose size is less than 250×250; and 160×160 when these cases are not satisfied. Finally, the defined bounding box will be replicated throughout the volume, serving as a basis for the ROI extraction.

At the end of this process, the extracted ROIs are passed as input to the second step of the proposed method, responsible for generating the initial segmentation of the cardiac structures.

3.2. Initial segmentation

ROIs extracted in the previous step are passed as input to an FCN for the segmentation of the LVC, Myo, and RV structures. This network has contraction and expansion paths, and skip connections as well as U-Net. However, the proposed FCN has a modified structure. An overview of the proposed FCN architecture can be seen in Fig. 3.

Before being passed to the FCN, ROIS are preprocessed. First, zero padding is applied for resizing images to the same size. Next, the aforementioned outlier removal and normalization (Section 3.1) are performed. After these processes, the input images will have dimensions 160×160 and three channels.

The feature extraction process in the contraction path is performed by Efficient-Net B3 (Tan & Le, 2019). The Efficient-Net family is known for the high accuracy achieved in the ImageNet and ImageNet-V2 challenges (Recht, Roelofs, Schmidt, & Shankar, 2019). These architectures introduce a model scaling method based on a compound coefficient. Differently from traditional approaches in which the network dimensions are arbitrarily defined, Efficient-Nets were developed using a heuristic method based on Grid Search that uniformly scales the depth, width, and resolution of the feature maps with a fixed set of scaling coefficients.

The Efficient-Net models are composed of blocks called MBConv. They are a combination of the Inverted Bottleneck (IB) (Sandler, Howard, Zhu, Zhmoginov, & Chen, 2018) and Squeeze-and-Excitation (SAE) (Hu et al., 2018) blocks. IB blocks use depthwise convolutions (DWConv). They have less parameters to be optimized than the traditional convolution. And because of that, the computational cost is reduced. SAE blocks, in turn, generate weighted feature maps so that the most relevant ones gain greater weight to the detriment of the others in the learning process.

To compose the proposed FCN, Efficient-Net B3 is used due to the dimensions of the input slices (160×160) and their aspects. Since the input dimensions are not very large, the segmentation process does not require a very deep network. However, the slices have complex features, so it is necessary to use a robust network to recognize patterns in a more refined way. Therefore, among the architectures of the Efficient-Net family (B0 to B7), B3 was chosen as the most appropriate.

In the expansion path of the proposed FCN, five convolutional blocks called Decoder Blocks are used. They are composed of Inception blocks (Szegedy et al., 2015) coupled to SAE blocks that are followed by Upsampling 2×2. Along the expansion path, the number of generated feature maps is reduced by a half (from 512 to 32). Similar to the U-Net, the proposed FCN uses skip connections to concatenate extracted features from the contraction path with those generated in the expansion. The Decoder Blocks also use the Attention mechanism. Fig. 4 shows the inner structure of a Decoder Block as well as the mentioned concatenation process. So, let g be the output of a previous block and x be the skip connection. Both are passed to the Attention block and g is also submitted to Upsampling (scale factor = 2). Finally, these block outputs are concatenated and passed as input to the Inception block.

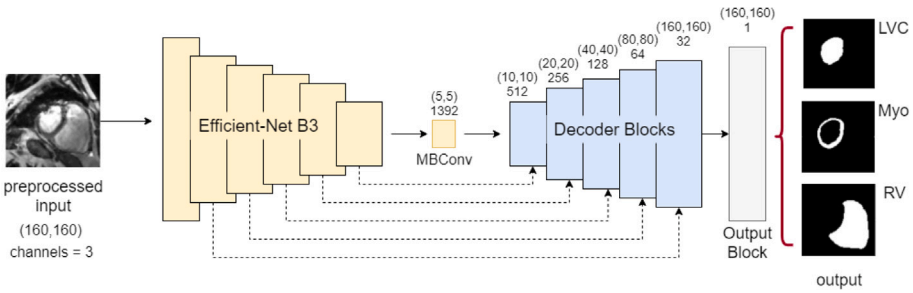

Fig. 3. The FCN proposed for the segmentation of cardiac structures.

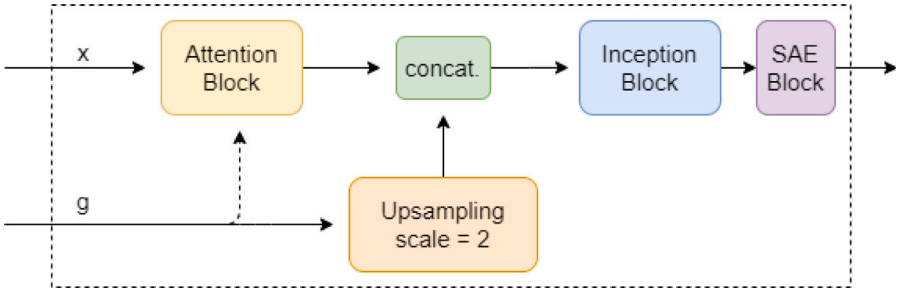

Fig. 4. Decoder block's internal structure.

The use of the Attention blocks was inspired by Oktay et al. (2018). These blocks allow to give weights to the internal regions of the feature maps transmitted by the skip connections. Thus, the most relevant regions of the maps will have more weight, while those of less relevance to the learning process of the network will receive lower weights. In relation to the Inception blocks, its use is aimed at learning features at various scales. With this, it is intended to identify the features of the smaller regions that belong to the structures of interest, as well as larger ones, and also to reduce false positives in the case of regions with similar spatial features. Moreover, the SAE blocks are used to improve the feature extraction process as mentioned before.

At the end of the expansion path, there is a 1×1 convolution layer followed by the Sigmoid activation, which therefore generates segmentation masks for the objects of interest.

It is important to note that, in the initial segmentation step, the proposed FCN is applied three times over the slice so that the LVC, RV, and Myo masks are generated separately. This strategy is based on the assumption that specialized training by structure of interest will produce better results compared to a single model that generates several different segmentation masks.

3.3. Refinement

The third step of the proposed method is called Refinement which comprises the using of techniques to improve the initial segmentation produced by the proposed FCN in the previous step.

As LVC, Myo, and RV have different aspects, mainly in relation to shape and size, the search for a unique set of techniques capable of improving the three segmentations becomes an exhausting task. Therefore, specific techniques were applied to each of these structures, in order to improve them individually.

The techniques used in the refinement step are grouped into two modules: reconstruction and postprocessing. The initial Myo and RV segmentations are submitted first to reconstruction and then to post-processing. The initial CV segmentation, in turn, is directed only to the latter. The explanation of how each module is performed will be presented below.

3.3.1. Reconstruction module

The reconstruction module was developed in order to improve the initial masks of Myo and RV. When analyzing the volume, it is observed that these masks have different shapes and sizes, being large and well defined in the medial slices and more irregular in the basal and apical slices.

Thus, the use of traditional techniques in image processing becomes more complex since many of them use fixed thresholds and many parameters whose estimation can be time-consuming and exhaustive. In addition, it is possible to obtain good results for certain slices and harm the segmentation of others at the same time.

Considering this scenario, the reconstruction module uses a deep convolutional network to produce improvements to the initial segmentations. Inspired by Souza et al. (2019), the idea is that the network learns the aspects of the ground-truth to correct the flaws presented by the initial segmentation. Thus, the cases of error can be improved without harming the cases of success. The reconstruction process can be seen in Fig. 5.

The reconstruction network receives an input with dimensions 160×160 and two channels, represented by the slice and its initial segmentation. The output will be the new reconstructed segmentation map. The architecture used in this process is the same U-Net used for the ROI extraction step (Section 3.1). For the training of this network, it was necessary to build a dataset composed of the slices, their ground-truth and their corresponding masks initially predicted by the proposed FCN. For this, the training, validation, and test sets were submitted to the initial segmentation step to generate the masks that serve as an input to the reconstruction network.

Therefore, specific models were trained for each structure (Myo and RV) so that there was no interference between the features, also reducing the complexity of the learning process. Thus, each network will focus on reconstructing its respective structure of interest.

