Myocardial Infarction Segmentation From Late Gadolinium Enhancement MRI By Neural Networks and Prior Information

Zhihao Chen*, Alain Lalande†, Michel Salomon*, Thomas Decourselle‡,
Thibaut Pommier§, Gilles Perrot*, and Raphaël Couturier*
*FEMTO-ST Institute, CNRS UMR 6174
Univ. Bourgogne Franche-Comte, Belfort, France
Email: {zhihao.chen,michel.salomon,gilles.perrot,raphael.couturier}@femto-st.fr
†ImVia Laboratory / Medical Imaging Department
Univ. Bourgogne Franche-Comte / CHU Dijon, Dijon, France
Email: alain.lalande@u-bourgogne.fr
‡CASIS Company, Dijon, France
Email: tdecourselle@casis.fr
§Cardiology Department
CHU Dijon, Dijon, France
Email: thibaut.pommier@chu-dijon.fr

***Abstract*—In this paper, we propose an automatic myocardial infarction segmentation framework from Delayed Enhancement cardiac MRI (DE-MRI) using a convolutional neural network (CNN) and prior information-based post-treatments. The work was conducted on our DE-MRI dataset, which is collected from daily clinical practice. 195 cases of DE-MRI examinations constitute this dataset, including on average 7 images per case with manually drawn contours by an expert. The objective is to automatically segment myocardial infarctions on both healthy and pathological images in the dataset. In the proposed framework, a downsampling-upsampling segmentation CNN firstly generates high recall segmentations of myocardial infarction from left ventricle DE-MR images, then the proposed prior information-based post-processing method identifies and removes false-positive segmentations from the CNN's prediction. To obtain a high recall prediction, two U-NET like semantic segmentation networks are investigated: CE-NET and its backbone with Dice loss and Stochastic Gradient Descent (SGD) using a batch size of value 1. The prior information-based post-processing evaluates every single contour in the CNN's segmentations: region features in each contour are compared to criteria which are firstly estimated based on the training set images and eventually fine-tuned based on the validation set images. All non-conforming contours are removed from the predictions to improve the accuracy of the segmentation. Combining the high recall networks and prior post-processing information, we achieve segmentation results comparable to those produced by human experts.**

***Index Terms*—segmentation, deep learning, prior information, myocardial infarction, DE-MRI**

This work was partly supported by the ADVANCES project founded by ISITE-BFC project (number ANR-15-IDEX-0003) and by the EIPHI Graduate School (contract ANR-17-EURE-0002). We also thank the Mesocentre of Franche-Comté for the computing facilities.

I. Introduction

Late Gadolinium Enhancement Magnetic Resonance Imaging (Delayed enhancement MRI or DE-MRI) has become the reference exam for myocardial infarction quantification. In such an image, acquired several minutes after the injection of the gadolinium contrast agent, normal myocardiums and ischemic tissues show different signals, that allow to distinguish the myocardial infarction (that appears in bright) from its surrounding healthy tissues (that appear in grey). Furthermore, this exam has been largely applied worldwide for accurate myocardial ischemic and non-ischemic pathology inspection [1] in routine clinical practice.

The myocardial infarction segmentation is mainly carried out by experts manually annotating the acquired images. This task is time-consuming and depends highly on the experience of the clinical experts. A robust fully automatic myocardial infarction segmentation approach for DE-MRI shows great potential for the clinical practice. With such an end-to-end solution, the physician only needs to validate the automatic segmentation con-

tours and add some minor corrections when necessary. Meanwhile, the design of an automatic segmentation method able to predict on mixed normal and pathological images is very challenging. The network should be able to predict on image-level and pixel-level at the same time, i.e., both the recall and the accuracy of the segmentation are demanding. Other difficulties include the presence of noises, the lack of contrast between normal and pathological areas due to the partial volume effect and also to the grey zone between myocardial infarction and normal myocardium. Finally, the heterogeneity in the signal within the dataset makes it difficult to design a robust method.

Current cardiac infarction segmentation approaches are principally rule-based models, including threshold-based approaches like Full-Width at Half-Maximum (FWHM) [2] or n-Standard Deviations (n-SD) [3], and clustering-based approaches like Gaussian Mixture Models (GMM) [4]. In the framework of a MICCAI challenge, Karim *et al.* [5] have compared most of the current rule-based models from human and animal datasets, showing that although they are time-saving, their segmentation quality generally lags far behind the intra-observation and inter-observation variations of the datasets.

