Convolutional Neural Network With Shape Prior Applied to Cardiac MRI Segmentation

Clement Zotti, Zhiming Luo, Alain Lalande, and Pierre-Marc Jodoin

***Abstract*—In this paper, we present a novel convolutional neural network architecture to segment images from a series of short-axis cardiac magnetic resonance slices (CMRI). The proposed model is an extension of the U-net that embeds a cardiac shape prior and involves a loss function tailored to the cardiac anatomy. Since the shape prior is computed offline only once, the execution of our model is not limited by its calculation. Our system takes as input raw magnetic resonance images, requires no manual preprocessing or image cropping and is trained to segment the endocardium and epicardium of the left ventricle, the endocardium of the right ventricle, as well as the center of the left ventricle. With its multiresolution grid architecture, the network learns both high and low-level features useful to register the shape prior as well as accurately localize the borders of the cardiac regions. Experimental results obtained on the Automatic Cardiac Diagnostic Challenge - Medical Image Computing and Computer Assisted Intervention (ACDC-MICCAI) 2017 dataset show that our model segments multislices CMRI (left and right ventricle contours) in 0.18 s with an average Dice coefficient of 0.91 and an average 3-D Hausdorff distance of 9.5 mm.**

***Index Terms*—Cardiac MRI segmentation, convolutional neural networks, shape prior.**

I. Introduction

NOWADAYS, cardiovascular diseases are the number one cause of death in the world with an estimation in 2015 of 31% of all deaths [1]. Cardiac MRI is a non-invasive imaging modality commonly used by physicians to detect and monitor cardiovascular diseases [2], [3]. One important advantage of MRI its high signal-to-noise ratio and good structural contrast between the blood, the myocardium and the surrounding organs. The use of kinetic MR images along the short axis orientation of the heart allows accurate evaluation of the function of the left and right ventricles [4], [5].

From kinetic images, several important parameters can be extracted to evaluate the cardiac anatomy and function, such as the volumes of the cavities in diastole and systole (and thus the ejection fraction), as well as the myocardial mass and the wall thickness and thickening [4]–[6]. The only reliable way to compute these values is through the delineation of the left ventricular endocardium (LV), the left ventricular epicardium (or myocardium - MYO) and the right ventricular endocardium (RV). As of today, the clinical use of cardiovascular MRI is hampered by the amount of data to be processed (often more than 10 short-axis slices and more than 20 phases per slice). Since the manual delineation of all images for all slices is clinically impracticable, several semi-automatic methods have been proposed, most of which being based on active contours, dynamic programming, graph cut or atlas fitting strategies [6]–[10].

So far, a limited number of fully-automatic cardiac segmentation methods have been proposed [7], [11], [12]. However, in the wake of the deep learning revolution, researchers started to apply deep convolution neural networks (CNN) to cardiac imagery [13]. Among the best CNN cardiac segmentation models are those involving a series of convolutions and pooling layers followed by one [14] or several [15], [16] deconvolution layers. The recent MICCAI 2017 ACDC challenge [17], [18] revealed that out of the 10 participants who proposed a fully-automatic segmentation method, 9 implemented a CNN and 8 implemented a variant of the U-Net of Ronneberger *et al.* [19].

In this paper we propose a CNN method specifically designed to segment the LV, RV and MYO without a third party segmentation method nor any manual intervention. Our approach incorporates a shape prior whose registration on the input image is learned by the model. Since the shape prior is computed only once, the execution of our model is not limited by its calculation. As such, our method is twice as fast as the winner [20] of the MICCAI 2017 ACDC challenge [17], [18] and barely slower than a plain U-Net [19], as shown in Table V. Our method implements a multi-resolution "grid" structure to guide the model at each resolution and, as such, allows to accurately register the shape prior and segment the multi-slices CMRI.

II. Previous Works

Most multi-slice cardiac MRI segmentation methods are semi-automatic. As mentioned in the survey papers of Petitjean *et al.* [7], [10], most of semi-automatic segmentation methods

Manuscript received December 22, 2017; revised May 14, 2018 and July 5, 2018; accepted August 5, 2018. Date of publication August 14, 2018; date of current version May 6, 2019. This research was funded by the Conseil franco-québécois de coopération universitaire (CFQCU), Nb:2015-FQ-186778. *(Corresponding author: Clement Zotti.)*

C. Zotti and P.-M. Jodoin are with the Department of Computer Science, University of Sherbrooke, Sherbrooke, QC J1K 2R1, Canada (e-mail: clement.zotti@usherbrooke.ca; pierre-marc.jodoin@usherbrooke.ca).

A. Lalande is with the Le2i laboratory, CNRS FRE 2005, University of Burgundy, Dijon 21079, France, and also with the MRI Department, University Hospital of Dijon, Dijon 21079, France (e-mail: alain.lalande@u-bourgogne.fr).

Z. Luo is with the Department of Cognitive Science, Xiamen University, Xiamen 361005, China, and also with the Department of Computer Science, University of Sherbrooke, Sherbrooke, QC J1K 2R1, Canada (e-mail: zhiming.luo@usherbrooke.ca).

Digital Object Identifier 10.1109/JBHI.2018.2865450

2168-2194 © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Authorized licensed use limited to: University of Electronic Science and Tech of China. Downloaded on February 06,2025 at 14:58:24 UTC from IEEE Xplore. Restrictions apply.

are based on active contours, dynamic programming, graph cut or some atlas fitting strategies [6], [8], [9], [21]. Unfortunately, these methods are far from real time due to the manual interaction they require. Also, most of it are ill-suited for segmenting simultaneously the LV, the RV, and the MYO as their model is tailored to one cardiac region.

As of today, a handful of fully-automatic cardiac segmentation methods have been proposed. Several approaches implemented an atlas registration. For example, Zuluaga *et al.* and Ou *et al.* [22], [23] used a multi-atlas framework where a target image is registered to the intensity images of a training database. The labeled images from the atlases are then combined throughout a majority vote. Although accurate, registration methods are known for being very slow. Other authors proposed a cascade of basic low-level image processing methods [11], [24] involving various edge detection filters, optimal thresholding techniques, morphological operators and post-processing methods. These methods often have a strong heuristic flavor and do not generalize well to images that slightly violate their initial assumptions. Level-set methods have also been proposed such as that of Liu *et al.* [12] which segments both the LV and the RV and that of Queiros *et al.* [25] which is both fast and accurate. However, while the level-set segmentation methods are fully-automatic, their initialization is usually not. Also, like the semi-automatic methods, these approaches are often designed to segment one specific cardiac region and, as such, do not generalize well to other cardiac regions.

Shape priors have been widely used by non-machine-learning MRI cardiac segmentation methods such as active contours [26], [27], feature-fusion methods [28], and active shape and appearance model (ASM/AAM) approaches [29]. Woo *et al.* [26] proposed an implicit dual shape prior to model the relationship between two circular regions of endocardium and epicardium. Wu *et al.* [27] integrated a circular prior into the snake model to extract the endocardium and then used it as a prior for the epicardium. Bai *et al.* [28] proposed a multi-atlas segmentation method using intensity, gradient and contextual features. In [29], the authors used AAM to optimize a global appearance model and ASM for finding local structures. They also introduced a multistage hybrid model. An early AAM stage was used to determine the pose and general shape, then a hybrid ASM/AAM stage was followed to refine the positions, and a final AAM stage was invoked for escaping a possible local minimum during previous ASM/AAM stage.