3.3.2. Postprocessing

The postprocessing module consists of a set of techniques used specifically to improve the initial LVC segmentations generated by the proposed FCN and also those produced by the reconstruction module in the case of the Myo and RV structures.

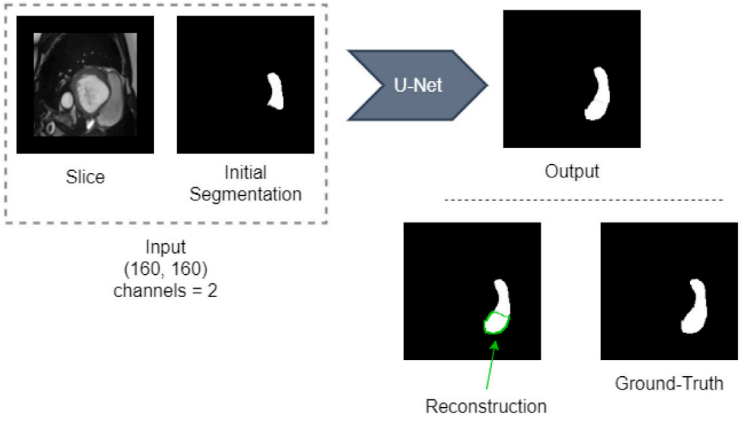

Fig. 5. An overview of the reconstruction process.

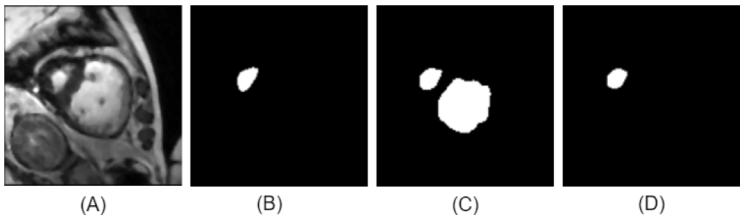

Fig. 6. Refinement step: postprocessing applied to the LVC initial segmentation.

Regarding LVC, it was verified in the tests that the initial segmentation step obtained consistent results for this structure, so that few techniques were sufficient for its improvement. The major failures detected were the prediction of false positives in regions of the slices with a similar texture to the LVC, as well as some gaps. Then, to correct these problems, the following techniques are used: (1) area filter and (2) closing. The area filter is useful for removing false positive regions. In general, these regions have an area larger than the LVC, and are more frequently predicted in apical slices. Thus, based on the prior knowledge that LVC is the smallest area, the area filter is used to maintain it. After that, closing is applied to fill possible internal gaps. This operation uses an elliptical structuring element with dimensions 15×15.

Fig. 6 shows the postprocessing applied to the LVC initial segmentation in the refinement step. In 6(A) and (B), slice and ground-truth are shown respectively. In 6(C), two predicted regions can be seen: the smallest is the LVC and the largest contains false positives. And, in 6(D), the result of the postprocessing with the elimination of the wrongly predicted region is shown.

In relation to the Myo and RV structures, their initial segmentations are first submitted to the reconstruction module in order to be improved. This process is important for reducing errors. However, even after reconstruction, some failures may persist. Therefore, postprocessing becomes necessary to correct them.

When analyzing the Myo segmentations generated by the reconstruction module, it was observed that the major remaining failures were elongations and discontinuities (Fig. 7). Elongations are groups of false positives connected to the generated mask that occur due to the similarity of texture between these pixels and those belonging to Myo. And discontinuities are gaps caused by predicted false negatives, which disconnect correctly predicted regions.

Therefore, the postprocessing applied to correct these failures consists of performing the following procedures: (1) elongation removal, (2) area filter, and (3) correction of discontinuities in the polar space.

Elongation removal comprises calculating horizontal and vertical projection histograms of the binary masks to remove lines or columns that contain less than five pixels with value 1. This threshold was

Fig. 7. An example of elongations and discontinuities.

determined by experiments. After that, the area filter is used to remove small objects and keep the largest one. This process is demonstrated in Fig. 8.

Finally, the output of elongation removal is passed to the process of correction of discontinuities in the polar space. The transformation to polar space is a step inspired by Tan et al. (2017). However, the authors use this technique as a previous processing to segmentation. On the other hand, in the proposed method, it is used as postprocessing. This technique consists of, given a center and a radius, mapping each pixel of the mask, represented on the Cartesian space, to a new image in the polar space. The center is obtained using the Hough transform and the radius is determined as half the width of the image ($160/2 = 80$ pixels). After obtaining the polar image, the horizontal histogram projection is calculated to find which lines do not have a pixel of value 1, which indicates a discontinuity. The lines before and after the discontinuity are also located. Then, the extreme points of these lines are found to thus establish the connection between them. This process can be seen in Fig. 9. At the end, the image is transformed from the polar to the Cartesian space, ending the postprocessing of the Myo segmentation.

In the case of RV segmentation, the observed problems that persisted after reconstruction are less complex compared to Myo masks. Thus, postprocessing aims to eliminate small regions of false positives

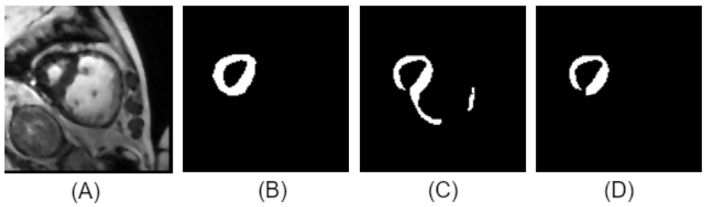

Fig. 8. Elongation removal: a sample of slice (A), its ground-truth (B), mask before (C) and after (D) the elongation removal process.

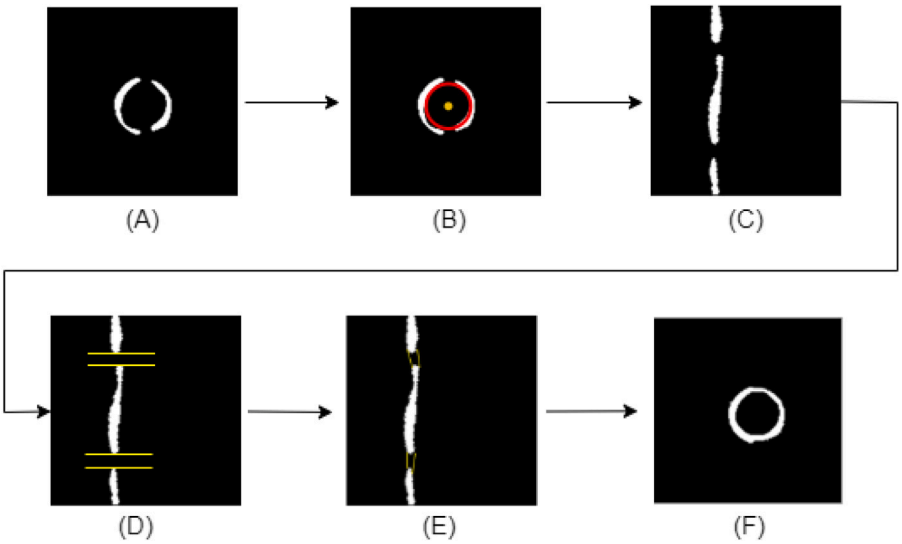

Fig. 9. Correction of discontinuities: input mask (A); circle detection (B); polar space mask (C); detection of discontinuities (D); connecting points (E); and the final result (F).

Fig. 10. Postprocessing of RV masks: a sample of slice (A), its ground-truth (B), mask before (C), and after (D) the postprocessing.

and gaps using (1) area filter and (2) closing. The first retrieves the largest object (RV) and the second is used to fill possible gaps, as shown in Fig. 10.

At the end of the refinement step, after the reconstruction and postprocessing modules are applied, the Myo, LVC, and RV masks are joined to produce the final segmentation map, which is the output generated by the proposed method.