Learning-based approaches for medical imaging segmentation have been widely researched and applied to cardiac MRI. However, few works concentrate on the myocardial infarction segmentation. For example, De La Rosa *et al.* [6] applied first a CNN classifier to detect pathological cases followed by a morphological treatment to generate the segmentation of infarctions. More works have been conducted for myocardium segmentation on LGE MRI. E.g., Avendi et al. [7] combine deep learning with a deformable model for left ventricle segmentation.

For the general-purpose semantic segmentation by deep learning, Fully Convolutional Networks (FCNs) [8] use deconvolution as the upsampling operation for detail reconstructions. To achieve a higher spatial precision of pixel-level semantic segmentation, U-Net [9] and its variants M-net [10], FusionNet [11], CE-Net [12], etc. optimize the upsampling phases through skip connections: such connections share the feature maps from a downsampling phase to the corresponding upsampling one so that localization information can be taken into account during the reconstruction of segmentation. Besides the design of the network architecture, some optimization tricks of the loss function and convolutional blocks have been proved meaningful on some application fields. Metric-based loss functions such as Generalized Dice [13] and the combination of Dice and Cross Entropy (CE) [14] can thus outperform the CE or the L1/L2 distance on some specific data. The substitution of convolutional blocks can also improve the segmentation results under some circumstances. However, this benefit can be mitigated by the larger number of network weights. Apart from the downsampling - upsampling approaches, Generative Adversarial Networks (GAN) have also been employed for medical image segmentation. Majurski *et al.* [15] and Xue *et al.* [16] have thus investigated such generative models showing good results on cell image and brain MRI segmentation.

It should be noted that in most of the previous works, the authors assume that the images to be segmented are always infarction images. The problem of image-level prediction (if an image is showing an infarction or not) has not been taken into account, which means that most of these approaches could not guarantee the correctness of the prediction on non-pathological images. Additional designs are potentially necessary to overcome this constraint, e.g., the classification of an image before using a segmentation network, or alternatively to classify the image after segmentation.

In the light of what has been said above, this paper proposes an automatic myocardial infarction segmentation framework. This framework receives the left ventricle myocardium from mixed normal and pathological DE-MRI in short-axis orientation as input and generates segmentation contours of myocardial infarction tissues. Our proposal includes three steps as shown in Fig. 1 from left to right: image preprocessing, image segmentation by an ensemble [17] of downsampling - upsampling segmentation CNNs, and prior information based post-processing. The contributions of this work are twofold. First, we propose an efficient framework for the infarction segmentation on highly class-imbalanced DE-MRI. The main strategy is the generation of high recall predictions through image preprocessing and CNN image segmentation, followed by the selection of good segmentation contours from the CNN's output during the prior information-based post-processing. Second, the design of a novel prior information-based post-processing.

The remainder of the paper is organized as follows. In Section II is explained how our dataset was built, giving in particular some details on the acquisition process. In the next section we describe the proposed segmentation framework, focusing on the segmentation CNN and the post-processing method. Experimental results are then presented in Section IV and more deeply discussed in

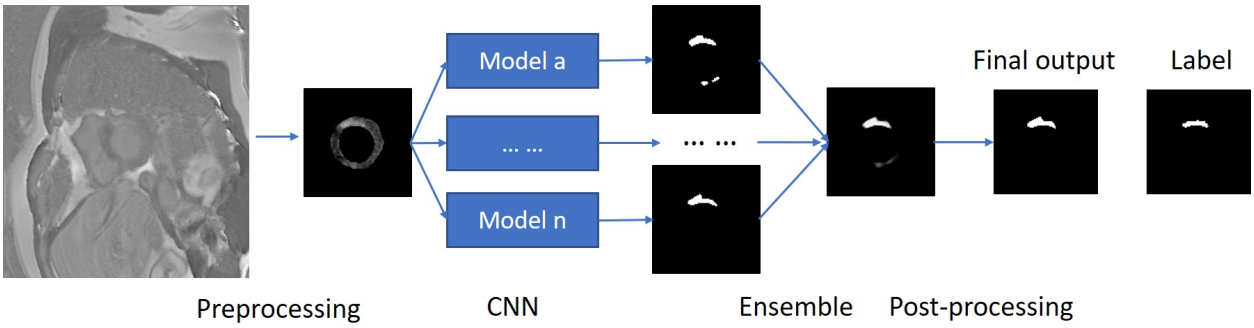

Fig. 1. Illustration of the proposed framework considering a single 2D input image. For a 2.5D input, three neighbouring images replace the single input image.

the final section, as well as directions for future work.

