applied sciences

MDPI

Article

Region-of-Interest-Based Cardiac Image Segmentation with Deep Learning

Raul-Ronald Galea [1,*], Laura Diosan [1,2], Anca Andreica [1,2], Loredana Popa [1], Simona Manole [1] and Zoltán Bálint [1,3]

[1] IMOGEN Research Institute, County Clinical Emergency Hospital, Clinicilor, 1-3, Cluj-Napoca, 400008 Cluj, Romania; lauras@cs.ubbcluj.ro (L.D.); anca@cs.ubbcluj.ro (A.A.); paplory@yahoo.com (L.P.); simona.manole@gmail.com (S.M.); zoltan.balint@phys.ubbcluj.ro (Z.B.)
[2] Faculty of Mathematics and Computer Science, Babes-Bolyai University, Mihail Kogălniceanu 1, Cluj-Napoca, 400084 Cluj, Romania
[3] Faculty of Physics, Babes-Bolyai University, Mihail Kogălniceanu 1, Cluj-Napoca, 400084 Cluj, Romania
* Correspondence: raulronaldgalea@gmail.com

Abstract: Despite the promising results obtained by deep learning methods in the field of medical image segmentation, lack of sufficient data always hinders performance to a certain degree. In this work, we explore the feasibility of applying deep learning methods on a pilot dataset. We present a simple and practical approach to perform segmentation in a 2D, slice-by-slice manner, based on region of interest (ROI) localization, applying an optimized training regime to improve segmentation performance from regions of interest. We start from two popular segmentation networks, the preferred model for medical segmentation, U-Net, and a general-purpose model, DeepLabV3+. Furthermore, we show that ensembling of these two fundamentally different architectures brings constant benefits by testing our approach on two different datasets, the publicly available ACDC challenge, and the imATFIB dataset from our in-house conducted clinical study. Results on the imATFIB dataset show that the proposed approach performs well with the provided training volumes, achieving an average Dice Similarity Coefficient of the whole heart of 89.89% on the validation set. Moreover, our algorithm achieved a mean Dice value of 91.87% on the ACDC validation, being comparable to the second best-performing approach on the challenge. Our approach provides an opportunity to serve as a building block of a computer-aided diagnostic system in a clinical setting.

Keywords: cardiac image segmentation; deep learning; MRI image analysis; convolutional neuronal networks; artificial intelligence

Citation: Galea, R.; Diosan, L.; Andreica, A.; Popa, L.; Manole, S.; Bálint, Z. Region-of-Interest Based Cardiac Image Segmentation with Deep Learning. *Appl. Sci.* **2021**, *11*, 1965. https://doi.org/10.3390/app11041965

Academic Editor: Wei-Chang Yeh
Received: 12 January 2021
Accepted: 15 February 2021
Published: 23 February 2021

1. Introduction

Cardiac image segmentation is a laborious manual task which requires both working hours of the radiologist and accurate analysis. Therefore, the development of automatic methods for cardiac image analysis is of major importance [1]. This demonstrates the high potential of deep learning-based segmentation methods, for applications where manual or semiautomatic contouring is still a bottleneck in the workflow [2]. While deep learning methods have recently shown great promise for the task of cardiac image segmentation [1], one of their most significant drawbacks is their need for a large amount of data to provide good results. There are a number of publicly available medical image datasets [1], among which the most popular being Multi-Modality Whole Heart Segmentation—MMWHS [3] and the Automated Cardiac Diagnosis challenge—ACDC [4]. However, in most cases, the data are limited in numbers (i.e., couple of hundreds at most), since medical data acquisition, and cardiac Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) are especially difficult to perform on a large population [5].

Several methods for handling the scarceness of labeled images have been proposed [6,7], but they can address the problem of limited annotations only partially. For instance, other methodologies (CT-based) to apply semi- and fully automated image

analysis are proposed in [8,9], where the importance of epicardial fat segmentation on Computed Tomography are analyzed by taking into account either the classic image-processing methods or the deep learning methods. Furthermore, image acquisition methods play an important role in medical image segmentation. It has been shown in [1,10] that factors such as slice thickness and large inter-slice gaps can determine 2D approaches to work better than 3D ones.

In this paper, we study the feasibility of applying deep learning methods for the problem of cardiac image segmentation on a pilot dataset. We attempt to make the most out of the available data via a number of techniques, which include: data augmentation, transfer learning, region of interest (ROI) localization, and ensembling. The pipeline we apply processes volumes in an entirely 2D, slice-by-slice manner, obtaining promising results for a possible clinical pipeline application.

We start from two popular segmentation networks, the preferred model for medical segmentation, U-Net, and a general purpose model, DeepLabV3+. Novel modifications and integrations to U-Net (like those proposed in [11–13]) will be explored in the future. In [14] a deep learning-based segmentation method that automatically configures itself (including preprocessing, network architecture, training, and post-processing), called nnU-net, is proposed for a large range of segmentation tasks in the biomedical domain. Although nnU-Net has been shown to find high-quality configurations on new datasets robustly, task-specific empirical optimization may have the potential to further enhance the performance of segmentation. We will show how our adapted pipeline is able to better solve the investigated problem.