4. Experiments

This section presents the dataset used in experiments and the results obtained by the proposed method for the segmentation of cardiac structures in short-axis cine-MRI. The evaluation metrics are also presented, as well as the details of the experiments and the discussion of the results obtained through the exposition of case studies and comparison with related works.

4.1. Dataset

The dataset used to validate the proposed method is provided by the Automated Cardiac Disease Diagnosis Challenge (ACDC) (Bernard et al., 2018). This dataset comprises short-axis cine-MRI of 150 patients divided into five groups evenly distributed: (1) healthy patients; (2) patients with previous myocardial infarction; (3) dilated cardiomyopathy; (4) hypertrophic cardiomyopathy; and (5) abnormal right ventricle.

The short-axis cine-MRI volumes were captured in breath hold partially or completely covering the cardiac cycle. Other acquisition parameters are: slice thickness of 5 to 8 mm, inter-slice gap of 5 to 8 mm, and spatial resolution from 1.37 to 1.68 mm^2/pixel.

The ACDC dataset is split into training and testing sets. The training set consists of exams of 100 patients along with the manual annotations (ground-truth) made by specialists during the ED and ES phases. The ground-truth delimits the right ventricle (RV), myocardium (Myo), and

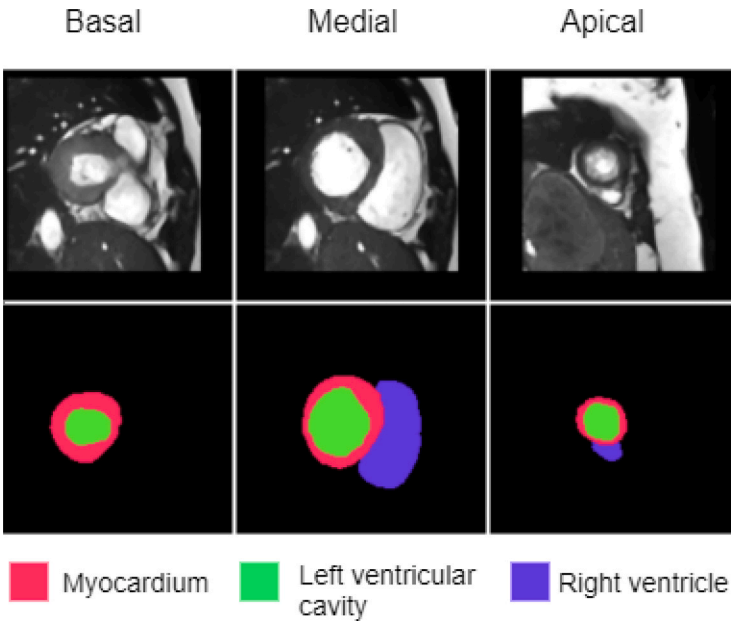

Fig. 11. Examples of the ACDC dataset with their respective annotations.

left ventricular cavity (LVC) in all the basal, medial, and apical slices. The test set contains exams of 50 patients without ground-truth. Thus, to validate a method, it is necessary to submit the generated masks to the challenge's online platform. Some examples of slices and their respective annotations can be seen in Fig. 11.

4.2. Results

The proposed method was developed in Python 3.6[1] using the machine learning frameworks Keras (Chollet et al., 2015) and Tensorflow.[2] The experiments were performed on a computer with Intel Core i5-7300HQ CPU, 8 GB RAM, NVIDIA GeForce GTX 1050 GPU, and Windows 10 system.

The metrics used for evaluating the similarity between the segmentation masks generated by the proposed method and the ground-truth are: Dice Coefficient, Jaccard Index, also called Intersection over Union (IoU), Sensitivity (SEN), and Precision (PRC). These metrics are widely used and accepted by the scientific community for evaluating medical image segmentation methods (Bland, 2015).

As previously mentioned, the ACDC dataset contains exams and their ground-truth for 100 patients. For the experiments, this dataset was randomly split into training, validation, and testing sets. Training set has 60 patients. Validation and testing have 20 patients each. Thus, a larger sample is provided for the training of network models, favoring generalization. It is also important to note that these sets did not change during the experiments. More details on the experiments and results achieved are presented in the following subsections.

4.2.1. ROI extraction

As aforementioned, the ROI extraction process is based on the Myo segmentation. In the experiments related to this step, two U-Nets (called 1 and 2) were tested. U-Net 1 is used to segment Myo only in the medial slices while U-Net 2 aims to segment that structure in all slices throughout the exam. Thus, it is possible to investigate through the comparison between these two approaches if the specialized training is necessary to generate the reference segments to be used in the ROI extraction process.

[1] https://www.python.org/.
[2] https://www.tensorflow.org.

Table 2
Comparison between approaches used in the segmentation process as part of the ROI extraction step.

	Dice	IoU
U-Net 1	**0.9368**	**0.8820**
U-Net 2	0.9041	0.8261

Table 3
MAE results for the ROI extraction process.

	MAE	STD	Max	Min
U-Net 1	**2.93**	**4.27**	**18.25**	**0.00**
U-Net 2	6.93	6.18	24.00	0.75

For the training of U-Net 1, only the medial slices of each exam were selected. However, this process significantly reduces the size of the training sample (only 120 slices). Therefore, Data Augmentation was used to increase the training set. So, the following operations were applied: rotation $[-5°, 5°]$, vertical and horizontal translations and horizontal flip. Thus, a total of 1080 slices were generated.

U-Net 1 was trained for 300 epochs using the Jaccard loss function and Adam optimizer (Kingma & Ba, 2015) with learning rate of 10^{-3}, decay factor of 0.1, and Early Stopping strategy (Prechelt, 1998). In relation to U-Net 2, this network was trained using the same hyperparameters as U-Net 1. However, the training set used in this case is composed of all types of slices, totaling 1154 slices. Both U-Net 1 and U-Net 2 have a total of 18,634,467 trainable parameters.

For evaluating, only medial slices of the test set were used, totaling 100 cases. Table 2 presents the results obtained for Myo segmentation in medial slices as part of the ROI extraction process.

U-Net 1, which represents specialized training with medial slices, generated the best segmentation results for the test set, obtaining Dice of 0.9368 and an IoU of 0.8820. U-Net 2, on the other hand, presented less favorable results due to having been trained with a more diversified set of slices.

Table 3 presents another analysis that was performed using the Mean Absolute Error (MAE) as a metric to measure the similarity between the ground-truth bounding boxes, which simulate the ROI extracted manually, and those generated automatically from the reference segmentations produced by the U-Nets 1 and 2.

Table 4
Cardiac structures segmentation: left ventricular cavity (LVC), myocardium (Myo), and right ventricle (RV). Comparison between the results obtained by the proposed FCN and other segmentation methods.

Test	Methods		Dice	IoU	SEN	PRC
1	U-Net		0.8903	0.8400	0.8867	0.9129
2	FCN (Efficient-Net B3+ std. convolutions)	**LVC**	0.9209	0.8807	0.9318	0.9166
3	**Proposed FCN**		**0.9283**	**0.8866**	**0.9374**	**0.9299**
1	U-Net		0.8016	0.7002	0.8141	0.8074
2	FCN (Efficient-Net B3+ std. convolutions)	**Myo**	0.8443	0.7608	0.8344	0.8711
3	**Proposed FCN**		**0.8474**	**0.7672**	**0.8355**	**0.8657**
1	U-Net		0.7025	0.6249	0.7905	0.6695
2	FCN (Efficient-Net B3+ std. convolution)	**RV**	0.7779	0.7175	0.8343	0.7586
3	**Proposed FCN**		**0.8306**	**0.7730**	**0.8515**	**0.8256**

It is observed that the ROI extraction approach based on the Myo segmentation in medial slices (U-Net 1) is capable of generating ROIs more similar to manual extraction since the average MAE is closer to zero. The results also show a low standard deviation, indicating consistency. There are cases in which the MAE is zero, i.e, the ROIs extracted automatically via U-Net 1 are identical to those extracted based on ground-truth.