II. Dataset

The dataset is composed of 195 patients' DE-MRI examination in short-axis orientation, with on average 7 slices that cover the left ventricle from the base to the apex. MRI acquisitions were performed on 1.5 T and 3 T magnet (Siemens Medical Solution, Erlangen, Germany) with a phased thoracic coil. Images were acquired 10 minutes after the injection of gadolinium-based contrast agent (Gd-DTPA; Magnevist, Schering- AG, Berlin, Germany) using a T1-weighted phase-sensitive inversion recovery (PSIR) sequence. The mean pixel spacing is 1.55 mm² ± 0.20 mm² according to the patient, thus the left ventricular myocardium is always included in a square of 80×80 pixel². Annotations of the myocardium and myocardial infarction of each slice of MR image were also provided. The myocardial contours and the myocardial infarction areas (if present) were manually drawn by an expert in the domain (with more than 15 years of experience) on each slice, allowing the accurate definition of the myocardium and the infarction without considering the cavity.

For the experiments, the dataset is randomly divided into training, validation, and test sets, corresponding respectively to 888, 241, and 194 images. One entire examination only belongs to one separated set. To evaluate manual annotations' variation, the expert repeated the manual segmentation on 35 cases randomly selected. Another expert was invited to annotate on the same cases to check the inter-observation agreement, in order to validate the reliability of the dataset and quantify the human expertise. The test set includes the same cases as the intra-observer and the inter-observer variation studies.

As the MR images were collected during daily clinical practice, it can objectively represent the challenges the medical imaging experts are facing. Difficulties such as artifact, poor contrast, blurry effect cannot be completely avoided. Besides the biases produced during image acquisition, the internal variations among images are more challenging: images on the apex or base could appear very differently; patients' physical state results in different contrast ratios, etc.

III. Methods

Applying existing segmentation frameworks on mixed normal-pathological images showed that on pathological images the result is satisfying. But on normal slices, non-pathological tissues are frequently segmented as infarctions. In this paper we will mainly present the solution for false-positive prediction elimination (normal tissue is predicted as infarction). Indeed our novel prior information-based post-processing method is specially designed for mixed normal-pathological images segmentation.

The proposed framework includes preprocessing, segmentation by CNN, and post-processing. Preprocessing first transforms data from whole MRI acquisition to focus on the myocardium (without cavity) as networks' inputs. The resulting images are then normalized to enhance the contrast. The obtained preprocessed images are then fed to the CNN for myocardial infarction segmentation. The different CNNs we considered are trained using Stochastic Gradient Descent (SGD) (with a batch size of one) instead of mini-batch SGD to get a higher prediction recall. A set of different network models' predictions avoids uncertainty among different training epochs. Finally, prior information-based post-processing identifies each contour of a segmentation such that only conforming contours are kept as final contours.

A. CNNs of High Recall Segmentation

Motivated by the huge success of the recently published CE-Net architecture on multiple kinds of data (cell, 2D CT images, optic disc etc.) [12], we chose it as our segmentation CNN model. For comparison, the backbone of CE-Net (CE-Net without dense atrous convolution (DAC) and residual multi-kernel pooling (RMP) blocks), which is very similar to FusionNet [11], has also been tested on same data. CE-Net is a variation of U-Net architecture [9]. Like the latter it uses skip connections between mirrored layers in both downsampling and upsampling stacks in order to recover the lost spatial information during downsampling. As in FusionNet , residual blocks [18] replace the repeated convolutions so that the feature extraction is more efficient. The novelty of CE-Net is the context extractor which is formed of DAC [19] and RMP blocks. In DAC block, densely connected convolutions extract information of different depths, allowing to decrease feature map size without losing semantic information. While in RMP block, residual pooling of different kernel size optimizes predictions of both big and small objects.

Clinical practice shows that neighbouring images can be referred to assist segmentation decision. Inspired by it, we arrange 3 successive images in 3 channels as a 2.5D input. To understand such 2.5D input, an 1D convolution layer is added at the beginning of the network. This layer merges the information between neighbouring images. Fig. 2 shows the 2.5D input CE-Net based network structure.