The structure of this work is as follows: Section 2 covers previous relevant research on the topic, Section 3 presents the datasets used for the experiments, Section 4 is dedicated to the proposed segmentation method, while Section 5 summarizes the details of the performed experiments and results. Finally, Section 6 provides the conclusion.

2. Related Work

Deep learning-based methods have previously been applied successfully to a range of medical imaging problems [15–19], including heart segmentation [1]. Two important challenges, ACDC [4] and MMWHS [3], are good benchmarks for measuring segmentation performance. In both cases, variations of the popular U-Net [20] is the preferred choice of researchers.

A significant point of study was the comparison of 2D and 3D methods. It has been shown that the performance depends highly on the acquisition methods [1,10]. Three-dimensional models are negatively affected by large inter-slice gaps and misalignments, whereas 2D models cope better in these situations. The results of Baumgarter et al. [10] confirm this on the ACDC dataset, with the ACDC data having a thickness of 5–8 mm and sometimes an inter-slice gap of 5 mm. The main drawback of using 2D models is that they cannot learn any inter-slice dependencies since they work slice-by-slice. On the other hand, on the MMWHS dataset, the best-performing models are 3D-based [3,21]. In [22] the authors proposed a novel intelligent framework designed ad hoc for enhancing MR image segmentation results of MRI data characterized by an underlying bimodal histogram.

Another significant topic is the two-step segmentation, where the region of interest (containing the heart) is first detected via a localization procedure, which is subsequently fed to the segmentation network. Different localization procedures were used, such as: regressing to the center of the heart, then cropping a fixed-size 3D region around the center, large enough to enclose all segmentation labels on every image from the training set. This strategy was proved effective as it was used to obtain the best results of the MMWHS challenge by Payer et al. [21].

Several other works experimented with similar ideas, for instance, by using a Faster RCNN network [23] or constructing a bounding box using three CNNs [24] that determine the presence of the heart independently in axial, coronal, and sagittal slices of the image volume.

These methods operate in a 3D scenario, aiming to obtain a 3D box which encompasses the entire heart. Other methods proposed to locate the heart by focusing on motion or heart beats, based on the changes in pixel intensities [25,26]. A somewhat simpler approach is proposed by Wang et al. in [27], where they use the output of a standard segmentation model to extract the ROI for another model, using the boundaries of the initial prediction. They used information from adjacent slices to compute the ROI for a single 2D slice.

3. Datasets

3.1. ACDC Dataset

The ACDC challenge [4] provides a total of 150 short-axis cine-MRI volumes from both diastolic and systolic phases, of which 100 are available for training and validating, 50 being reserved for testing. In this dataset, segmentation masks are provided for the following heart substructures: right ventricle (RV), myocardium (Myo), and left ventricle (LV).

3.2. imATFIB Dataset

The datasets were acquired in our in-house performed clinical study, imATFIB—IMaging-based, non-invasive diagnosis of persistent ATrial FIBrilation. The study is registered under the number NCT03584126 at clinicaltrials.gov (accessed on 22 February 2021). The study obtained ethical approval from the local Ethics committee (No. 20117/04.10.2016) and recruited subjects between 2017–2020. All subjects gave their written informed consent to participate in the study. Patients and healthy volunteers underwent cardiological evaluation using ECG and echocardiography, followed by cardiac MRI measurements.

All subjects were imaged with a 3T whole-body MRI system (3.0T Discovery MR750w General Electric MRI scanner) using a dedicated body coil for signal reception. Cardiac MRI required synchronous cardiac gating with ECG and breath-holding techniques to overcome motion artifacts. The cardiac MRI protocol consisted of dark blood sequences (Black Blood SSFSE), FIESTA cine sequences (ALL FIESTA CINE AST), and post-contrast sequences: rest perfusion (FGRE Time Course), angiography imaging (Aorta CEMRA Asset) and LGE/MDE images (2D MDE, 2D PSMDE) in three different planes: short axis, four-chamber, and two-chamber view. The MDE/LGE sequences were acquired with a time delay of 10 min after the injection of gadolinium, showing contrast between normal myocardium (dark) and abnormal myocardium (hyperenhancement). 3D HEART sequence was also acquired and used to assess coronary abnormalities. The participants were constantly monitored during the examination, and no immediate adverse effects were reported.

Ten datasets of healthy volunteers and ten datasets of patients with atrial fibrillation were selected and the heart was manually traced by two radiologists. The results were annotated manually by a radiologist with more than 10 years of experience and cross-checked by another, senior radiologist, with more than 20 years of experience in the field. The results were exported and stored in DICOM format.

Within the frame of our in-house performed clinical study, imATFIB, we obtained 20 cardiac MRI volumes with segmentation masks for the whole heart. The reason why the ACDC dataset was deemed a good benchmark is the similarity with the imATFIB dataset in characteristics such as resolution (256×256 for imATFIB), inter-slice gaps and slice thickness.