Therefore, it is possible to conclude that the U-Net 1 trained in a specialized way proved to be the best approach for the ROI extraction process. This network produced the best segmentation results from which ROIs could be extracted similarly to those obtained manually. These results also indicate that all the structures of interest (LVC, Myo, and RV) will be preserved to be submitted to the next step of the proposed method.

4.2.2. Initial segmentation

For the training process of the proposed FCN, it was used a version of the training dataset containing only the ROIs extracted using ground-truth as reference. Data augmentation was performed for increasing the training set in order to avoid overfitting and favor generalization. The applied operations are the same as those used in the experiment of the ROI extraction. As a result, 10 386 slices were generated for the training set.

As mentioned in Section 3.2, three proposed FCN models were trained for producing the Myo, LVC, and RV segmentations separately. Each network was trained for 300 epochs using Jaccard loss and Adam optimizer with learning rate of 10^{-3}. Also, a decay factor of 0.1 was used as a way to avoid local minima. The number of trainable parameters for each proposed FCN model is 53,138,869.

Two sets were used for the testing phase: one is composed of ROIs manually extracted and the other contains the output from the ROI extraction step, simulating the proposed method's pipeline. Both are composed of the same patients, totaling 386 slices.

Table 4 shows the results achieved by the proposed FCN and also a comparative analysis between this and two other approaches: the traditional U-Net (Ronneberger et al., 2015) and a version of the proposed FCN with the expansion path based on standard convolutions (without the proposed decoder blocks). U-Net serves as baseline for the comparison due to its wide use in medical image segmentation. Besides, the proposed method uses this network to extract ROIs, which motivates its testing also in the segmentation step. In turn, the experiment involving the proposed FCN with standard convolutions is intended to validate the isolated use of Efficient-Net B3 as an improvement for the segmentation process.

These results are related to the test set with ROIs extracted automatically by the first step of the proposed method. Test 2 shows that the use of Efficient-Net B3 for feature extraction in the contraction path improved the segmentation results in comparison to the U-Net (test 1). This improvement was confirmed for the three structures, and for RV masks, the values of Dice and IoU increased significantly by approximately 0.07 and 0.09, respectively. In test 3 (with the proposed FCN), this increase is approximately 0.12 and 0.14, surpassing the other approaches.

Table 5
Experiments with the proposed FCN in two test sets: (C1) with the ROIs extracted manually and (C2) ROIs extracted by Step 1 of the proposed method.

Methods		Dice	IoU	SEN	PRC
C1	**LVC**	0.9216	0.8810	0.9276	0.9215
C2		**0.9283**	**0.8866**	**0.9374**	**0.9299**
C1	**Myo**	0.8470	0.7648	0.8378	0.8643
C2		**0.8474**	**0.7672**	**0.8355**	**0.8657**
C1	**RV**	0.8324	0.7764	0.8477	0.8347
C2		**0.8306**	**0.7730**	**0.8515**	**0.8256**

In general, the proposed FCN presents the best results. It is possible to observe a considerable increase in the values obtained for the segmentation of the three structures. For the LVC, the initial segmentation via the proposed FCN obtains 0.9283 of Dice, 0.8866 of IoU, 0.9374 of Sensitivity, and 0.9299 of Precision; representing an average increase of approximately 3.8% compared to U-Net.

For the Myo segmentation, the proposed FCN obtains 0.8474 of Dice, 0.7672 of IoU, 0.8355 of Sensitivity, and 0.8657 of Precision. This shows an average increase of 4.8%. And for the RV, the proposed FCN obtained 0.8306 of Dice, 0.7730 of IoU, 0.8515 of Sensitivity, and 0.8256 of Precision (average increase of 12.8% in relation to the U-Net).

Thus, it is observed that the proposed combination of the Efficient-Net B3 with Inception, Attention, and SAE blocks is more promising in relation to the standard convolution blocks used by the other approaches on the expansion path (tests 1 and 2).

As previously mentioned, in order to verify the impact of the ROI extraction process, another experiment was carried out using the test set containing the same patients, but with the ROI extracted manually. These results can be seen in Table 5. In general, it is possible to infer that the impact of automatic extraction on the segmentation process is minimal, since the proposed FCN achieves very similar results, with a small statistical difference, which shows a good interoperability between steps 1 and 2 of the method proposed.

At last, a comparative analysis was performed considering three scenarios: (A) the proposed method with steps 1 and 2, (B) the proposed FCN, and (C) U-Net, the last two are used to segment the slices without the previous step of ROI extraction. This comparison can be seen in Table 6.

In scenarios B and C, the dimensions of the slices were reduced to 160 × 160 due to hardware limitations. The obtained results are considerably inferior in relation to the proposed method (Scenario A), because the other networks process the entire slice and, therefore, are subject to deal with a high pixel class imbalance. This problem is increased by reducing the dimensions of the slices because the size of the structures of interest is considerably reduced, especially in the apical and basal regions of the exam, where these structures are naturally small. Thus, it is possible to infer that the scope reduction provided by Step 1 of the proposed method helps to generate better segmentation results because it reduces the imbalance and preserves the size of the structures of interest.

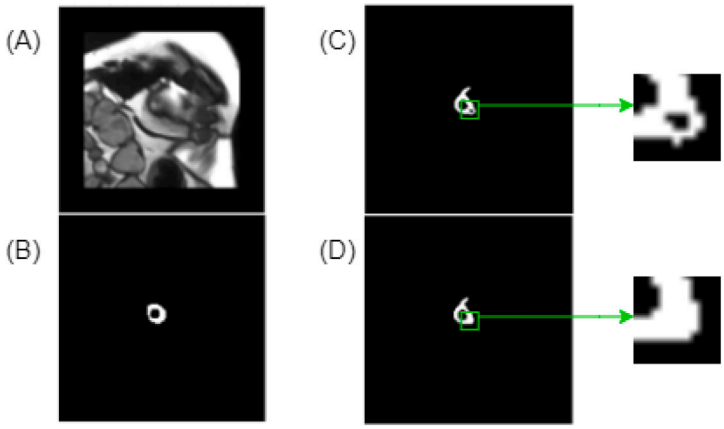

Fig. 12. Segmentation improvement produced by the reconstruction module: (A) input slice, (B) ground-truth, (C) initial segmentation with a Dice of 0.7368, and (D) segmentation after the reconstruction module, with a Dice of 0.7528.

Table 6
Comparative analysis between the proposed method (A) that divides segmentation into steps; the proposed FCN (B); and U-Net (C). For B and C, the ROI extraction step was discarded.

Methods		Dice	IoU	SEN	PRC
A		**0.9283**	**0.8866**	**0.9374**	**0.9299**
B	**LVC**	0.8540	0.7784	0.7915	0.9476
C		0.8592	0.7837	0.7970	0.9532
A		**0.8474**	**0.7672**	**0.8355**	**0.8657**
B	**Myo**	0.4399	0.3177	0.3278	0.8943
C		0.4457	0.3213	0.3277	0.9097
A		**0.8306**	**0.7730**	**0.8515**	**0.8256**
B	**RV**	0.6309	0.5353	0.5607	0.7864
C		0.5975	0.4995	0.5201	0.7626

4.2.3. *Refinement: Reconstruction and postprocessing*

The third step of the proposed method is responsible for the refinement of the initial segmentations using two modules: reconstruction and postprocessing. Therefore, this section presents the final results obtained by the proposed method after the application of these techniques.

As mentioned in Section 3.3, the inputs of the reconstruction module are the ROI and its initial segmentations. This module uses an U-Net to generate as output an improved mask. Inputs and outputs have dimensions 160 × 160. The reconstruction network was trained with the same hyperparameters defined for the previous training sessions, e.g., the number of epochs, optimizer, and the learning rate. This U-Net architecture is similar to the one used in the ROI extraction process, so the number of trainable parameters is the same. Only the loss function was changed to the Binary Crossentropy (BCE) + Jaccard Loss. The incorporation of the BCE aims to reinforce the penalty for the prediction of false positives and false negatives (Jadon, 2020). The results of the proposed reconstruction module are presented in Table 7. This table also shows a comparison with two other approaches based on that proposed by Souza et al. (2019).