In most CNN applications, mini-batch SGD makes training more robust and efficient [20]. Mini-batch SGD takes into account all samples in the batch and returns the average of the gradients. On one hand a large batch size smoothes the update path during training by reducing the variance of weight updates, but the gradient of noisy samples has little impact. On the other hand a small batch size has a higher risk of misleading the update path. However, preliminary experiments indicate that SGD (batch size of value 1) produces better segmentations. In fact, the images in our dataset have large variations in their characteristics, base and apex images in particular can be quite different. As the mini-batch SGD tends to ignore noisy features during backpropagation, the variations could be identified as noises. That means that some useful information is dropped out during training, while with SGD all features in a image contribute equivalently to the weights update. Despite learning both features of data variations and noises, and decelerating the training, SGD outperforms mini-batch SGD on our dataset.

CE is one of the most common loss functions in semantic segmentation networks. Nevertheless, it does not work properly on our data because of the important class imbalance. Weighted CE can alleviate this problem, but the weight parameter is very difficult to determine. Hence, we adopted the Dice loss in our experiments.

B. Post-processing with Prior Information

We use the ensemble technique to merge multiple network models' predictions to avoid the uncertainty of the different training epochs. Hence, during the training stage, multiple weights of the same network structure are saved before early stopping. A saving occurs at the end of a series of epochs whose interval depends on that of the batch size used during training. For the validation and test stages, these models of the same network structure perform independent segmentations on each input image before eventually merging their results. Finally, the merged results are converted to binary images and the binary threshold value is determined using the validation set.

The prior information-based post-processing refers to the rules that clinical experts employ for manual annotation drawing and inspection. Based on these rules and relative prior information in the data, criteria are constructed to eliminate false-positive contours in the CNN's segmentations. To design such criteria, we firstly investigated the image characteristics which can be considered as criteria, then we inspected the statistics of prior information in the training set to study the relevance of the proposed criteria. Since our data have significant variations, the criteria should be universal enough for the whole dataset. On the strength of this inspiration, three criteria were designed to improve the accuracy of the segmentation.

The first criterion focuses on the infarction size. In clinical practice, a validated myocardial infarction should have a minimum area. Otherwise, it will be considered as a noisy region, e.g. an artifact. Therefore, this criterion compares the area of each single segmentation in CNN's outputs to a threshold value of the area. All segmentation contours which do not meet the condition:

$$Area(C) > \theta_{area} \tag{1}$$

are removed, where C is one contour belonging to a coarse segmentation and θ_{area} is the threshold area. Impacts of noises on the CNN's segmentations such as the partial volume of fat or cavity, artifacts of high signal should be eliminated by this criterion.

The second and third criteria take advantage of the contrast agent's nature. The gadolinium contrast agent

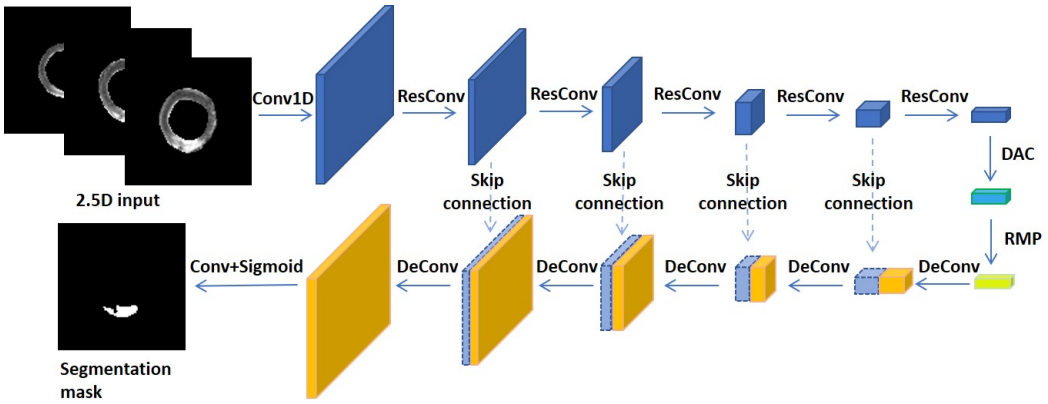

Fig. 2. Illustration of experimented 2.5D CE-NET. ResConv doubles the number of feature maps and halves their resolution. Skip connection concatenates corresponding encoder feature maps and decoder feature maps. For 2D input, the 1D convolution layer is removed; backbone network keeps same architecture as above except DAC and RMP blocks which are removed.