As far as we know, no deep learning method was applied to cardiac MRI segmentation before 2014. The first extensive use of deep learning on cardiac data occurred in 2015 during the Kaggle Second Annual Data Science Bowl challenge.[1] Unfortunately, since that challenge provided no proceedings, the specifics of these methods was not published.

Since then, people started using deep learning in conjunction with none-deep-learning segmentation methods. Let us mention the work by Ngo *et al.* [30] which used a deep belief network to initialize and guide a level-set segmentation model of the LV. Avendi *et al.* [31] used a CNN and an autoencoder to initialize a level-set method. In the same line of thoughts, Rupprecht *et al.* [32] combined a patch-based CNN with an active contour method. They used a CNN to predict a vector field which, in turn, is used to "push" on the contour. Yang *et al.* [33] proposed a multi-atlas segmentation method whose label fusion and feature extraction is done by a deep learning method.

Other authors used a standalone CNN to segment cardiac structures from CMRI data. Tran *et al.* [13] applied the fully convolutional neural network of Long *et al.* [34] to segment the LV and the RV. Poudel *et al.* [35] proposed a U-Net similar to that of Ronneberger *et al.* [19] but with a gated recurrent unit (GRU) on the high-level features. The GRU is used to propagate feature information from one slice to another. Finally, Oktay *et al.* [36] proposed a super-resolution approach based on a residual convolutional neural network to reconstruct high resolution 3D volumes from 2D image stacks for more accurate image analysis.

Recently, machine learning methods [37]–[39] were used to predict high level LV/RV informations (like volume and ejection fraction) without the need to segment images. Gu *et al.* [37] proposed a sparse regression with an output correlation method to estimate the ejection fraction. Xue *et al.* [38] developed a network called Indices-Net which can directly estimate the LV area, the myocardium area and the 6 regional wall thicknesses. Indices-Net contains a deep convolution autoencoder for image representation and a CNN for regression. Xue *et al.* [39] further improved the model which can simultaneously estimate LV indices and the cardiac phase by using a CNN for image representation and two recurrent neural networks for modeling the temporal dynamics of cardiac sequences. Let us mention that although one can quantify the parameters of the LV/RV without localizing the contours, since cardiac diagnostics are based of well-known physiological parameters, physicist usually need contour maps to assess that cardiac quantification values come from actual physiological measurements.

In the wake of the 2017 ACDC MICCAI challenge [17], [18], 10 fully-automatic segmentation methods were proposed, 9 of which implemented a CNN. Most of the top performing methods implemented a modified version of the U-Net [19]. The method that ended up winning the ACDC segmentation challenge is that of Isensee *et al.* [20] which used an ensemble of 2D and 3D U-nets. In second position was Khened *et al.* [40] which proposed a 2D version of the U-net with dense and residual connections [41] instead of skip connexions. Baumgartner *et al.* [42] tried multiple architectures of 2D and 3D U-net with crossentropy and dice losses. Their best implementation is a 2D U-net with a crossentropy loss which ended up in third place, on par with Zotti *et al.* [43] and their grid net architecture that will be described in the next section. Another architecture proposed by Jang *et al.* [44] is the M-net [45] which is identical to the U-net but with multiple inputs and outputs and trained with a weighted crossentropy loss.

III. Our Method

The goal of our method is to segment the LV, the RV and the MYO from a 3D $N \times M \times H$ raw MR image volume X created from a stack of N 2D short-axis slices with M lines and H columns. This is done by predicting a 3D label map T also of

[1] www.kaggle.com/c/second-annual-data-science-bowl

Authorized licensed use limited to: University of Electronic Science and Tech of China. Downloaded on February 06,2025 at 14:58:24 UTC from IEEE Xplore. Restrictions apply.

size $N \times M \times H$ and whose voxels $\vec{v} = (i, j, k)$ contain a label $T_{\vec{v}} \in \{\text{Back, LV, RV, MYO}\}$, where "Back" stands for tissues different than the other three. Here we assume that the input image X has been acquired following the common clinical SSFP cine-MRI sequence with a series of short-axis slices starting from the mitral valves down to the apex of the left ventricle [46].

In this section, we give the probabilistic framework for our model that is an extension of the method proposed by Zotti *et al.* [43]. We then describe the core architecture of our *Grid-net* model, and then show how a shape prior, a multiresolution structure and a Dice loss can be integrated to it.

A. Probabilistic Formulation

Our training set contains N multi-slice images $\mathbf{X} = \{X_1, \ldots, X_N\}$ with their associated label fields $\mathbf{T} = \{T_1, \ldots, T_N\}$. In order to improve the segmentation accuracy near the border of the LV, RV and MYO, the training set also includes the contour map of each region $\mathbf{C} = \{C_1, \ldots, C_N\}$ where $C_{n,\vec{v}} = T_{n,\vec{v}}$ for every voxel $\vec{v}$ located on the border of the LV, RV, or MYO, and $C_{n,\vec{v}} = 0$ otherwise.

The best parameters $\mathbf{w}$ (*i.e.* the weights of our CNN) are those that maximize the *a posteriori* formulation

$$\mathbf{w} = \arg\max_{\mathbf{w}} P(\mathbf{w}|\mathbf{X}, \mathbf{T}, \mathbf{C}) \tag{1}$$

which can be reformulated according to the Bayes theorem as

$$\mathbf{w} = \arg\max_{\mathbf{w}} \frac{P(\mathbf{X}, \mathbf{T}, \mathbf{C}|\mathbf{w})P(\mathbf{w})}{P(\mathbf{X}, \mathbf{T}, \mathbf{C})}. \tag{2}$$

Since $P(\mathbf{X}, \mathbf{T}, \mathbf{C})$ is independent of $\mathbf{w}$ and following the principles of the joint and conditional probabilities, we can reformulate the optimization criterion as:

$$\mathbf{w} = \arg\max_{\mathbf{w}} P(\mathbf{C}|\mathbf{X}, \mathbf{T}, \mathbf{w})P(\mathbf{T}|\mathbf{X}, \mathbf{w})P(\mathbf{X}|\mathbf{w})P(\mathbf{w}).$$

Since the input image $\mathbf{X}$ is fixed, we can assume that $P(\mathbf{X}|\mathbf{w})$ is constant and remove it from the equation

$$\mathbf{w} = \arg\max_{\mathbf{w}} P(\mathbf{C}|\mathbf{X}, \mathbf{T}, \mathbf{w})P(\mathbf{T}|\mathbf{X}, \mathbf{w})P(\mathbf{w}). \tag{3}$$