4. Method

This section describes the main parts composing the proposed method: architectures, data preprocessing, and ROI localization. We apply CNNs for the task of segmentation, namely, the U-Net [20] and DeepLab [28] architectures. Following several preprocessing steps, the input images are fed to the networks which, in turn, are trained to predict the semantic label associated with each image voxel. In case ROI extraction is used, the models are applied only on the selected part of the image.

4.1. Architectures

Most semantic segmentation CNN architectures follow the same general pattern: using a contracting component which gradually shrinks the height and width of the image, obtaining a coarse representation which has strong semantic value. This is subsequently followed by an upsampling component that rebuilds the original size of the image with the help of higher resolution features previously computed in the contracting part.

The U-Net we use differs from the standard version presented in [20] via the use of simple bilinear interpolation instead of transposed convolutions, the use of batch normalization [29] and the use of padding after each 3×3 convolution. Thus, the network outputs the same resolution as the input. The motivation behind opting to use plain interpolation is lowering complexity and memory restrictions, as no additional parameters are required for upsampling.

The DeepLabV3+ [28] model also follows this general pattern, usually applying a standard classification CNN as a backbone, followed by a simple upsampling path consisting of consecutive convolutional layers and bilinear upsampling. The backbone used in this work is ResNeXt50 [30], a variation of the standard ResNet50 [30] that makes use of grouped convolutions to boost performance, while maintaining model complexity and parameter numbers. The ASPP [31] module is replaced by a single convolution to save memory and computation, as well as to reduce the chance of overfitting.

Based on the ideas from [32], variations of these models were also tried, mainly reducing the width of the model in a bid to avoid overfitting, as well as trying multiple input resolutions in an attempt to find an optimal setup. Similar to [33], we use an extra parameter, the width multiplier, to control the width of the models. The number of filters in each layer is multiplied by the value of this parameter.

4.2. Data Preprocessing

Training volumes are split into slices and processed in a 2D manner, in batches, in a random order. The motivation behind using a 2D approach is given by previous research finding the performance of 2D vs. 3D approaches highly depends on the data characteristics. Specifically, large inter-slice gaps and misalignments (which is the case for the imATFIB dataset) negatively impact 3D models, resulting in slice-by-slice approaches, which are basically oblivious to inter-slice relations, to perform better. Data normalization is done by computing the mean value and standard deviation of each particular slice and using those values to normalize. All slices are resampled bilinearly to a specific height and width before being fed to the models. It is worth mentioning that per-slice normalization seems to be the best-performing method when it comes to MRI images [21,34], and it is the general conclusion from our experiments as well. However, there are also exceptions to this rule, where the best-performing DeepLab model on the ACDC validation set was trained using the mean and standard deviation of the training set to normalize slices.

With little data, augmentation is especially important, but there should also be a limit when applying augmentation, as too much of it can affect performance. We found that simple augmentations, such as horizontal mirroring and rotations, generally help improve segmentation, whereas elastic deformations, Gaussian noise and blurring can occasionally further boost performance. However, tuning the parameters regarding the intensity/strength of applying these augmentation proved difficult, resulting in inconsistent results (sometimes badly hurting performance). Notably, a popular augmentation technique, random crop, considerably hurts performance. A possible reason could be that heart structures always appear contiguously on MRI images, so a patch comprising only parts of it is not realistic. To be consistent, results presented in this work only use mirroring and rotations.

4.3. ROI Localization

The region of interest (ROI) is defined as the labelled area of an image slice, which in most cases is practically a small part of the image. Naturally, the question arises whether

models would benefit by focusing only on this region. To explore this idea, we denoted two types of segmentation models: Standard and ROI models. The first one expects a regular image as input, while the second one needs an ROI as input.

4.3.1. Perfect ROI

During training of the ROI models, the ground truth is used to extract the (perfect) ROI (the perfect ROI box is the rectangle that encompasses the ground truth segmentation mask). During inference, we do not have access to the ground truth. The localization step aims to compute the ROI by using the prediction of a Standard segmentation model (the process is similar, but this time the rectangle is computed around the mask resulting from the model prediction, instead of the ground truth mask).

We have first explored the theoretical benefit of a perfectly selected ROI. In this context, image slices are first upsampled bilinearly to 512×512, a bounding box of the labelled area is computed using the ground truth mask and the area delimited by this box is cropped, resized to 224×224, and then fed to the model. At the time of evaluation, the ROI was reinserted into the original image slice, resized to its original dimensions. Thus, the 3D volume is processed slice by slice, with the model itself processing only the ROI of each slice. The procedure is depicted in Figure 1.