can enhance the signal intensity on our T1-weighted MR images and agent-enriched tissues show high signals. Blood transports the contrast agent in cardiac tissues. The absorption and the release rate of the agent depend on the type of tissue, which allows us to distinguish each tissue in DE-MRI. Therefore, 10 minutes after the injection of the contrast agent, infarction tissues and cavity show high signals, while normal myocardium tissues appear dark. According to the above rules, the second criterion compares the median signal value of the cavity (signal of the blood) and the candidate area:

$$M(C) - M(cavity) > \theta_{signal} \tag{2}$$

where $M()$ means the median signal value in the region, C stands for the region of the candidate contour, and $cavity$ stands for the region of the cavity. If the signal in the region of the candidate contour is high enough compared to the cavity, this area will be kept. In this criterion, the intensity signal of the blood inside the cavity indicates the reference of low intensity. Indeed, the cavity is a reliable reference tissue because its contour is given in the dataset and the signal in the cavity is relatively homogeneous.

However, the absolute difference between the infarction tissue and cavity is not always consistent between examination cases as shown in the left part of Fig. 3. The delayed acquisition time (the images are not acquired exactly at the same delay time between examinations), the patient's physical state, characteristics of MRI machine, etc. can affect the optimal threshold value of our criterion. To reduce the impact of such variations, the third criterion refers to both intensities of the cavities and non-pathological myocardium tissues, which constitute a scope of signal. The candidate infarction area is then compared to this scope:

$$(M(C) - M(cav))/(M(cav) - M(myo)) > \theta_{\%} \tag{3}$$

where $M(myo)$ refers to the median signal of non-infarction myocardium. The threshold value $\theta_{\%}$ becomes a percentage of the scope rather than an absolute difference (θ_{signal}) as in (2). Fig. 3 calculated on the training set prior information proves that the index of scope on the right is more consistent that the absolute difference on the left. Nevertheless, the contours of healthy myocardium tissues are not given for the infarction segmentation task (only the whole myocardium contour annotation is provided). Reminding that our neural network ensures high recall, we assume that all the myocardium except the regions segmented by the neural network is healthy, which means that the non-selected myocardium is certainly non-pathological. The final segmentation results testify the relevance of this hypothesis.

According to the statistics of the prior information and experiments, our framework employs the first and third criteria to select true-positive segmentations. The final post-processing is summed-up in Algorithm 1. All threshold values are roughly determined on the training set's prior information 3, and then fine-tuned on the validation set to obtain the best performance on the test set and avoid data leakage.

IV. Experiments and Results

To validate if the neighbouring image information can be correctly interpreted by the CNN to improve the segmentation accuracy, the 2.5D backbone and 2D

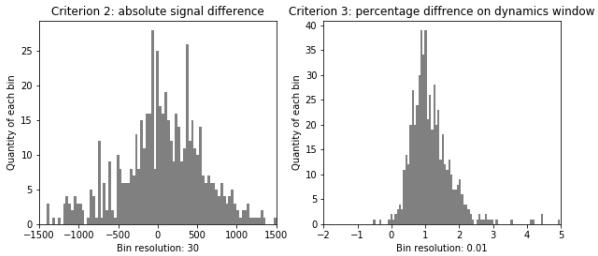

Fig. 3. Prior statistics supporting criterion (2) and criterion (3). The absolute difference of signal (left) is less consistent than the proportional difference on scope (right).

Algorithm 1: Post-processing

Result: removal of false-positive segmentation contour(s) of one image
Input : Set of infarction contours $\{S_i\}$
Input : Criterion (1) C_1; Criterion (3) C_3
Output: Final segmentation contour(s) S_{final}
$S_{ensemble} \leftarrow Avg(\{S_i)\}$;
$S_{ensemble} \leftarrow Binary(S_{ensemble})$;
for $aContour \in S_{ensemble}$ **do**
  if $aContour$ *fulfills conditions* $\{C_1 \cup C_3\}$ **then**
    $S_{final} \leftarrow S_{final} + aContour$
  end
end
return S_{final}

backbone models are compared. The same parameters are used, except for the dimension of the input images and the 1D convolution layer. As shown in Tab. I, the experiments show that no evident gain is obtained with the 2.5D input.