If we assume that the random variables of X, C and T are spatially independent (*i.e.* C_i vs C_j, T_i vs T_j and X_i vs X_j are independent when $i \neq j$) and that $P(\mathbf{w})$ follows a zero-centered Gaussian $\mathcal{N}(\mathbf{w}; 0)$, we get that the best $\mathbf{w}$ is the one that maximizes the following loss:

$$\mathcal{L} = \prod_i P(C_i|X_i, T_i, \mathbf{w})P(T_i|X_i, \mathbf{w})\mathcal{N}(\mathbf{w}; 0) \tag{4}$$

or, equivalently, minimizes the following log of the loss:

$$\mathcal{L} = \sum_i \underbrace{-\ln P(C_i|X_i, T_i, \mathbf{w})}_{\mathcal{L}_C} \underbrace{-\ln P(T_i|X_i, \mathbf{w})}_{\mathcal{L}_T} + \underbrace{\gamma_w ||\vec{w}||^2}_{\mathcal{L}_w} \tag{5}$$

where $\mathcal{L}_T$ and $\mathcal{L}_C$ are the log-likelihood of the predicted labels and predicted contours and $\mathcal{L}_w$ is the prior loss. $\mathcal{L}_T$ implements the following cross-entropy

$$\mathcal{L}_T = -\ln P(T_i|X_i, \mathbf{w}) = -\gamma_T \sum_{l=1}^{4} \sum_{\vec{v}} T_{i,l,\vec{v}} \ln \hat{T}_{l,\vec{v}}$$

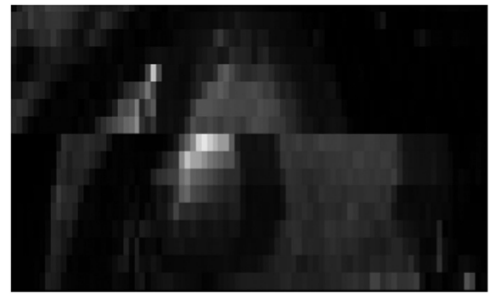


Fig. 1. Multi-slices reconstruction with horizontal shift along the short axis plane due to different breath hold during acquisition.

where l is the class index,i the slice index, $\vec{v}$ is a pixel location, γ_T is a constant, $T_{i,l,\vec{v}}$ is the true probability that pixel $\vec{v}$ is in class l, and $\hat{T}_{i,l,\vec{v}}$ is the output of our model for pixel $\vec{v}$ and class l. The contour loss is also a cross-entropy

$$\mathcal{L}_C = -\ln P(C_i|X_i, T_i, \mathbf{w}) = -\gamma_C \sum_{l=1}^{4} \sum_{\vec{v}} C_{i,l,\vec{v}} \ln \hat{C}_{i,l,\vec{v}}$$

where C_i and $\hat{C}_i$ are contours extracted from T_i and $\hat{T}_i$, and γ_C is a constant. The contours are extracted with a Sobel filter on the ground truth and the prediction to compute this loss. Note that having two pixel-wise losses $\mathcal{L}_C$ and $\mathcal{L}_T$ impose a larger penalty on erroneous contour pixels (which affects both $\mathcal{L}_C$ and $\mathcal{L}_T$) than any other pixel (which only affects $\mathcal{L}_T$). This increases the gradient steps towards solutions with accurate contours, i.e. solutions for which the heart is well delineated. Such contour loss has also been used successfully for saliency detection [47].

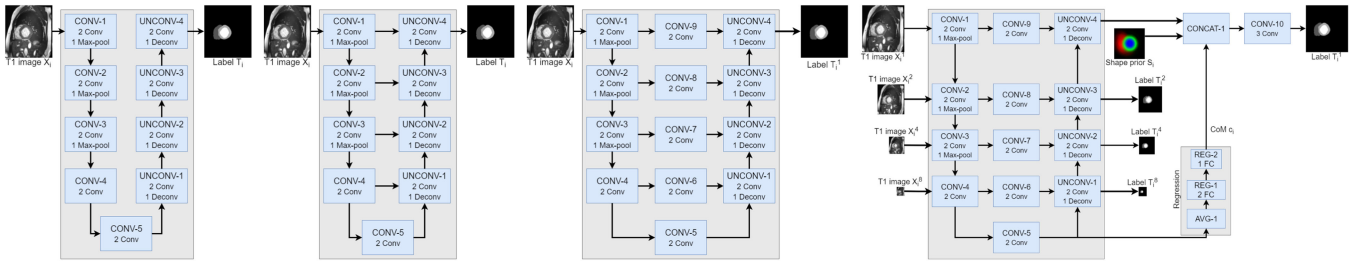
B. Core Architecture

The core of our model is the "grid-net" architecture illustrated in Fig. 2(c). This architecture can be seen as an extension of other well-known encoder-decoder networks such as the ConvDeconv by Noh *et al.* [15], the FCNet [34], and the U-Net [19]. One important advantage of these networks is their ability to extract features that account for both the global and the local context. For CMRI, global features help to differentiate the heart from the surrounding organs while the local features ensure accurate segmentation.

Our grid-net CNN has 3 columns and 5 rows (c.f. Fig. 2(d)). Each row extracts features specific to a given scale. The input of our model (upper left) is a 256×256 MR image X_i while the output is the label field T_i (top right) also of size 256×256. Note that a common issue with MR cardiac images is the fact that in the 2D short-axis image, the location of the heart sometimes get shifted from one slice to another due to different breath-holds during successive acquisitions (c.f. Fig. 1). That shift as well as the large inter-slice gap (often 10 mm) make the use of 3D convolution operations ill-suited [48]. As a consequence, our network processes each 2D slices independently and reshape the output 3D volume T_i by stacking up the resulting 2D label fields.

As we get deeper in the network (from CONV-1 to CONV-5), the extracted features involve a larger context of the input image.

Authorized licensed use limited to: University of Electronic Science and Tech of China. Downloaded on February 06,2025 at 14:58:24 UTC from IEEE Xplore. Restrictions apply.



(a) ConvDeconv (b) U-Net (c) GridNet-core (d) Our proposed GridNet-MD architecture

Fig. 2. Four different CNN architectures. (a) ConvDeconv [34]. (b) U-Net [19]. (c) Gridnet (sec.III.B). (d) The proposed multiresolution gridnet.