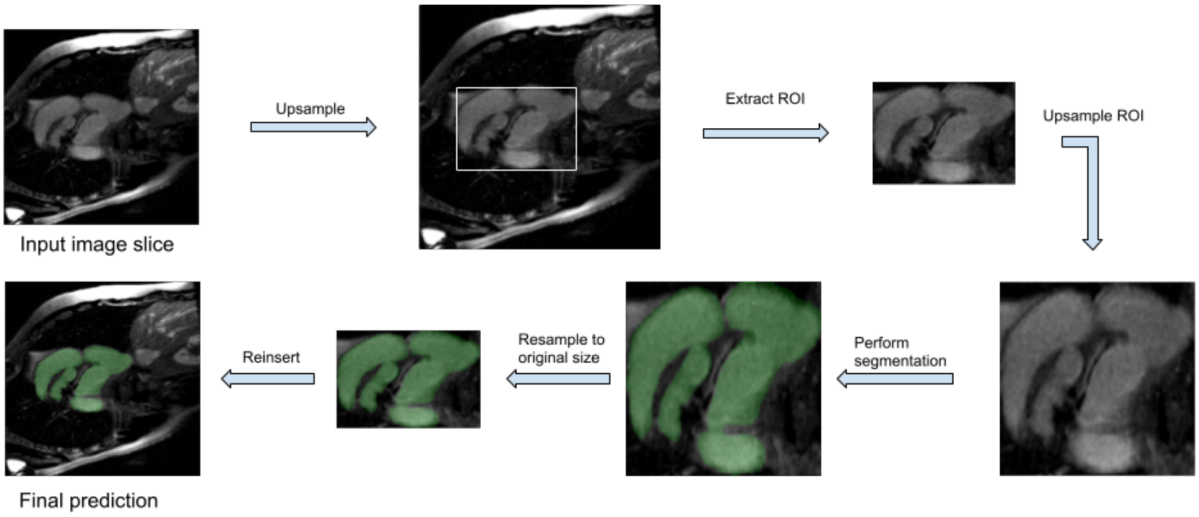

Figure 1. The flow-diagram that supports our approach. Before the training process starts, two scaling steps and a cropping step are performed: each original image slice is resized to 512×512 and, from this enlarged image, a ROI is cropped (by using the corresponding ground-truth) and resized to 224×224. The segmentation model is trained on these processed images and a set of masks are predicted for each input. Finally, the masks are re-scaled to the original dimensions and re-inserted in the original image.

4.3.2. ROI Extraction Mechanism

The ROI extraction mechanism we used is very similar to the one presented in [27], but in this work we simplified it to operate in an entirely 2D manner and further optimize training for prediction from dynamically computed ROIs. We concatenate two segmentation models (Standard and ROI) to perform segmentation, with the prediction of the former being used to compute the ROI for the latter.

Essentially, the steps are exactly the same as depicted in Figure 1, but instead of using the ground truth mask to compute the ROI box coordinates, the segmentation mask predicted by the Standard model is used.

The training setup is the same for all models: The models are optimized using Adam [35], with a learning rate of 0.001 which is divided by 10 after 60% of the epochs are complete, and again at 90%, batch size of 12 for the U-Net (16 for DeepLab) and a weight de-

cay of 0.0001. The training duration is 300 epochs (there was no extensive hyper-parameter search done).

4.3.3. Standard Model Training

Since cardiac image segmentation is perhaps less complex than other segmentation tasks [36], we found that dividing the number of filters in the DeepLab model by a factor of 3.2, coupled with increasing input resolution to 480×480 generally improved performance, which is in line with the original purpose of the model's design (multiple-class, high resolution image segmentation). On the other hand, U-Net did best on the imATFIB dataset with a standard number of channels and 224×224 input resolution, while on the ACDC dataset it did benefit from a higher input resolution, 320×320, paired with a width multiplication factor of 0.5. Experiments presented in section 5 were performed with the settings mentioned above.

4.3.4. ROI Model Training

Training an ROI model was done by computing the target ROIs using the boxes from the ground truth masks. Empirical observations show that the ROI boxes obtained by the Standard model are larger and usually do not respect the ratio of the true ROI box. In an attempt to account for the mistakes of the Standard model, we trained the ROI models to successfully perform segmentation from larger than perfect ROIs.

To achieve this, we specifically optimize the training of the ROI model, by perturbing the ground truth ROI boxes randomly to reflect the shapes and sizes of dynamically computed ROI boxes. Perturbation was performed both on width and height starting from a minimum value up to a certain threshold (usually a fifth of the size of the box). Moreover, we also imposed a minimum size of the ROI box, to avoid artifacts produced by aggressive bilinear upsampling afterwards. In order to avoid randomness in validation, the perturbation was set to a fixed size at time of evaluation (half the value of the max threshold). All ROI models had the same input resolution, namely 224×224.

Accordingly, the perturbation during training is done as follows: we sampled an individual perturbation value pi within the range (min_p, max_p), for each of the four coordinates of the bounding box: $(x1, y1)$ top left and $(x2, y2)$ bottom right, respectively. Then, the perturbation operations are:

$$x1 = x1 - p1;\ y1 = y1 - p2$$
$$x2 = x2 + p3;\ y2 = y2 + p4.$$

4.3.5. Ensembling

Given the variety of methods to perform segmentation, we also use ensembling techniques in an effort to further boost performance. Softmax predictions from a number of models are summed and then averaged to obtain the final ensemble prediction.