The original CE-Net architecture takes as input larger images than ours. Passing through 5 convolutional blocks, the feature map size is pooled as $\frac{1}{32}$ of the input image, which means that the DAC block will receive a 3×3 feature map if we feed the network with a 96×96 pixel² input image. To ensure the functionality of DAC and RMP blocks without reducing the depth of the network, in order to assess their benefit, we have tested CE-Net on interpolated images having a five times larger resolution than the original 96×96 ones. The nearest neighbour interpolation guarantees unchanged semantic information. The obtained results indicate two facts. First, CE-Net does not bring us a better result. The additional dense convolution and multi-kernel pooling on high-level semantic information do not provide useful information to the final segmentation. Second, interpolated images slightly improve the performance after post-processing. Indeed, interpolation equivalently changes the size of the receptive field of convolution kernels.

Most of the healthy images are wrongly segmented as false-positive by the CNN. The best CNN we tested obtains 80.92% on accuracy and 100% on recall (2D interpolated inputs, CE-Net), which means that all pathological images are segmented, whereas without the post-treatment 19.08% of the healthy test images are segmented as infarctions. The prior information-based post-processing obtains significant improvement on the segmentation results, especially on healthy images. After post-processing, the segmentation remains false on 8.25% of the images, including 5.81% false negative and 2.44% false-positive. On pathological images, the segmentation of infarction is also slightly improved. This improvement majorly comes from the removing of false-positive contours from noisy areas as shown in the base image in Fig 4. Notice that the threshold values of the criteria can affect the post-treatment results: more strict thresholds decrease the rate of false-positive segmentations, but increase the false-negative and vice versa. The threshold values we used are fine-tuned on the validation set by inspecting the Dice values and the difference of proportion. Two metrics that can increase or decrease differently depending on the threshold values.

Fig. 4 shows a segmentation result of an MRI examination. The result is globally satisfying in middle images. Tab. I gives the results for each CNN model and the effect of the post-processing on the test set. The metric Difference is calculated as the absolute area difference between the segmentation and the label divided by the area of the myocardium. Healthy images that are correctly segmented return 100% Dice and 0% Difference. Values in bracket stand for the results before post-processing. Hausdorff distance is not taken as our metric because in a single image multiple contours of infarction could be present. Comparing the Hausdorff distance between 2 series of contours is meaningless. The metrics obtained before and after the post-processing stage indicate that the proposed post-processing efficiently removes most wrong segmentation contours. Tab. II shows the final results according to the categories of the slice position. Apex and base images result in more failed segmentation because of their higher uncertainty and less agreement between clinical experts. Focusing on middle images, the results are quite satisfying compared to the annotation's internal variations.