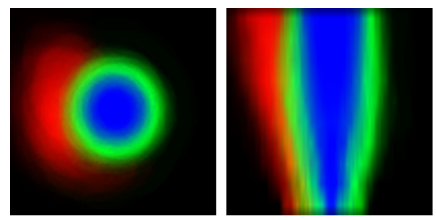
TABLE I
DETAILS OF THE *GridNet_core* ARCHITECTURE

Block	Layer	Kernel	Stride	Output
CONV-1	2 conv + max-pool	3x3 2x2	1 2	256x256x32 128x128x32
CONV-2	2 conv + max-pool	3x3 2x2	1 2	128x128x64 64x64x64
CONV-3	2 conv + max-pool	3x3 2x2	1 2	64x64x128 32x32x128
CONV-4	2 conv + max-pool	3x3 2x2	1 2	32x32x256 16x16x256
CONV-5	2 conv	3x3	1	16x16x512
CONV-6	2 conv	3x3	1	32x32x256
CONV-7	2 conv	3x3	1	64x64x128
CONV-8	2 conv	3x3	1	128x128x64
CONV-9	2 conv	3x3	1	256x256x32
UNCONV-1	deconv + 2 conv	2x2 3x3	 1	32x32x768 32x32x256
UNCONV-2	deconv + 2 conv	2x2 3x3	 1	64x64x384 64x64x128
UNCONV-3	deconv + 2 conv	 3x3	 1	128x128x192 128x128x64
UNCONV-4	deconv + conv + conv	2x2 3x3 1x1	 1 1	256x256x96 256x256x32 256x256x4

The second column contains 4 convolution layers (all without max-pooling) used to compute features at various resolutions. The last column aggregates features from the lowest to the highest resolution. The UNCONV-4 layer contains both global and local features which we use to segment the image. Note that this grid structure is similar to the U-Net except for the middle CONV-6 to 9 layers. Each conv layer has a 3×3 receptive field (detailed architecture in Table I) and its feature maps have the same size than their input (zero padding). We also batch normalize each feature map, use the ReLU activation function, and dropout [49] to prevent overfitting. For the remaining of this paper, we will refer to this network as the *GridNet_core*.

C. Shape Prior

In order to prevent the model from generating anatomically impossible segmentation, we added a shape prior to our CNN. The shape prior S is a 3D volume which encodes the probability of a 3D location $\vec{v} = (i, j, k)$ to be a member of a certain class (Back, LV, RV, or MYO). We estimate this probability by computing the pixel-wise empirical proportion of each class based on the groundtruth label fields T_i of the training dataset:



(a) short axis (b) long axis

Fig. 3. Our shape prior probability map.

$$P(\mathcal{C}|\vec{v}) = \frac{1}{N_t} \sum_{i=1}^{N_t} \mathbb{1}_{\mathcal{C}}\left(T_{i,\vec{v}}\right)$$

where $\mathbb{1}_{\mathcal{C}}\left(T_{i,\vec{v}}\right)$ is an indicator function which returns 1 when $T_{i,\vec{v}} = \mathcal{C}$ and 0 otherwise, and N_t is the total number of training images (c.f. Fig. 3).

These probabilities are put into a $15 \times 100 \times 100 \times 3$ volume S where 3 stands for the 3 classes (RV, MYO, LV)[2], 15 stands for the number of interpolated slices (from the base to the apex) and 100×100 is the inplane size of the shape prior S. Note that before to compute S, we registered the training label fields T_i into a common space and cropped a 100×100 region around that center. Since the size and the orientation of the heart does not vary much from one patient to another, registration was done by aligning the cardiac center of mass (CoM) $\vec{c}_i$ of each training label field X_i. The CoM is computed based on the location of the pericardium (obtained from MYO and RV) in each slice.

The training set is now augmented with the left ventricle CoM $\mathbf{c} = \{\vec{c}_1, \ldots, \vec{c}_N\}$, as well as a precomputed shape prior S. We can change Eq. (1) as follows:

$$\mathbf{w} = \arg\max_{\mathbf{w}} P(\mathbf{w}|\mathbf{X}, \mathbf{T}, \mathbf{c}, \mathbf{C}, S) \tag{6}$$

where the $\mathbf{c}$ and $\mathbf{S}$ are the new included CoM and shape prior. We can now derive the new loss with respect to the new input and get:

$$\mathbf{w} = \arg\max_{\mathbf{w}} P(\mathbf{C}|\mathbf{X}, \mathbf{T}, \mathbf{w}) P(\mathbf{T}|\mathbf{X}, S, \mathbf{c}, \mathbf{w}) P(\mathbf{c}|\mathbf{X}, \mathbf{w}) P(\mathbf{w}).$$

[2]No need to store the probability of "Back" since the 4 probabilities sum up to 1.

Authorized licensed use limited to: University of Electronic Science and Tech of China. Downloaded on February 06,2025 at 14:58:24 UTC from IEEE Xplore. Restrictions apply.

TABLE II
EXTRA LAYERS OF OUR GridNet-MD MODEL

Block	Layer	Kernel	Stride	Output
AVG-1	Average pooling			512
REG-1	1 fully connected + 1 fully connected +			256 128
REG-2	1 fully connected			2
CONCAT-1	align prior + concat			256x256x7
CONV-10	3 conv + 1 conv	3x3 1x1	1 1	256x256x32 256x256x4

If we assume that $P(\vec{c}_i|X_i, \mathbf{w})$ follows a Gaussian distribution centered on the true CoM $\hat{c}_i$, the resulting loss becomes:

$$\mathcal{L} = \sum_i \underbrace{-\ln P(C_i|X_i, T_i, \mathbf{w})}_{\mathcal{L}_C} \underbrace{-\ln P(T_i|X_i, S, c_i, \mathbf{w})}_{\mathcal{L}_T} + \underbrace{\gamma_c ||\vec{c}_{i,\mathbf{w}} - \hat{c}_i||^2}_{\mathcal{L}_c} + \underbrace{\gamma_w ||\vec{w}||^2}_{\mathcal{L}_w}. \quad (7)$$

Since at test time $\vec{c}$ is not provided with the input image, our CNN comes with a regression module designed to predict the CoM of the input image X as shown in Fig. 2(d).

D. Multi Resolution

To enforce a better accuracy on the borders of the cardiac regions, we added an output at each resolution of our network (c.f. t_i^2, t_i^4, t_i^8 in Fig. 2(d)). These outputs are added to the loss according and compared to rescaled versions of the full-scale groundtruth map T_i^1. The resulting optimization loss is

$$\mathcal{L}_M = \sum_i \underbrace{-\ln P(C_i|X_i, T_i^1, \mathbf{w})}_{\mathcal{L}_C} \underbrace{-\ln P(T_i^1|X_i, S, c_i, \mathbf{w})}_{\mathcal{L}_{T\,1}} - \sum_{s\in\{2,4,8\}} \underbrace{(\ln P(T_i^s|X_i, c_i, \mathbf{w}))}_{\mathcal{L}_{T\,s}} + \underbrace{\gamma_c ||c_{i,\mathbf{w}} - \hat{c}_i||^2}_{\mathcal{L}_c} + \underbrace{\gamma_w ||\vec{w}||^2}_{\mathcal{L}_w}$$

where T_i^s is the groundtruth for class i and scale s. The Table II give a detailed description of the new layers added to the *GridNet_core*.