The approach proposed by Souza et al. (2019) (Experiment 1) uses an FCN based on ResNet (He, Zhang, Ren, & Sun, 2016) that has a fully connected layer with number of neurons equal to the input dimensions. The loss function used consists of two BCE terms: one is used to analyze the prediction of background pixels and the other for analyzing the prediction of true positives. Due to hardware limitations, the input slices were resized to 96 × 96 in the experiment with this approach. However, it did not obtain satisfactory results for the reconstruction of cardiac structures, since its performance was inferior than that achieved by the initial segmentation (without reconstruction). In Experiment 2, U-Net was trained with the same loss function used in Experiment 1. This network showed better results compared to the previous experiment, but they were still unsatisfactory.

In Experiment 3, in which U-Net was trained with the combined loss function, the best overall result was obtained for the reconstruction of cardiac structures. Improvements can be noticed for the Myo and RV masks. Regarding the reconstruction of the LVC masks, all metrics decreased, which led to the discard of the use of the reconstruction module for that structure. For Myo, all metrics increases. And for RV, the Dice and IoU values had a tiny decrease of 0.0002. On the other hand, the sensitivity increased, indicating the predisposition of the model to classify pixels as being part of the region of interest. Therefore, from this analysis, it was decided to use the reconstruction module only for the Myo and RV structures. Examples of the improvement provided by the reconstruction module can be seen in Fig. 12.

At last, the postprocessing module is evaluated. Table 8 shows the general result obtained by the complete pipeline of the proposed method with the test set going through all three steps. And Table 9 shows the obtained results grouped by the end of diastole (ED) and the end of systole (ES) cardiac phases.

The tables show the results obtained with the execution of steps 1 and 2 of the proposed method in comparison with its complete pipeline. It is noticed that, in the general evaluation, the use of the refinement module ensured the generation of an improved final segmentation. This is confirmed by the significant increase that can be observed for the evaluation metrics.

In the analysis of slices grouped by cardiac phases, it is observed that the method achieves consistent results for the ED group, and presents an improvement in the ES segmentations. This is because, in this phase, the structures are more retracted and less evident, which causes the main failures corrected by the refinement step. Thus, the improvement observed quantitatively indicates the important contribution of this step to the proposed method.

The qualitative results of the final segmentation for the three types of slices can be seen in Fig. 13. In this figure, it is possible to observe the improvement of the initial segmentation provided by the refinement step, with emphasis on the correction of discontinuities shown in the medial and basal cases. The apical example, in turn, shows good results despite a small growth in the region predicted for the RV. In this case, there is a similarity between the texture of this region with other adjacent ones, since they are closer because of the systole movement.

In the basal example, it is noticed that the proposed method considers both textural and shape aspects. Because, although there are in those slices a structure with texture and location similar to the RV, but with a different shape, this region were correctly predicted as true negative thus corroborating the generalization ability.

Finally, regarding computational complexity, the average execution time of the proposed method is 4 s per patient. Among the three steps,

Table 7
Experiments carried out with different approaches for the reconstruction module.

Exp.	Methods		Dice	IoU	SEN	PRC
1	Souza et al. (2019)		0.8222	0.7477	0.7734	0.9025
2	U-Net + BCE	**LVC**	0.9057	0.8473	0.8640	0.9647
3	**U-Net + BCE + Jaccard loss**		**0.9230**	**0.8812**	**0.9328**	**0.9238**
1	Souza et al. (2019)		0.4216	0.2813	0.3195	0.6410
2	U-Net + BCE	**Myo**	0.8145	0.7144	0.7328	0.9271
3	**U-Net + BCE + Jaccard loss**		**0.8528**	**0.7727**	**0.8420**	**0.8700**
1	Souza et al. (2019)		0.4388	0.3192	0.3364	0.6943
2	U-Net + BCE	**RV**	0.7962	0.7164	0.7486	0.8706
3	**U-Net + BCE + Jaccard loss**		**0.8304**	**0.7728**	**0.8523**	**0.8248**

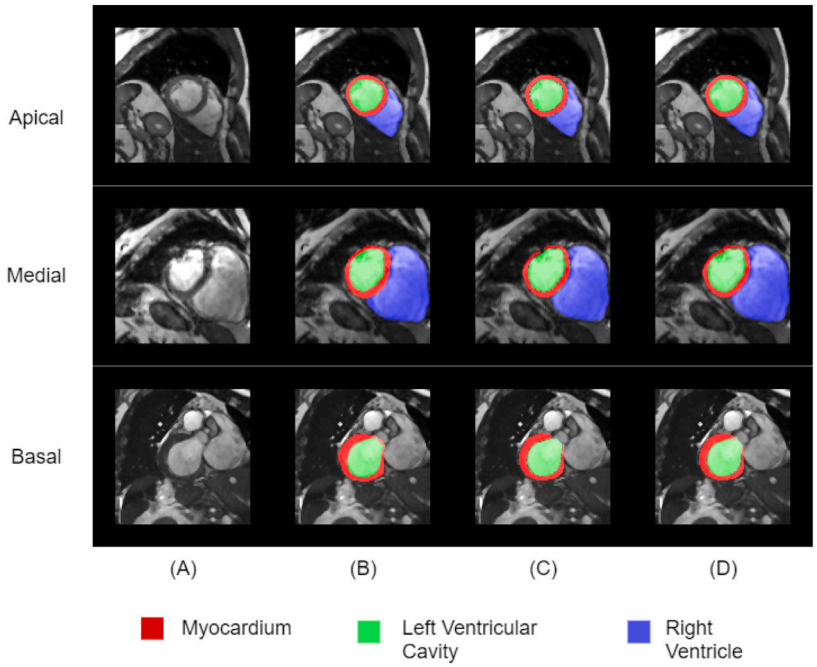

Fig. 13. Qualitative results of the proposed method: (A) the input, (B) the ground-truth, (C) the initial segmentation, and (D) the final segmentation.

Table 8
Results obtained by the proposed method in two scenarios: execution of Steps 1 and 2, without refinement (Step 3), and the complete pipeline.

Proposed method		Dice	IoU	SEN	PRC
Steps 1 and 2	**LVC**	0.9283	0.8866	0.9374	0.9299
Steps 1, 2, and 3		**0.9303**	**0.8886**	**0.9416**	**0.9285**
Steps 1 and 2	**Myo**	0.8474	0.7672	0.8355	0.8657
Steps 1, 2, and 3		**0.8552**	**0.7747**	**0.8449**	**0.8730**
Steps 1 and 2	**RV**	0.8306	0.7730	0.8515	0.8256
Steps 1, 2, and 3		**0.8312**	**0.7740**	**0.8512**	**0.8269**

Table 9
Evaluation of the proposed method from the Dice and IoU values calculated for the slices in the ED and ES cardiac phases.

Proposed method			Dice	IoU
Steps 1 and 2	**LVC**	ED	0.9602	0.9248
Steps 1, 2, and 3		**ED**	**0.9602**	**0.9248**
Steps 1 and 2		ES	0.9246	0.8708
Steps 1, 2, and 3		**ES**	**0.9337**	**0.8795**
Steps 1 and 2	**Myo**	ED	0.8870	0.7986
Steps 1, 2, and 3		**ED**	**0.8871**	**0.7986**
Steps 1 and 2		ES	0.8875	0.8006
Steps 1, 2, and 3		**ES**	**0.8902**	**0.8041**
Steps 1 and 2	**RV**	ED	0.9262	0.8653
Steps 1, 2, and 3		**ED**	**0.9257**	**0.8646**
Steps 1 and 2		ES	0.8517	0.7477
Steps 1, 2, and 3		**ES**	**0.8543**	**0.7520**

the initial segmentation is the slowest, with an average execution time of 2 s. This is because it is necessary to run three proposed FCNs for the specific segmentation of each structure of interest. Despite the different techniques used in each step, it is observed that the proposed method achieves promising results in a time considered satisfactory.