To summarize, 2.5D input images can not be correctly

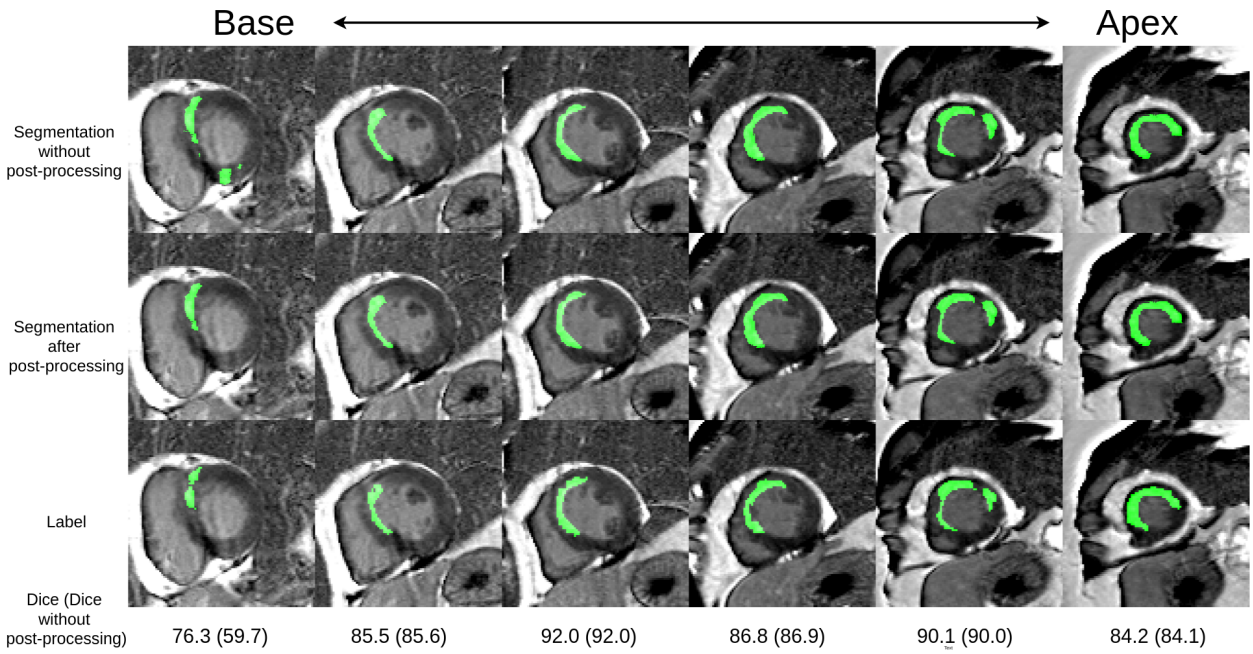

Fig. 4. Segmentation result from the 2D backbone network. From left to right, a whole examination is segmented (base to apex). From top to bottom, the contours are respectively the merged segmentations (the set of predictions, without the post-processing), post-treated segmentations, and their labels. Dice values are given at the bottom of each slice (Dice value without post-processing in bracket). The artifact due to the MRI acquisition on the first image is wrongly segmented as an infarction by CNN, then the post-processing successfully removes this false-positive segmentation so that its Dice improves from 59.7 to 76.3.

TABLE I
EXPERIMENTS RESULTS ON TEST SET

Model[1]	Dice(%)	Difference(%)	Recall(%)	Accuracy(%)
Intra-observation	81.00	4.14	85.45	71.20
Inter-observation	77.42	8.66	90.26	67.50
2D BB B1	74.19 (66.29)	7.72 (8.29)	83.20 (85.62)	65.00 (56.28)
2D BB B32	71.81 (57.88)	8.87 (11.65)	77.41 (81.69)	61.32 (46.61)
2.5D BB B1	75.71 (69.01)	7.70 (7.83)	84.72 (86.80)	66.41 (59.15)
2.5D BB B32	71.05 (59.09)	8.09 (11.22)	75.94 (81.20)	60.29 (42.62)
2D BB 5x B1	77.98 (65.2)	5.33 (7.93)	88.93 (90.34)	68.39 (55.97)
2D BB 5x B32	73.12 (59.51)	7.25 (11.98)	79.40 (85.97)	63.34 (48.73)
2D CE 5x B1	77.93 (65.67)	5.47 (7.64)	89.13 (90.56)	68.07 (56.38)
2D CE 5x B32	73.18 (59.93)	7.39 (11.19)	78.86 (84.96)	63.57 (49.17)

TABLE II
RESULTS ON DIFFERENT IMAGE POSITIONS

	Middle		Apex + Base	
	Dice(%)	Difference(%)	Dice(%)	Difference(%)
Proposal	81.37 (69.70)	4.29 (6.03)	65.63 (48.86)	9.21 (14.36)
Intra-observation	81.96	3.89	76.72	6.19

interpreted by the CNN. SGD achieves the best results with all network structures and input image resolution. Interpolation slightly optimizes segmentation details. The prior information-based post-processing largely improves the final segmentation results by solving the problem of false-positive segmentation on healthy images; the proposed framework provides results comparable to clinical experts, especially on middle slices.

V. DISCUSSION

An automatic myocardial infarction segmentation framework from cardiac DE-MRI has been designed using the CE-Net architecture and prior-information-based post-treatments. We have shown that SGD is more suited than its widely adopted mini-batch variant to train the CNN. SGD can efficiently increase the recall of prediction on highly biased data, which is very favorable for post-processing. 2.5D input does not provide more correctly interpreted semantic information. One possible explication is the delocalization between neighbouring images. Although each myocardium is relocalized in the center of the input image, the evolution of their form along the z axis is uncertain, hence the 1D convolution could wrongly merge inter-channel information. Another possible reason is the amplified noise. Since noise in one single image will corrupt both 3 channels due to

[1]BB: Backbone; CE: CE-NET; 5x: interpolated; B1/B32: batch size

the inter-channel convolution. CE-Net shows no difference to its backbone version, which could mean that more feature extraction on high semantic information is not helpful for the small and irregular region to be segmented. The improvement observed with interpolated images reveals also the effect of the receptive field size.

The prior information-based post-processing provides a huge improvement, particularly on mixed healthy-infarcted data. To generalize this prior information-based post-processing on other kinds of medical imaging, first, two (or more) labeled or region-confirmed tissues would be chosen to create a dynamics window. Second, the region signal in candidate contour would be compared to a proportional threshold on this window. Only the conformed contours would be finally kept as the result of the segmentation. The representative value of the region could be the median of its histogram. The difficulty of this approach is the selection of characteristic tissues.