E. Dice Coefficient

A common issue in medical image segmentation is the fact that classes are often imbalanced. In our case, the number of pixels associated to the background is up to 100 times larger than for the other classes altogether. One can compensate for this issue by adding a loss term that is not affected by the unbalanced nature of the classes. Empirical validation reveal that the Dice loss [50] is an effective solution. The multi-class Dice loss is

$$\mathcal{D}(T, \hat{T}) = 1 - \frac{1}{\sum_i \lambda_i} \left[\sum_i \lambda_i \frac{2 \sum_n T_{in} \hat{T}_{in}}{\sum_n T_{in}^2 + \hat{T}_{in}^2} \right] \quad (8)$$

TABLE III
RANKING* ON THE TEST SET BETWEEN METHODS FOLLOWING THE ONLINE CHALLENGE RANK COMPUTATION. THESE METHODS WERE TRAINED WITH THE SAME OPTIMIZER, THE SAME DATA AUGMENTATION, AND THE SAME HYPERPARAMETERS GRID SEARCH

Methods	LV				RV				MYO			
	Dice		HD		Dice		HD		Dice		HD	
	ED	ES	ED	ES	ED	ES	ED	ES	ED	ES	ED	ES
GridNet-MD	**0.96**	**0.91**	6.2	**8.4**	**0.94**	**0.89**	11.1	12.7	**0.89**	**0.90**	9.6	**9.2**
GridNet-M	**0.96**	**0.91**	**6.0**	8.8	**0.94**	0.88	**10.2**	**12.2**	**0.89**	**0.90**	**8.3**	9.7
GridNet [43]	**0.96**	**0.91**	6.7	8.7	**0.94**	0.88	10.3	14.1	0.88	**0.90**	8.7	**9.2**
GridNet-core	0.95	0.88	7.2	9.9	0.91	0.85	12.4	13.3	0.87	0.89	8.8	9.6
U-Net	0.96	0.90	6.7	9.9	0.89	0.83	13.6	15.1	0.83	0.85	10.7	12.8
ConvDeconv	0.94	0.88	7.5	10.4	0.87	0.81	13.7	15.1	0.80	0.83	11.7	13.2

* *Although clinical metrics not shown, they were considered for the ranking.*

where i is a class index, λ_i is the weight associated to class i, n the number of pixels, T the groundtruth label map, and $\hat{T}$ is the predicted label map. The λ_i are computed given the number of pixels of each class. Since on average the background has 100 times more pixels, we set $\lambda_0 = 0.1$ and the other to $\lambda_i = 10$ where $i \in \{1, 2, 3\}$. The new loss now becomes,

$$\mathcal{L}_{MD} = \mathcal{L}_M + \sum_{s\in\{1,2,4,8\}} \mathcal{D}(T_i^s, \hat{T}_i^s) \quad (9)$$

where $\mathcal{D}$ is the Dice loss defined in Eq. (8), and s is the scale index. Our final system implements the multi-resolution GridNet + shape prior shown in Fig. 2(d) that we train with the loss of Eq. (9). We refer to this method as *GridNet-MD*.

IV. EXPERIMENTAL SETUP AND RESULTS

A. Dataset

We trained and tested our system on the ACDC MICCAI 2017 dataset [17], [18]. This dataset is made of 150 exams from different patients acquired from two MRI scanners (1.5 T and 3.0 T, Siemens Medical Solution, Erlangen, Germany) over a 5 year period, 100 exams were used for training and 50 for testing. The exams are divided into 5 evenly distributed groups: dilated cardiomyopathy, hypertrophic cardiomyopathy, myocardial infarction with altered left ventricular ejection fraction, abnormal right ventricle and patients without cardiac disease. Cine MR images were acquired in breath-hold with a retrospective or prospective gating and with a SSFP sequence in short-axis orientation. More specifically, a series of short axis slices cover the heart from the base to the apex of the left ventricle, with a slice thickness of 5 to 8 mm and sometimes an interslice gap of 5 mm. The short-axis in-plane spatial resolution goes from 0.83 to 1.75 mm^2/pixel.

The reconstructed 3D images of the diastolic and systolic phases are provided with a pixel-accurate manual delineation by two independent medical experts of each cardiac region. The delineation of these regions for each slice follows well-defined rules (such as for example the inclusion of the papillary muscles into the cavity) widely accepted in cardiology [51].

The experts had to come to an agreement when their outline diverged by more than 1 pixel.

Authorized licensed use limited to: University of Electronic Science and Tech of China. Downloaded on February 06,2025 at 14:58:24 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
P-VALUE SIGNIFICANCE FROM A WILCOXON TEST ON THE ACDC VALIDATION SET BETWEEN GridNet-MD AND U-Net/ConvDeconv

Methods	LV				RV				MYO			
	Dice		HD		Dice		HD		Dice		HD	
	ED	ES	ED	ES	ED	ES	ED	ES	ED	ES	ED	ES
U-Net	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.13	<0.01
ConvDeconv	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.011	<0.01

B. Evaluation Criteria and Implementation Details

We first compare our proposed *GridNet-MD* method with four sub-architectures namely : the convDeconv CNN by Noh *et al.* [15], the U-Net by Ronneberger *et al.* [19], the GridNet-core (Fig. 2(c)), and the *GridNet-M* (the proposed method without the Dice loss). All theses models were trained with data augmentation involving rotation from -60 to $+60$ degrees from the center of the image, and a gamma correction with $\gamma \in [0.5, 1.5]$. We also used the Adam [52] optimizer with a learning rate of 10^{-4}, an early stopping based on the Dice coefficient and a grid search to set our hyper parameters (number of feature maps, number of neurons for the regression, with or without dropout, l2 regularizer coefficient). For the grid search, we trained the systems with 15 patients per group (75 patients in total) and validated on the remaining 25 patients. Our code[3] is based on the Keras library [53]. Note that we also trained and tested the U-Net and the ConvDeconv with their original structure (33M parameters) but got worse results.

We also compared our method to the ones reported on the ACDC MICCAI leaderboard [17]. As mentioned before, nine of those methods implemented a CNN, most of which a U-Net. The metrics used to compare the methods are the Dice score and the Hausdorff distances at the end systolic and end diastolic time points, as well as the correlation, the bias and the standard deviation for the ejection fraction, the end diastolic volume and the myocardial mass. The former metrics are so-called *clinical metrics*.

C. Experimental Results

Table III shows results obtained by our method *GridNet-MD* compared with five subversions including ConvDeconv [34] and U-Net [19]. The GridNet [43] results are those obtained by Zotti *et al.* [43] during the ACDC challenge. As one can see, the successive updates proposed in section III all contribute to improve results. In particular replacing the skip connections by convolution layers and adding a contour loss improve results (c.f. GridNet-core vs U-Net). Also, the multiresolution structure combined with the Dice loss increases further the accuracy of our model. GridNet-MD outperforms convDeconv and U-Net on every metric. While the Dice score of convDeconv and U-Net is close to our's for the LV, our method clearly outperforms it for the RV and MYO.

TABLE V
PERCENTAGE OF ANATOMICALLY PLAUSIBLE SLICES PRODUCES BY EACH MODEL ON THE ACDC VALIDATION SET, AND THE AVERAGE PROCESSING TIME PER 3D VOLUME ($\pm$STD)

Methods	slices (%)	time (s)
GridNet-MD	**94.2**	0.18 $\pm$0.2
U-Net	84.2	0.17 $\pm$ 0.2
ConvDeconv	75.6	0.15 $\pm$ 0.2

TABLE VI
GENERALIZATION RESULTS OF OUR METHOD ON THE ACDC TEST SET AND THE MICCAI2009 SUNNYBROOOKE DATASET

Datasets	LV			
	Dice		HD	
	ED	ES	ED	ES
ACDC (Test)	0.96	0.91	6.2	8.4
MICCAI 2009 (Training)	0.94	0.89	9.3	11.1

We ran a statistical test on these results to measure how significantly better our results are compared to those from U-Net and ConvDeconv. We implemented a Wilcoxon test [54] since the data do not follow a Gaussian distribution. The p-values are reported in Table IV. As can be seen, our method is significantly better than the other two (p-value below 0.01) on every metric except for the Hausdorff distance of MYO-ED.