4.3. Performance evaluation on ACDC challenge: Comparison with related works

The ACDC challenge provides a test dataset containing exams from 50 patients (without ground-truth) exclusively for performance evaluation. Dice coefficient and the Hausdorff distance (HD) are used for measuring segmentation accuracy of the cardiac structures in ED and ES phases. And for a clinical evaluation, correlation, bias, and standard deviation values are calculated from the following measurements: the ED volumes of LVC and RV; the ejection fractions of LVC and RV; and the Myo mass in ED. Tables 10, 11, and 12 show the results obtained by the proposed method in experiments with the ACDC test dataset as well as the comparison with other related works.

The proposed method presented a promising performance obtaining good scores for the LVC, Myo, and RV segmentation. For the LVC, the proposed method obtains a Dice value in ED similar to that achieved

Table 10
Results in the segmentation of the Left Ventricular Cavity (LVC). Comparison of the proposed method with other competing approaches on the ACDC test dataset.

Method	Dice ED	Dice ES	HD ED	HD ES	EF corr.	EF bias	Vol. ED corr.	Vol. ED bias
Proposed method	**0.963**	**0.912**	**8.062**	**10.432**	**0.975**	**1.030**	**0.994**	**0.110**
Simantiris and Tziritas (2020)	0.967	0.928	6.366	7.573	0.993	−0.360	0.998	2.032
Isensee et al. (2017)	0.967	0.928	5.476	6.921	0.991	0.490	0.997	1.530
Zotti, Luo, Lalande, and Jodoin (2018)	0.964	0.912	6.180	8.386	0.990	−0.476	0.997	3.746
Painchaud et al. (2019)	0.961	0.911	6.152	8.278	0.990	−0.480	0.997	3.824
Khened et al. (2019)	0.964	0.917	8.129	8.968	0.989	−0.548	0.997	0.576
Baumgartner et al. (2017)	0.963	0.911	6.526	9.170	0.988	0.568	0.995	1.436
Calisto and Lai-Yuen (2020)	0.958	0.903	5.592	8.644	0.981	0.494	0.997	3.072
Wolterink, Leiner, Viergever, and Išgum (2017)	0.961	0.918	7.515	9.603	0.988	−0.494	0.993	3.046
Rohé, Sermesant, and Pennec (2017)	0.957	0.900	7.483	10.747	0.989	−0.094	0.993	4.182

Table 11
Results in the segmentation of the Myocardium (Myo). Comparison of the proposed method with other competing approaches on the ACDC test dataset.

Method	Dice ED	Dice ES	HD ED	HD ES	Vol. ES corr.	Vol. ES bias	Mass ED corr.	Mass ED bias
Proposed method	**0.894**	**0.905**	**7.906**	**9.912**	**0.980**	**−1.100**	**0.988**	**−1.820**
Isensee et al. (2017)	0.904	0.923	7.014	7.328	0.988	−1.984	0.987	−2.547
Simantiris and Tziritas (2020)	0.891	0.904	8.264	9.575	0.983	−2.134	0.992	−2.904
Calisto and Lai-Yuen (2020)	0.873	0.895	8.197	8.318	0.988	−1.79	0.989	−2.100
Zotti et al. (2018)	0.886	0.902	9.586	9.291	0.980	1.160	0.986	−1.827
Painchaud et al. (2019)	0.881	0.897	8.651	9.598	0.979	0.296	0.987	−2.906
Khened et al. (2019)	0.889	0.898	9.841	12.582	0.979	−2.572	0.990	−2.873
Patravali, Jain, and Chilamkurthy (2017)	0.882	0.897	9.757	11.256	0.986	−4.464	0.989	−11.586
Baumgartner et al. (2017)	0.892	0.901	8.703	10.637	0.983	−9.602	0.982	−6.861
Zotti et al. (2017)	0.884	0.896	8.708	9.264	0.960	−7.804	0.984	−12.405
Wolterink et al. (2017)	0.875	0.894	11.121	10.687	0.971	0.906	0.963	−0.960

Table 12
Results in the segmentation of the Righ Ventricle (RV). Comparison of the proposed method with other competing approaches on the ACDC test dataset.

Method	Dice ED	Dice ES	HD ED	HD ES	EF corr.	EF bias	Vol. ED corr.	Vol. ED bias
Proposed method	**0.900**	**0.860**	**14.660**	**17.560**	**0.743**	**1.810**	**0.931**	**7.370**
Isensee et al. (2017)	0.951	0.904	8.205	11.655	0.910	−3.750	0.992	0.900
Calisto and Lai-Yuen (2020)	0.936	0.884	10.183	12.234	0.899	−2.118	0.989	3.550
Simantiris and Tziritas (2020)	0.936	0.889	13.289	14.367	0.894	−1.292	0.990	0.906
Zotti et al. (2018)	0.934	0.885	11.052	12.650	0.869	−0.872	0.986	2.372
Zotti et al. (2017)	0.941	0.882	10.318	14.053	0.872	−2.228	0.991	−3.722
Painchaud et al. (2019)	0.933	0.884	13.718	13.323	0.865	−0.874	0.986	2.078
Khened et al. (2019)	0.935	0.879	13.994	13.930	0.858	−2.246	0.982	−2.896
Baumgartner et al. (2017)	0.932	0.883	12.670	14.691	0.851	1.218	0.977	−2.290
Wolterink et al. (2017)	0.928	0.872	11.879	13.399	0.852	−4.610	0.980	3.596
Rohé et al. (2017)	0.916	0.845	14.049	15.926	0.781	−0.662	0.983	7.340

by Baumgartner et al. (2017) surpassing this by 0.001 of Dice in ES phase. In Fig. 14 is shown a comparison between Dice coefficients obtained by the proposed method and by the other approaches for the LVC segmentation on the ACDC test dataset. However, the proposed method is surpassed by approaches that use techniques with high computational cost. Among these, the one that stands out is the method developed by Isensee et al. (2017) that is based on an ensemble of U-Nets 2D and 3D and achieves the best score in two out of three rankings.

In relation to the Myo segmentation, the proposed method obtains one of the three best Dice values in ED and ES phases (0.890 and 0.905, respectively), and surpasses the method of Simantiris and Tziritas (2020). The HD values obtained for Myo are also among the best results, meaning that the segmentations generated by the proposed method for the exams at the ED and ES phases are considerably similar to the ground-truth.

A comparison chart for Dice coefficients related to the Myo segmentation can be seen in Fig. 15. It should also be noted that, although the proposed method has been designed using few hardware resources, it overcomes more robust methods such as AdaEn-Net (Calisto & Lai-Yuen, 2020), Grid-Net (Zotti et al., 2018), and the densely connected residual network proposed by Khened et al. (2019). The first two are more complex because they process the entire slices of the exam and the last one, despite proposing a cascade segmentation approach, uses a more dense and deep network architecture.

At last, for the RV, the proposed method obtained less good results compared to LVC and Myo, but it is still competitive with those achieved by the other ranked approaches. In the comparison chart shown in Fig. 16, it can be observed that, excepts (Isensee et al., 2017), the most approaches obtain Dice values between 0.84 and 0.89 for the ES phase. It shows that the RV segmentation is a challenging task. In this case, the proposed method achieves 0.86 and surpasses that proposed by Rohé et al. (2017).

5. Discussion

The division of the segmentation process into three steps is promising, since the techniques developed for each step present a good interoperability to generate more accurate masks, solving some problems in the application domain.