5. Experiments and Results

Both the imATFIB and ACDC datasets are split 80/20 into a training and validation set. Evaluation is done at the original resolution of the images. The mean Dice is computed by averaging over the Dice of each volume. For the imATFIB dataset, we measure the Dice on two classes only: heart vs background. An example set of images from the imATFIB dataset is presented in Figure 2. Evaluation scripts provided by the ACDC challenge were used to compute metrics.

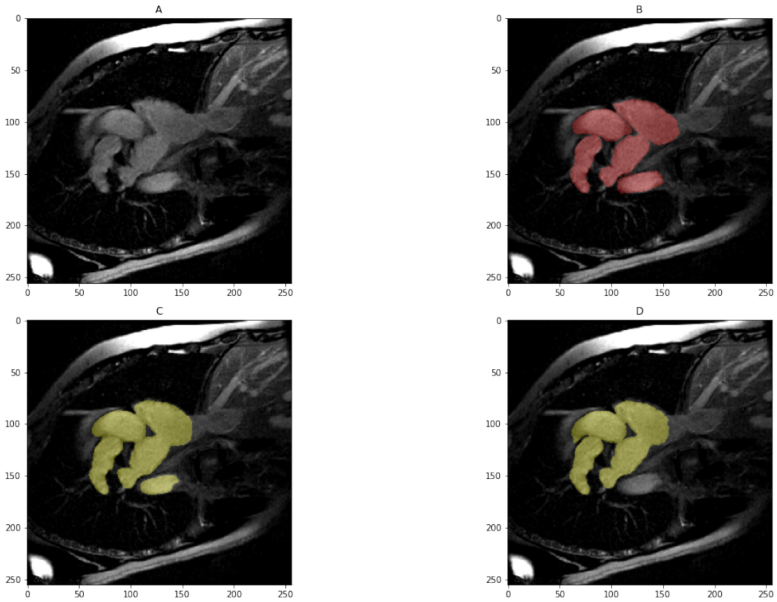

Figure 2. Sample results from the imATFIB validation set (Red—ground truth, Yellow—model prediction). (**A**) Input image, (**B**) Label mask, (**C**) All Ensemble model (M_9), (**D**) U-net model (M_4).

Table 1 presents reference results for ROI models validated using the ROI computed from the ground truth masks. These results should set a theoretical upper limit to segmentation performance when using ROIs with the models applied, as well as target values to achieve with dynamically computed ROIs. These are the same ROI models that will be used for validation and testing in the following steps. Therefore, any performance drop is certain to be caused by imperfections in computing the ROI boxes. Although no statistically significant differences were observed in terms of the mean Dice value, the U-net is consistently more robust, especially on the ACDC dataset, having a much better result even on the most difficult validation volumes.

Table 1. Perfect ROI results on the ACDC and imATFIB validation sets. The imATFIB dataset "mean" refers to the average Dice of all volumes (whole-heart vs background), whereas the ACDC dataset "mean" refers to averaging the dices on the three segmented heart substructures as well (StDev—Standard deviation, Min–Max—range determined by the minimum, respective maximum dice values obtained on the volumes).

Model	Dataset	Mean Dice	StDev	Min–Max
DeepLab	ACDC	92.6	5.4	67.4– 97.5
Unet	ACDC	92.9	3.6	83.8–97.5
DeepLab	imATFIB	89.7	1.0	88.4–90.8
Unet	imATFIB	90	1.4	88.5–92.3

Tables 2 and 3 show the results obtained on the imATFIB and ACDC validation sets, respectively. On both datasets, the U-net consistently maintains a lower standard deviation and a higher dice value on its worst volumes, which indicates better generalization and more stable performance than the DeepLab model. Furthermore, in Table 2 the last model has the best mean dice value, but also the narrowest Min-Max interval and the second the smallest StdDev.

Table 2. Results on the imATFIB validation set (Standard Ensemble: Ensemble of best standard Unet and DeepLab, ROI Ensemble: Ensemble of ROI pipelines (M_5 and M_6), All Ensemble: Ensemble of the previous two ensembles (M_7 and M_8), Ensembling is done by averaging over the predicted softmax values of the models).

Model	Pretraining	Mean Dice	StDev	Min–Max
M_1: DeepLab	No	85.5	2.0	82.5–88.1
M_2: DeepLab	Yes	86.6	1.9	83.4–88.6
M_3: Unet	No	88.2	1.2	85.8–88.9
M_4: Unet	Yes	88.2	1.2	85.8–88.9
M_5: Unet & Unet	-	89.34	1.1	87.8–90.4
M_6: Unet & DeepLab	-	88.70	0.9	87.5–90.0
M_7: Standard Ensemble	-	89.25	0.8	87.9–90.2
M_8: ROI Ensemble	-	89.41	1.2	88.0–91.3
M_9 : All Ensemble	-	89.89	0.9	88.8–91.2

Table 3. Results on the ACDC validation set (RV—right ventricle, LV—left ventricle, Myo—myocardium).