Compared to the intra-observation variations, the results demonstrate the possible application of the proposed approach in clinical practice. Moreover, in terms of Dice value, our automatic segmentation method has reached the level of a human expert. Let us also emphasize that the combination of a high recall neural network and prior information-based post-processing can be adopted by other medical image segmentation frameworks. Future work includes two objectives. First, false predictions must be further reduced when dealing with healthy and pathological images. Second, a more complete framework able to segment myocardium, pathological tissue (infarction and no-reflow areas), as well as peripheral pathological tissue should be developed.

REFERENCES

[1] J. Vogel-Claussen, C. Rochitte, K. Wu, I. Kamel, T. Foo, J. Lima, and D. Bluemke, "Delayed enhancement mr imaging: Utility in myocardial assessment1," *Radiographics*, vol. 26, pp. 795–810, 05 2006.

[2] L. C. Amado, B. L. Gerber, S. N. Gupta, D. W. Rettmann, G. Szarf, R. Schock, K. Nasir, D. L. Kraitchman, and J. A. Lima, "Accurate and objective infarct sizing by contrast-enhanced magnetic resonance imaging in a canine myocardial infarction model," *Journal of the American College of Cardiology*, vol. 44, no. 12, pp. 2383–2389, 12 2004.

[3] R. Kim, D. Fieno, T. Parrish, K. Harris, E. Chen, O. Simonetti, J. Bundy, J. Finn, F. Klocke, and R. Judd, "Relationship of mri delayed contrast enhancement to irreversible injury, infarct age, and contractile function," *Circulation*, vol. 100, no. 19, pp. 1992–2002, 11 1999.

[4] A. Lalande, V. Valindria, M. Angue, N. Vignon, A. Cochet, and F. Brunotte, "Automatic evaluation of the peri-infarct area of myocardial infarction from delayed enhancement mri," in *European Society for Magnetic Resonance in Medicine and Biology (ESMRMB), Lisbonne, Portugal*, 10 2012.

[5] R. Karim *et al.*, "Evaluation of state-of-the-art segmentation algorithms for left ventricle infarct from late gadolinium enhancement mr images," *Medical Image Analysis*, vol. 30, pp. 95 – 107, 2016.

[6] E. de la Rosa, D. Sidibé, T. Decourselle, T. Leclercq, A. Cochet, and A. Lalande, "Myocardial infarction quantification from late gadolinium enhancement mri using top-hat transforms and neural networks," *Computers methods and programs in biomedicine [Under review]*, 2019, unpublished. [Online]. Available: https://arxiv.org/abs/1901.02911

[7] M. Avendi, A. Kheradvar, and H. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac mri," *Medical Image Analysis*, vol. 30, pp. 108 – 119, 2016.

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 3431–3440.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," vol. 9351, 10 2015, pp. 234–241.

[10] R. Mehta and J. Sivaswamy, "M-net: A convolutional neural network for deep brain structure segmentation," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, April 2017, pp. 437–440.

[11] T. M. Quan, D. G. C. Hildebrand, and W. Jeong, "Fusionnet: A deep fully residual convolutional neural network for image segmentation in connectomics," *CoRR*, vol. abs/1612.05360, 2016. [Online]. Available: http://arxiv.org/abs/1612.05360

[12] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2281–2292, Oct 2019.

[13] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support - DLMIA*, 2017, pp. 240–248. [Online]. Available: https://doi.org/10.1007/978-3-319-67558-9_28

[14] S. Taghanaki, Y. Zheng, S. K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, and G. Hamarneh, "Combo loss: Handling input and output imbalance in multi-organ segmentation," *Computerized Medical Imaging and Graphics*, vol. 75, 05 2019.

[15] M. Majurski, P. Manescu, S. Padi, N. Schaub, N. Hotaling, C. Simon Jr, and P. Bajcsy, "Cell image segmentation using generative adversarial networks, transfer learning, and augmentations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*.

[16] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, "Segan: Adversarial network with multi-scale L1 loss for medical image segmentation," *Neuroinformatics*, vol. 16, no. 3-4, pp. 383–392, 2018.

[17] D. Opitz and R. Maclin, "Popular ensemble methods: An empirical study," *Journal of artificial intelligence research*, vol. 11, pp. 169–198, 1999.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[19] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[20] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2016.