As observed in the results, the first step helps to mitigate the class imbalance. In addition, it reduces the processing cost by extracting

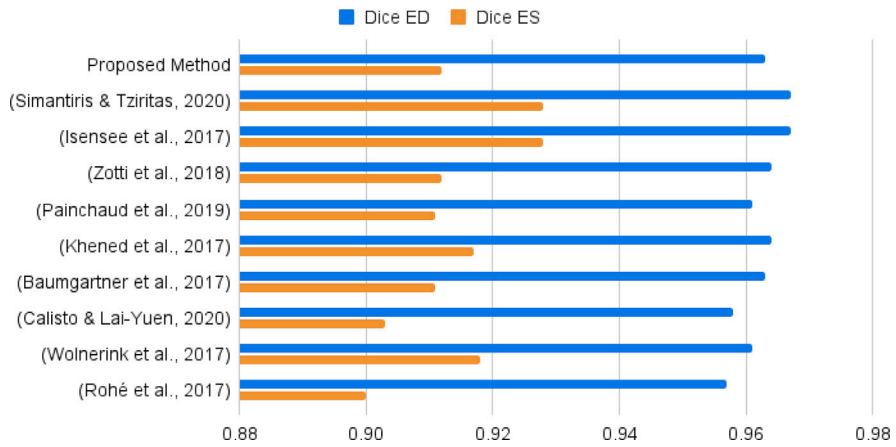

Fig. 14. Comparison chart of the Dice results in the LVC segmentation.

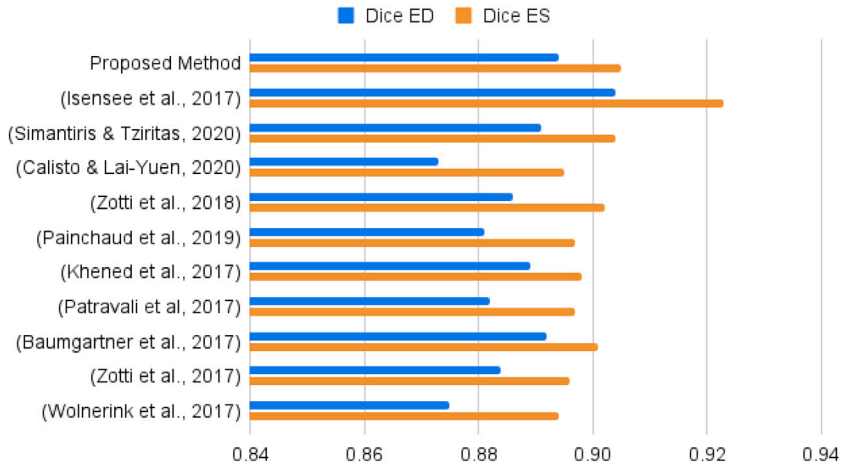

Fig. 15. Comparison chart of the results in the Myo segmentation.

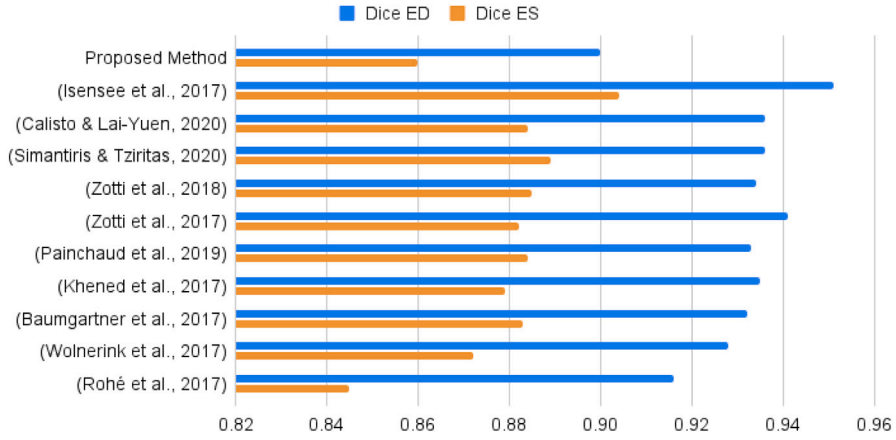

Fig. 16. Comparison chart of the results in the RV segmentation.

the ROI instead of resizing the input slices, which could change their aspects and hinder the learning process. The second step generates initial segmentations based on the ROI extracted using a proposed FCN that combines mechanisms that have been well evaluated in the literature. And the third step improves the initial segmentation using the reconstruction and post-processing modules that were developed specifically for each cardiac structure. It is also noteworthy that the reconstruction module is a contribution to the process of improving the masks since it avoids the search and testing of different post-processing techniques, which can be an exhausting and time-consuming process.

In general, the proposed method obtained good results for the segmentation of cardiac structures and this is confirmed through performance evaluation, with the ACDC test dataset, in which the achieved results are competitive and can be well ranked. Most of the related works proposed methods based on variations of U-Net or ensembles and perform the segmentation process in a single step. These methods were developed using more robust hardware resources such as the

NVIDIA GTX Titan X or V high-performance GPUs. On the other hand, the proposed method divides the segmentation process into smaller steps, each aimed at a specific part of the problem, which allowed its development on more limited hardware with an NVIDIA GTX 1050 GPU that has 2 GB on-board memory and 640 CUDA cores, about 80% less than those previously mentioned.

Regarding the segmentation metrics, it is observed that the right ventricle obtained less good results in comparison with the other structures. It is understood that this happens due to the variability of shapes that the RV can present throughout the volume. This structure is usually better defined in the medial slices, but, as observed in the tests, in the apical slices the RV can assume a very small size, so that the method misclassified pixels as background, generating false negatives. In the case of basal slices, there were instances when a cardiac structure similar to the RV was also generated as false positives. Nevertheless, the proposed method was able to produce good results for the RV, with a Dice of 0.9257 (ED) and 0.8543 (ES) obtained in the local test and 0.900 (ED) and 0.860 (ES) in the ACDC test dataset.

It is important to note that the results achieved for the LVC segmentation can be ranked among the top eight in the ACDC challenge, and, for the Myo, among the top five. These quantitative and qualitative analysis indicate that the proposed method produces competitive results and also show that its application is feasible in the real scenario.

6. Conclusion

This work presented a method proposed for the segmentation of cardiac structures in short-axis cine-MRI images. These structures are the left ventricular cavity (LVC), myocardium (Myo), and the right ventricle (RV). This segmentation process is an important task in the context of early diagnosis and treatment of cardiovascular diseases.

The proposed method is composed of three steps. In the first, an ROI is extracted from a reference segmentation generated by U-Net, in order to reduce processing and mitigate the pixel class imbalance problem. The second step consists of submitting the ROI to an FCN to produce an initial segmentation. The proposed architecture for this FCN uses Efficient-Net B3 in the contraction path and a combination between Inception, Attention, and Squeeze-and-Excitation blocks with skip connections in the expansion path. Finally, the initial segmentations are passed on to the third step called refinement. In this step, the reconstruction module based on U-Net is used to correct failures in RV and Myo masks. Moreover, all masks are passed to the postprocessing module in which image processing techniques are used to improve the segmentation results.

The experiments show that the proposed method obtains promising results. The division into steps showed a better performance in relation to segmentation process in a single stage. The proposed FCN used in the second step surpasses U-Net, which is one of the major deep learning approaches for medical image segmentation. And the refinement step, represented by reconstruction and postprocessing modules, demonstrates to be essential for improving segmentation. This can be confirmed by the performance evaluation in the ACDC test dataset, in which the proposed method achieves good results being competitive compared to the best approaches in the segmentation part of the challenge, with emphasis on the LVC and Myo segmentations.

As future works, it is intended to study other architectures, such as Generative Adversarial Networks (Goodfellow et al., 2014) and also other types of Attention blocks, such as the multi-scale (Sinha & Dolz, 2020), in order to apply them both for the initial segmentation and for the reconstruction module aiming to improve the overall results, mainly those related to the right ventricle, for which the proposed method presented more cases of failure.

CRediT authorship contribution statement

Italo Francyles Santos da Silva: Conceptualization, Methodology, Software, Writing – original draft, Investigation. **Aristófanes Corrêa Silva:** Conceptualization, Supervision, Validation. **Anselmo Cardoso de Paiva:** Conceptualization, Supervision. **Marcelo Gattass:** Conceptualization, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the Brazilian foundations Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for financial support.

References

Abdeltawab, H., Khalifa, F., Taher, F., Alghamdi, N. S., Ghazal, M., Beache, G., et al. (2020). A deep learning-based approach for automatic segmentation and quantification of the left ventricle from cardiac cine MR images. *Computerized Medical Imaging and Graphics*, *81*, Article 101717.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SEGNET: A deep convolutional encoder–decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*, 2481–2495.