Model	Dice RV	Dice LV	Dice Myo	Mean Dice	StDev	Min–Max
M_1: DeepLab	90.50	94.50	89.10	91.37	3.2	80.6–95.2
M_3: Unet	89.32	95.07	89.47	91.29	3.0	83.1–95.5
M_{10}: DeepLab & DeepLab	90.58	94.72	89.53	91.61	3.4	77.5–95.9
M_{11}: DeepLab & Unet	90.65	95.02	89.94	91.87	3.0	81.7–95.7
M_{12}: Standard Ensemble (M_1 and M_3)	90.48	95.12	89.89	91.83	3.0	82.5–95.9
M_{13}: ROI Ensemble (M_{10} and M_{11})	90.88	94.98	90.11	91.99	3.2	80.0–95.9
M_{14}: All Ensemble (M_{12} and M_{13})	91.26	95.39	90.49	92.38	2.9	81.7–96.2

Related to the first four scenarios presented in Table 2, two of them involve a transfer learning approach: starting from a pretrained set of weights obtained by first training on the ACDC dataset, all layers are initially frozen, except the final multi label classification layer (in this case only two labels, heart and background). Layers are then progressively unfrozen up until the 50% of the number of epochs are complete. The training setup is the same as before.

Notably, consistent improvements are made via ROI localization and ensembling on both datasets. For ROI extraction, all combinations of model ordering were tested, but only the best-performing ones were written in the tables. The rule for applying ROI extraction is simple, the prediction of the best-performing standard model is used for extraction. This rule determines the Standard Model whose prediction is used for ROI extraction, and then both ROI models are validated with it. Consistent with Table 1, the best results as an ROI model are obtained by the U-Net. Transfer learning did not have a major impact. While bringing an improvement to the performance of the DeepLab, it failed to lead to better convergence for the U-Net.

In the case of both datasets, the performance of the ROI segmentation (from computed boxes) worsened minimally compared to the perfect ROI results, which was expected. It should be mentioned that the ROI models were trained specifically to account for imperfect ROIs, so it is most likely possible to obtain better perfect ROI performance with optimizations in this sense.

Promisingly, the performance gained through ensembling is almost as much as through ROI localization, which proves that ensembling fundamentally different architectures is highly beneficial. Specifically, DeepLab differs from U-Net principally in depth, as well as the stride of the contracting path (the value by which the input image dimensions get divided to obtain the smallest feature map: 16 by default for DeepLab, 32 for U-Net).

Finally, to officially benchmark our proposed approach, we made two submissions to the ACDC challenge. The first one is the best-performing DeepLab model on the ACDC

validation set, which is then compared to the All Ensemble. Results are presented in Tables 4 and 5. The Hausdorff metric is a measure of distance between sets (the smaller the better), especially sensitive to outliers, and it is one of the two main metrics of the ACDC challenge (along with the Dice coefficient). Improvements remain consistent with the validation set, and the end result of a mean Dice of 91.87% is decent, being comparable to the second best-performing approach on the challenge [4].

Table 4. Results on the ACDC validation set (RV—right ventricle, LV—left ventricle, Myo—myocardium).

Model	Dice RV	Dice LV	Dice Myo	Mean Dice	StDev	Min–Max
M_1: DeepLab	90.50	94.50	89.10	91.37	3.2	80.6–95.2
M_3: Unet	89.32	95.07	89.47	91.29	3.0	83.1–95.5
M_{10}: DeepLab & DeepLab	90.58	94.72	89.53	91.61	3.4	77.5–95.9
M_{11}: DeepLab & Unet	90.65	95.02	89.94	91.87	3.0	81.7–95.7
M_{12}: Standard Ensemble (M_1 and M_3)	90.48	95.12	89.89	91.83	3.0	82.5–95.9
M_{13}: ROI Ensemble (M_{10} and M_{11})	90.88	94.98	90.11	91.99	3.2	80.0–95.9
M_{14}: All Ensemble (M_{12} and M_{13})	91.26	95.39	90.49	92.38	2.9	81.7–96.2

Table 5. Results on the ACDC test set—Hausdorff Distance (RV—right ventricle, LV—left ventricle, Myo—myocardium).

Model	Hausdorff RV (mm)	Hausdorff LV (mm)	Hausdorff Myo (mm)
M_1: DeepLab	24.27	11.93	14.46
M_{14}: All Ensemble	13.21	10.99	12.08

6. Conclusions

In this paper, we presented data on training and testing fundamentally different model architectures, one of which not being explicitly designed for medical image segmentation, DeepLab, and showed that ensembling of such models is beneficial. Starting from an existing idea [27], we optimized training for ROI segmentation in an entirely 2D slice-by-slice manner, obtaining close results to theoretical bests. We are aware that our study has limitations—the size of our acquired pilot dataset is very small. However, we also tested our method on a publicly available, significantly larger dataset, on which we obtain good results, also. Thus, we provide stronger validation for our method.