We also observe that our method produces more anatomically plausible results while being barely slower than U-Net and ConvDeconv. In order to quantify this observation, we report the percentage of anatomically plausible slices and the processing time for each method in Table V. An anatomically implausible 2D segmentation result in our case is one for which a region has a hole in it, or when the RV is disconnected from the myocardium, or when the LV cavity is in contact with the background, or when a region has two (or more) disconnected regions, or when a model produces no segmentation results on a slice that do contain a section of LV, RV or MYO.

In order to illustrate the generalization power of our method, we tested it (without finetuning) on the training set of the Sunnybrooke 2009 MICCAI Cardiac MR Left Ventricle Segmentation Challenge.[4] As shown in Table VI our Dice and HD scores on the MICCAI 2009 dataset are very close to that of the ACDC test set which shows the good generalization power of our method. Let us also mention that since the HD is computed in 3D and

[3]https://bitbucket.org/vitalab/vitalabai_public/src/master/VITALabAI/model/acdc/

[4]http://smial.sri.utoronto.ca/LV_Challenge/Home.html

Authorized licensed use limited to: University of Electronic Science and Tech of China. Downloaded on February 06,2025 at 14:58:24 UTC from IEEE Xplore. Restrictions apply.

TABLE VII
LV TEST RESULTS FROM THE ACDC DATASET FOR OUR GridNet-MD METHOD AND ACDC CHALLENGERS

	Dice		Hausdorff		EF			Volume ED		
	ED	ES	ED	ES	corr.	bias	std-dev.	corr.	bias	std-dev.
Isensee *et al.* [20]	0.968	0.931	7.4	6.9	0.991	0.2	3.1	0.997	2.7	5.7
GridNet-MD	0.963	0.912	6.2	8.4	0.990	-0.5	3.1	0.997	3.7	5.1
Khened *et al.* [40]	0.964	0.917	8.1	9.0	0.989	-0.5	3.4	0.997	0.6	5.5
Jang *et al.* [44]	0.959	0.921	7.7	7.1	0.989	-0.3	3.3	0.993	-0.4	8.7
Baumgartner *et al.* [42]	0.963	0.911	6.5	9.2	0.988	0.6	3.4	0.995	1.4	7.6
Rohe *et al.* [57]	0.957	0.900	7.5	10.8	0.989	-0.1	3.2	0.993	4.2	8.6
Wolterink *et al.* [58]	0.961	0.918	7.5	9.6	0.988	-0.5	3.4	0.993	3.0	8.7
Zotti *et al.* [43]	0.957	0.905	6.6	8.7	0.987	-1.2	3.6	0.997	9.6	6.4
Jain *et al.* [59]	0.955	0.885	8.2	10.9	0.971	1.7	5.5	0.997	9.9	6.7
Tziritas-Grinias [60]	0.948	0.865	8.9	11.6	0.975	-1.6	5.0	0.992	2.0	11.7
Yang *et al.* [61]	0.864	0.775	47.9	53.1	0.926	1.5	8.1	0.894	12.2	32.0

TABLE VIII
RV TEST RESULTS FROM THE ACDC DATASET FOR OUR GridNet-MD METHOD AND ACDC CHALLENGERS

	Dice		Hausdorff		EF			Volume ED		
	ED	ES	ED	ES	corr.	bias	std-dev.	corr.	bias	std-dev.
Isensee *et al.* [20]	0.946	0.899	10.1	12.2	0.901	-2.7	6.2	0.988	4.4	10.8
GridNet-MD	0.934	0.885	11.1	12.7	0.869	-0.9	6.8	0.986	2.4	11.5
Zotti *et al.* [43]	0.941	0.882	10.3	14.1	0.872	-2.2	6.8	0.991	-3.7	9.2
Khened *et al.* [40]	0.935	0.879	14.0	13.9	0.858	-2.2	6.9	0.982	-2.9	12.6
Baumgartner *et al.* [42]	0.932	0.883	12.7	14.7	0.851	1.2	7.3	0.977	-2.3	15.1
Jang *et al.* [44]	0.929	0.885	12.9	11.8	0.793	-3.2	8.3	0.986	-10.8	11.6
Wolterink *et al.* [58]	0.928	0.872	11.9	13.4	0.852	-4.6	6.9	0.980	3.6	15.2
Roh *et al.* [57]	0.916	0.845	14.1	15.9	0.781	-0.7	9.9	0.983	7.3	13.4
Jain *et al.* [59]	0.911	0.819	13.5	18.7	0.791	6.8	8.1	0.945	5.6	22.2
Tziritas-Grinias [60]	0.863	0.743	21.0	25.7	0.758	-0.5	9.1	0.930	18.6	25.4
Yang *et al.* [61]	0.789	0.770	30.3	31.1	0.793	-3.2	8.3	0.986	-10.8	11.6

TABLE IX
MYO TEST RESULTS FROM THE ACDC DATASET FOR OUR GridNet-MD METHOD AND ACDC CHALLENGERS

	Dice		Hausdorff		Mass ED		
	ED	ES	ED	ES	corr.	bias	std-dev.
Isensee *et al.* [20]	0.902	0.919	8.7	8.7	0.989	-4.8	7.6
GridNet-MD	0.886	0.902	9.6	9.3	0.986	-1.8	8.6
Khened *et al.* [40]	0.889	0.898	9.8	12.6	0.990	-2.9	7.5
Jain *et al.* [59]	0.882	0.897	9.8	11.3	0.989	11.6	8.1
Baumgartner *et al.* [42]	0.892	0.901	8.7	10.6	0.982	-6.9	9.8
Zotti *et al.* [43]	0.884	0.896	8.7	9.3	0.984	-12.4	9.0
Wolterink *et al.* [58]	0.875	0.894	11.1	10.7	0.963	-1.0	14.6
Jang *et al.* [44]	0.875	0.895	9.9	8.9	0.968	11.5	12.9
Roh *et al.* [57]	0.867	0.869	11.5	13.0	0.967	-3.4	13.3
Tziritas-Grinias [60]	0.794	0.801	12.6	14.8	0.942	-28.9	28.0
Yang *et al.* [61]	N/A	N/A	N/A	N/A	N/A	N/A	N/A

that the interslice gap is 10 mm on average, the 3 mm increase on the MICCAI 2009 dataset corresponds to less than one pixel.