Baumgartner, C. F., Koch, L. M., Pollefeys, M., & Konukoglu, E. (2017). An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation. In *International workshop on statistical atlases and computational models of the heart* (pp. 111–119). Springer.

Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.-A., et al. (2018). Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE Transactions on Medical Imaging*, *37*, 2514–2525.

Bland, M. (2015). *An introduction to medical statistics* (4th ed.). Oxford University Press (UK), ISBN: 978–0–19–958992–0.

Caelles, S., Maninis, K.-K., Pont-Tuset, J., Leal-Taixé, L., Cremers, D., & Van Gool, L. (2017). One-shot video object segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 221–230).

Calisto, M. B., & Lai-Yuen, S. K. (2020). Adaen-net: An ensemble of adaptive 2D–3D fully convolutional networks for medical image segmentation. *Neural Networks*, *126*, 76–94.

Chollet, F., et al. (2015). Keras. https://keras.io (Accessed: April 17, 2021).

Faridah Abdul Aziz, Y., Fadzli, F., Rizal Azman, R., Mohamed Sani, F., Vijayananthan, A., & Nazri, M. (2013). State of the heart: CMR in coronary artery disease. *Current Medical Imaging Reviews*, *9*, 201–213.

Gao, L., Zhang, L., Liu, C., & Wu, S. (2020). Handling imbalanced medical image data: A deep-learning-based one-class classification approach. *Artificial Intelligence in Medicine*, *108*, Article 101935.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. In *NIPS'14, Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2* (pp. 2672–2680). Cambridge, MA, USA: MIT Press.

Grinias, E., & Tziritas, G. (2017). Fast fully-automatic cardiac segmentation in MRI using MRF model optimization, substructures tracking and b-spline smoothing. In *International Workshop on Statistical Atlases and Computational Models of the Heart* (pp. 91–100). Springer.

Hazra, A., Mandal, S. K., Gupta, A., Mukherjee, A., & Mukherjee, A. (2017). Heart disease diagnosis and prediction using machine learning and data mining techniques: a review. *Advances in Computational Sciences and Technology*, *10*, 2137–2159.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Hu, H., Pan, N., Wang, J., Yin, T., & Ye, R. (2019). Automatic segmentation of left ventricle from cardiac MRI via deep learning and region constrained dynamic programming. *Neurocomputing*, *347*, 139–148.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132–7141).

Isensee, F., Jaeger, P. F., Full, P. M., Wolf, I., Engelhardt, S., & Maier-Hein, K. H. (2017). Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. In *International workshop on statistical atlases and computational models of the heart* (pp. 120–129). Springer.

Jadon, S. (2020). A survey of loss functions for semantic segmentation. In *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)* (pp. 1–7). IEEE.

Khened, M., Kollerathu, V. A., & Krishnamurthi, G. (2019). Fully convolutional multi-scale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. *Medical Image Analysis*, *51*, 21–45.

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of the 3rd international conference for learning representations (ICLR 2015)* (pp. 1–15).

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431–3440).

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *J. Fürnkranz, & T. Joachims (Eds.), Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 807–814).

Nasr-Esfahani, M., Mohrekesh, M., Akbari, M., Soroushmehr, S. R., Nasr-Esfahani, E., Karimi, N., et al. (2018). Left ventricle segmentation in cardiac MR images using fully convolutional network. In *2018 40th annual international conference of the ieee engineering in medicine and biology society (EMBC)* (pp. 1275–1278). IEEE.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M. C. H., Heinrich, M. P., Misawa, K., et al. (2018). Attention U-net: Learning where to look for the pancreas. In *Proceedings of the 1st conference on medical imaging with deep learning (MIDL 2018)* (pp. 1–10).

Painchaud, N., Skandarani, Y., Judge, T., Bernard, O., Lalande, A., & Jodoin, P.-M. (2019). Cardiac MRI segmentation with strong anatomical guarantees. In *International conference on medical image computing and computer-assisted intervention* (pp. 632–640). Springer.

Patravali, J., Jain, S., & Chilamkurthy, S. (2017). 2D-3D fully convolutional neural networks for cardiac MR segmentation. In *International workshop on statistical atlases and computational models of the heart* (pp. 130–139). Springer.

Prechelt, L. (1998). Early stopping-but when? In *Neural networks: tricks of the trade* (pp. 55–69). Springer.

Radau, P., Lu, Y., Connelly, K., Paul, G., Dick, A., & Wright, G. (2009). Evaluation framework for algorithms segmenting short axis cardiac. *MRI the MIDAS Journal - Cardiac MR Left Ventricle Segmentation Challenge*, *49*, 1–7, URL: http://hdl.handle.net/10380/3070.

Recht, B., Roelofs, R., Schmidt, L., & Shankar, V. (2019). Do ImageNet classifiers generalize to ImageNet? In *International conference on machine learning* (pp. 5389–5400). PMLR.

Rohé, M.-M., Sermesant, M., & Pennec, X. (2017). Automatic multi-atlas segmentation of myocardium with SVF-net. In *International workshop on statistical atlases and computational models of the heart* (pp. 170–177). Springer.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 234–241). Springer.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510–4520).

Santurkar, S., Tsipras, D., Ilyas, A., & Madry, A. (2018). How does batch normalization help optimization? In *Proceedings of the 32nd international conference on neural information processing systems* (pp. 2488–2498).

Sara, L., Szarf, G., Tachibana, A., Shiozaki, A. A., Villa, A. V., Oliveira, A. C. d., et al. (2014). II diretriz de ressonância magnética e tomografia computadorizada cardiovascular da sociedade brasileira de cardiologia e do colégio brasileiro de radiologia. *Arquivos Brasileiros de Cardiologia*, *103*, 1–86.

Sechtem, U., Pflugfelder, P. W., Gould, R. G., Cassidy, M., & Higgins, C. B. (1987). Measurement of right and left ventricular volumes in healthy individuals with cine MR imaging. *Radiology*, *163*, 697–702.

Simantiris, G., & Tziritas, G. (2020). Cardiac MRI segmentation with a dilated CNN incorporating domain-specific constraints. *IEEE Journal of Selected Topics in Signal Processing*, *14*, 1235–1243.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556v6.

Sinha, A., & Dolz, J. (2020). Multi-scale self-guided attention for medical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, *25*, 121–130.

Souza, J. C., Diniz, J. O. B., Ferreira, J. L., da Silva, G. L. F., Silva, A. C., & de Paiva, A. C. (2019). An automatic method for lung segmentation and reconstruction in chest x-ray using deep neural networks. *Computer Methods and Programs in Biomedicine*, *177*, 285–296.

Suinesiaputra, A., Cowan, B. R., Al-Agamy, A. O., Elattar, M. A., Ayache, N., Fahmy, A. S., et al. (2014). A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images. *Medical Image Analysis*, *18*, 50–62.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–9).

Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th international conference on machine learning (ICML 2019)*, (pp. 6105–6114).

Tan, L. K., Liew, Y. M., Lim, E., & McLaughlin, R. A. (2017). Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine MR sequences. *Medical Image Analysis*, *39*, 78–86.

Tran, P. V. (2016). A fully convolutional neural network for cardiac segmentation in short-axis MRI. arXiv preprint arXiv:1604.00494v3.

Wolterink, J. M., Leiner, T., Viergever, M. A., & Išgum, I. (2017). Automatic segmentation and disease classification using cardiac cine MR images. In *International workshop on statistical atlases and computational models of the heart* (pp. 101–110). Springer.

Zotti, C., Luo, Z., Humbert, O., Lalande, A., & Jodoin, P.-M. (2017). GridNet with automatic shape prior registration for automatic MRI cardiac segmentation. In *International workshop on statistical atlases and computational models of the heart* (pp. 73–81). Springer.

Zotti, C., Luo, Z., Lalande, A., & Jodoin, P.-M. (2018). Convolutional neural network with shape prior applied to cardiac MRI segmentation. *IEEE Journal of Biomedical and Health Informatics*, *23*, 1119–1128.