In conclusion, we presented a simple, practical approach to cardiac image segmentation from MRI images, demonstrating that both ensembling of different architectures and using ROI localization improved performance. Moreover, this approach should be applicable to any segmentation problem where an ROI might be present. Our approach provides an opportunity to serve as a building block of a computer-aided diagnostic system in a clinical setting.

As future work, we plan to investigate a similar approach in the 3D context and a semi-supervised approach (when in the training set there are some labeled images and some unlabeled images). Furthermore, we plan to integrate the proposed approach in a real system (in this case, some quantisation methods could be required in order to reduce/scale down the learned model).

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

Funding: The authors highly acknowledge financial support from the Competitiveness Operational Program 2014-2020 POC-A1-A1.1.4-E-2015, financed under the European Regional Development Fund, project number P37_245.

Institutional Review Board Statement: The study obtained ethical approval from the local Ethics committee (Nr.20117/04.10.2016) and recruited subjects between 2017–2020.

Informed Consent Statement: All subjects gave their written informed consent to participate in the study. Patients and healthy volunteers underwent cardiological evaluation using ECG and echocardiography, followed by cardiac MRI measurements.

Data Availability Statement: Source code is available at: https://github.com/RonaldGalea/imATFIB.

Acknowledgments: The excellent technical assistance of Mihaela Coman, Cristina Szabo and Silviu Ianc is highly appreciated.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, C.; Qin, C.; Qiu, H.; Tarroni, G.; Duan, J.; Bai, W.; Rueckert, D. Deep Learning for Cardiac Image Segmentation: A Review. *Front. Cardiovasc. Med.* **2020**, *7*, 25, doi:10.3389/fcvm.2020.00025.
2. Alom, M.Z.; Yakopcic, C.; Hasan, M.; Taha, T.M.; Asari, V.K. Recurrent residual U-Net for medical image segmentation. *J. Med. Imaging* **2019**, *6*, 014006.
3. Zhuang, X.; Li, L.; Payer, C.; Štern, D.; Urschler, M.; Heinrich, M.P.; Oster, J.; Wang, C.; Smedby, O.; Bian, C.; et al. Evaluation of algorithms for Multi-Modality Whole Heart Segmentation: An open-access grand challenge. *Med. Image Anal.* **2019**, *58*, 101537, doi:10.1016/j.media.2019.101537.
4. Bernard, O.; Lalande, A.; Zotti, C.; Cervenansky, F.; Yang, X.; Heng, P.A.; Cetin, I.; Lekadir, K.; Camara, O.; González Ballester, M.A.; et al. Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? *IEEE Trans. Med. Imaging* **2018**, 1, doi:10.1109/TMI.2018.2837502.
5. Trullo, R.; Petitjean, C.; Dubray, B.; Ruan, S. Multiorgan segmentation using distance-aware adversarial networks. *J. Med. Imaging* **2019**, *6*, 014001.
6. Tajbakhsh, N.; Jeyaseelan, L.; Li, Q.; Chiang, J.N.; Wu, Z.; Ding, X. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Med. Image Anal.* **2020**, *63*, 101693.
7. Mărginean, R.; Andreica, A.; Dioşan, L.; Bálint, Z. Feasibility of Automatic Seed Generation Applied to Cardiac MRI Image Analysis. *Mathematics* **2020**, *8*, 1511.
8. Militello, C.; Rundo, L.; Toia, P.; Conti, V.; Russo, G.; Filorizzo, C.; Maffei, E.; Cademartiri, F.; La Grutta, L.; Midiri, M.; et al. A semi-automatic approach for epicardial adipose tissue segmentation and quantification on cardiac CT scans. *Comput. Biol. Med.* **2019**, *114*, 103424.
9. Commandeur, F.; Goeller, M.; Razipour, A.; Cadet, S.; Hell, M.M.; Kwiecinski, J.; Chen, X.; Chang, H.J.; Marwan, M.; Achenbach, S.; et al. Fully automated CT quantification of epicardial adipose tissue by deep learning: A multicenter study. *Radiol. Artif. Intell.* **2019**, *1*, e190045.
10. Baumgartner, C.F.; Koch, L.M.; Pollefeys, M.; Konukoglu, E. 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*; Springer: Cham, Switzerland, 2017.
11. Liu, L.; Cheng, J.; Quan, Q.; Wu, F.X.; Wang, Y.P.; Wang, J. A survey on U-shaped networks in medical image segmentations. *Neurocomputing* **2020**, *409*, 244–258.
12. Schlemper, J.; Oktay, O.; Schaap, M.; Heinrich, M.; Kainz, B.; Glocker, B.; Rueckert, D. Attention gated networks: Learning to leverage salient regions in medical images. *Med. Image Anal.* **2019**, *53*, 197–207.
13. Rundo, L.; Han, C.; Nagano, Y.; Zhang, J.; Hataya, R.; Militello, C.; Tangherloni, A.; Nobile, M.S.; Ferretti, C.; Besozzi, D.; et al. USE-Net: Incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets. *Neurocomputing* **2019**, *365*, 31–43.
14. Isensee, F.; Jaeger, P.F.; Kohl, S.A.; Petersen, J.; Maier-Hein, K.H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **2021**, *18*, 203–211.
15. Kavalcova, L.; Skaba, R.; Kyncl, M.; Rouskova, B.; Prochazka, A. The diagnostic value of MRI fistulogram and MRI distal colostogram in patients with anorectal malformations. *J. Pediatr. Surg.* **2013**, *48*, 1806–1809.
16. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88.
17. Altaf, F.; Islam, S.M.; Akhtar, N.; Janjua, N.K. Going deep in medical image analysis: Concepts, methods, challenges, and future directions. *IEEE Access* **2019**, *7*, 99540–99572.
18. Fourcade, A.; Khonsari, R. Deep learning in medical image analysis: A third eye for doctors. *J. Stomatol. Oral Maxillofac. Surg.* **2019**, *120*, 279–288.
19. Aldoj, N.; Biavati, F.; Michallek, F.; Stober, S.; Dewey, M. Automatic prostate and prostate zones segmentation of magnetic resonance images using DenseNet-like U-Net. *Sci. Rep.* **2020**, *10*, 1–17.
20. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015.