Table VII, VIII, and IX show results from the online ACDC leaderboard. Results reported in these tables follow the same ranking procedure than that used by the online ACDC plateform. First, each method gets a ranking position (from 1 to 11) for each metric. This leads to 8 ranks for every method in Table VII and VIII (std-dev are not considered in the ranking) and 6 for Table IX. These ranks are than averaged out leading to a global average ranking score. This ranking strategy is similar to that used in other challenges [55], [56]. For each category, our method *GridNet-MD* is almost always ranked second. In first position is Isensee *et al.* [20] which uses an ensemble of 2D and 3D U-Nets. In our case, we have only one 2D model that fits on a single GPU and processes a stack of 10 2D MR images in 0.18 seconds on average which is two times faster than Isensee *et al.* [20]. Our method is thus in second position overall but first among those using a single CNN.

Figure 4 shows results of our method obtained on the validation set. Segmentation maps with the worst, median and best

Authorized licensed use limited to: University of Electronic Science and Tech of China. Downloaded on February 06,2025 at 14:58:24 UTC from IEEE Xplore. Restrictions apply.

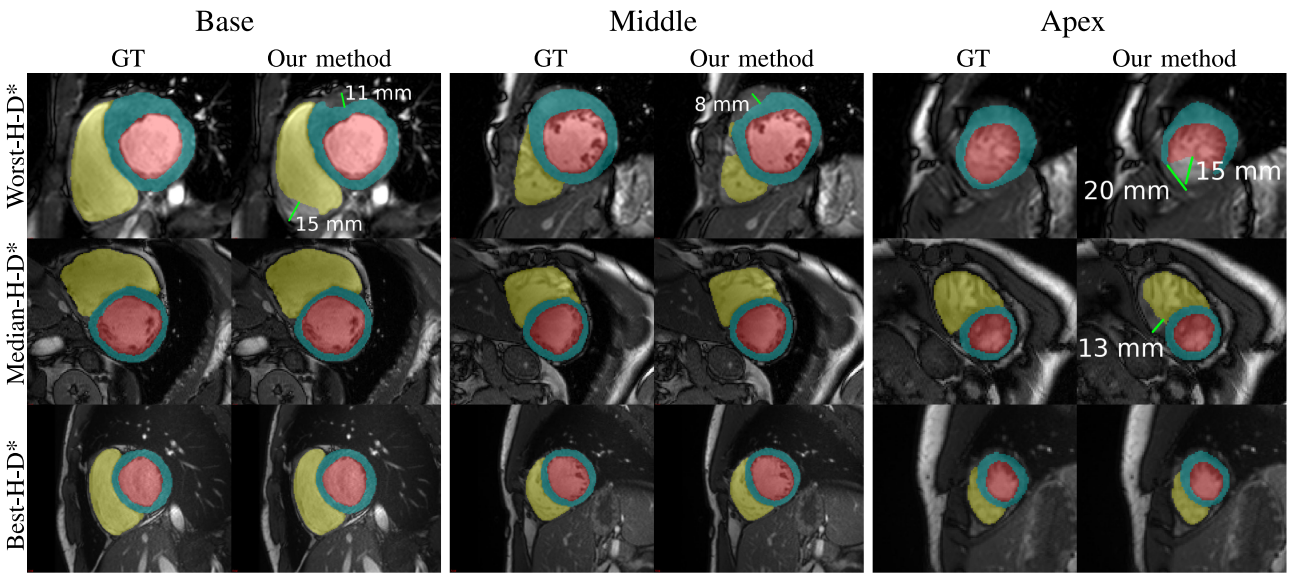


Fig. 4. Groundtruth (GT) from the validation set and results of our method. The columns show slices at the base, the middle, and the apex of the left ventricle. The first row shows the worst result according to the Hausdorff distance and the Dice coefficient; the second row show results with the median Hausdorff and Dice, and the third row is the most accurate result.
*H-D: Hausdorff and Dice.

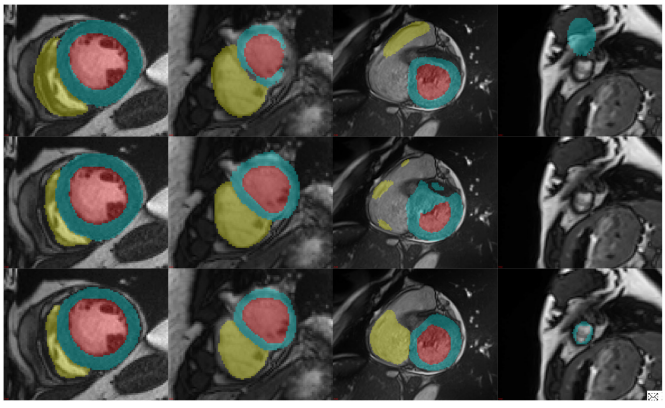


Fig. 5. Visual comparison results on the test set where each row shows the results of ConvDeconv, U-Net and GridNet-MD. Each column represent one slice of a patient.

segmentation accuracy according to the Dice and the Hausdorff metrics are shown. Even the worst results (those with Hausdorff distances of more than 10 mm on several slices) are anatomically plausible and do not suffer from isolated false positives and false negatives. Visual inspection reveals that the apex is a challenging section due to the small size of the cardiac anatomy at that position. The base is also challenging since the RV and the aorta are simultaneously visible due to partial volume effect, and also since there is not a clear separation between RV and the pulmonary artery, not between the RV and the right atrium. As for the results with median HD, usually only one or two slices have an HD above 10 mm, the other slices having an HD below 5 mm, especially for the LV and the MYO.

In Fig. 5, we compare results from our method to those from ConvDeconv and U-Net. While our method is not perfect (c.f. the bottom right image showing a myocardium without a LV inside) our method produces fewer false positives and false negatives, thanks to the shape prior.

TABLE X
PERCENTAGE OF SLICES OF THE ACDC VALIDATION SET WHOSE DICE COEFFICIENT IS BELOW 90% OR 2D HAUSDORFF DISTANCE IS ABOVE 5 mm

Methods	Dice			HD		
	RV	MYO	LV	RV	MYO	LV
ConvDeconv	65.2%	84.8%	20.4%	79.8%	43.1%	13.5%
U-Net	39.4%	64.2%	16.7%	52.7%	25.2%	11.5%
GridNet-MD	**32.9** %	**44.2**%	**13.1**%	**42.3**%	**14.4**%	**8.1**%

In order to further identify where our method best performs, we computed the average 2D Hausdorff distance and 2D Dice score of the LV, RV and MYO for each slice of the validation set. Results for our method, ConvDeconv and U-Net on the RV, MYO and LV are shown in Fig. 6. Our method is more accurate than the other two on each slice of the volume. Also, as the visual inspection suggested, our method is struggling more next to the apex and the valves. This is especially true for the RV which can be easily confused with the atrium and the pulmonary artery next to the valves.

To further inspect the limits of our method, we computed the percentage of slices for which the results could be considered outside the inner-observed margin. Although the ACDC dataset do not provide such margin, the two experts which annotated the images gave themselves an error margin of ± 1 pixel i.e. a 2D Hausdorff distance below 5 mm and a Dice score above 0.9. Results for the three methods are presented in Table X. Again, our method outperforms the other two, especially on the RV.

Authorized licensed use limited to: University of Electronic Science and Tech of China. Downloaded on February 06,2025 at 14:58:24 UTC from IEEE Xplore. Restrictions apply.