21. Payer, C.; Stern, D.; Bischof, H.; Urschler, M. Multi-label Whole Heart Segmentation Using CNNs and Anatomical Label Configurations. Statistical Atlases and Computational Models of the Heart. In Proceedings of the ACDC and MMWHS Challenges—8th International Workshop, STACOM 2017 (Held in Conjunction with MICCAI 2017), Quebec City, QC, Canada, 10–14 September 2017; Lecture Notes in Computer Science; Pop, M., Sermesant, M., Jodoin, P., Lalande, A., Zhuang, X., Yang, G., Young, A.A., Bernard, O., Eds.; Springer: Berlin/Heidelberg, Germany, 2017; Volume 10663, pp. 190–198, doi:10.1007/978-3-319-75541-0_20.
22. Rundo, L.; Tangherloni, A.; Cazzaniga, P.; Nobile, M.; Russo, G.; Gilardi, M.; Vitabile, S.; Mauri, G.; Besozzi, D.; Militello, C. A novel framework for MR image segmentation and quantification by using MedGA. *Comput. Methods Programs Biomed.* **2019**, *176*, 159–172, doi:10.1016/j.cmpb.2019.04.016.
23. Ren, S.; He, K.; Girshick, R.B.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv* **2015**, arXiv:1506.01497.
24. Zreik, M.; Leiner, T.; de Vos, B.D.; van Hamersvelt, R.W.; Viergever, M.A.; Išgum, I. Automatic segmentation of the left ventricle in cardiac CT angiography using convolutional neural networks. In Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 13–16 April 2016; pp. 40–43.
25. Nasr-Esfahani, M.; Mohrekesh, M.; Akbari, M.; Soroushmehr, S.M.R.; Nasr-Esfahani, E.; Karimi, N.; Samavi, S.; Najarian, K. Left Ventricle Segmentation in Cardiac MR Images Using Fully Convolutional Network. *arXiv* **2018**, arXiv:1802.07778.
26. Constantinides, C.; Chenoune, Y.; Mousseaux, E.; Roullot, E.; Frouin, F. Automated heart localization for the segmentation of the ventricular cavities on cine magnetic resonance images. In Proceedings of the 2010 Computing in Cardiology, Belfast, UK, 26–29 September 2010; pp. 911–914.
27. Wang, C.; MacGillivray, T.; Macnaught, G.; Yang, G.; Newby, D.E. A two-stage 3D Unet framework for multi-class segmentation on full resolution image. *arXiv* **2018**, arXiv:1804.04341.
28. Chen, L.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *arXiv* **2018**, arXiv:1802.02611.
29. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2015**, arXiv:1502.03167.
30. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
31. Chen, L.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *arXiv* **2016**, arXiv:1606.00915.
32. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv* **2019**, arXiv:1905.11946.
33. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
34. Isensee, F.; Jaeger, P.; Full, P.M.; Wolf, I.; Engelhardt, S.; Maier-Hein, K.H. Automatic Cardiac Disease Assessment on cine-MRI via Time-Series Segmentation and Domain Specific Features. *arXiv* **2017**, arXiv:1707.00587.
35. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization, 2014. In Proceedings of the 3rd International Conference for Learning Representations, San Diego, CA, USA, 7–9 May 2015.
36. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset for Semantic Urban Scene Understanding. *arXiv* **2016**, arXiv:1604.01685.