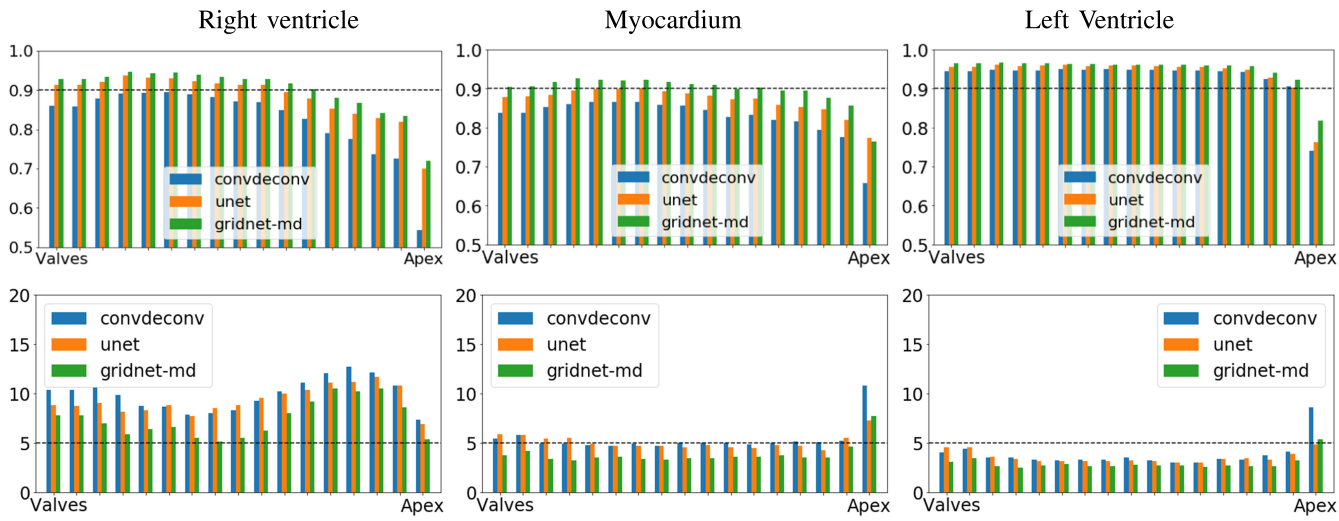


Fig. 6. [Row 1] Mean dice coefficient per slice (higher is better). [Row 2] mean Hausdorff distance per slice (lower is better). This evaluation is made on the validation set.

V. Conclusion

In this paper, we proposed a CNN architecture designed for kinetic cardiac MRI segmentation. Our method is an extension of the U-Net as it integrates 4 new aspects. First, the U-Net skip connexions are replaced by convolution layers so the concatenated feature maps contain more meaningful informations. Second, our system uses a shape prior whose registration is embedded within its structure. Since the shape prior is computed only once before training and testing, the execution of our network is not limited by its calculation. Third, our system incorporates a multiresolution input and output to further improve its ability to globally localize the heart as well as it segmentation accuracy. And forth, a five-term loss is used which, among other things, penalizes poorly aligned contours. All these improvements make our method less vulnerable to spurious false positives and false negatives. Results show that our method is highly accurate with an overall Dice score of 0.91 and a mean Hausdorff distance of 9.5 mm. Our method is also fast with an average processing time of 0.18 seconds to process a stack of 10 2D MR images. As of today, our method outperforms every state-of-the-art methods except for that of Isensee *et al.* [20] which involves an ensemble of CNN. Our approach is a good step towards the ultimate goal of having a fully-automatic method whose error rate is within the human annotator margin.

References

[1] WHO, "Cardiovascular diseases (cvds) - fact sheets," 2017, [Online]. Available: www.who.int/mediacentre/factsheets/fs317. Accessed on: Apr. 20, 2018.

[2] F. H. Epstein, "Mri of left ventricular function," *J. Nucl. Cardiol.*, vol. 14, no. 5, pp. 729–744, 2007.

[3] G. W. Vick, "The gold standard for noninvasive imaging in coronary heart disease: Magnetic resonance imaging," *Current Opinion Cardiol.*, vol. 24, no. 6, pp. 567–579, 2009.

[4] W. G. Hundley *et al.*, "Accf/acr/aha/nasci/scmr 2010 expert consensus document on cardiovascular magnetic resonance: A report of the american college of cardiology foundation task force on expert consensus documents," *J. Amer. College Cardiol.*, vol. 55, no. 23, pp. 2614–2662, 2010.

[5] C. M. Kramer *et al.*, "Standardized cardiovascular magnetic resonance (CMR) protocols 2013 update," *J. Cardiovascular Magn. Reson.*, vol. 15, no. 1: 91, 2013.

[6] P. Peng *et al.*, "A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging," *Magn. Reson. Mater. Phys., Biol. Med.*, vol. 29, no. 2, pp. 155–195, 2016.

[7] C. Petitjean *et al.*, "Right ventricle segmentation from cardiac mri: A collation study," *Med. Image Anal.*, vol. 19, no. 1, pp. 187–202, 2015.

[8] D. A. Auger *et al.*, "Semi-automated left ventricular segmentation based on a guide point model approach for 3d cine dense cardiovascular magnetic resonance," *J. Cardiovascular Magn. Reson.*, vol. 16: 8, 2014.

[9] D. Grosgeorge *et al.*, "Graph cut segmentation with a statistical shape model in cardiac MRI," *Comput. Vis. Image Understanding*, vol. 117, no. 9, pp. 1027–1035, 2013.

[10] C. Petitjean and J. N. Dacher, "A review of segmentation methods in short axis cardiac mr images," *Med. Image Anal.*, vol. 15, no. 2, pp. 169–184, 2011.

[11] L. Wang *et al.*, "Left ventricle: Fully automated segmentation based on spatiotemporal continuity and myocardium information in cine cardiac magnetic resonance imaging (lv-fast)," *BioMed Res. Int.*, vol. 2015, 2015, Art. no. 367583.

[12] Y. Liu *et al.*, "Distance regularized two level sets for segmentation of left and right ventricles from CINE-MRI," *Mag. Res. Img.*, vol. 34, no. 5, pp. 699–706, 2016.

[13] P. V. Tran, "A fully convolutional neural network for cardiac segmentation in short-axis mri," 2016, arXiv:1604.00494.

[14] L. K. Tan *et al.*, "Cardiac left ventricle segmentation using convolutional neural network regression," in *Proc. IEEE EMBS Conf. Biomed. Eng. Sci.*, 2016, pp. 490–493.

[15] H. Noh *et al.*, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1520–1528.

[16] T. A. Ngo *et al.*, "Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance," *Med. Image Anal.*, vol. 35, pp. 159–171, 2017.

[17] O. Bernard, "Acdc-miccai 2017 web plateform," 2017, [Online]. Available: acdc.creatis.insa-lyon.fr. Accessed on: Apr. 20, 2018.

[18] O. Bernard *et al.*, "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE Trans. Med. Imag.*, p. 1, 2018.

[19] O. Ronneberger *et al.*, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.

[20] F. Isensee *et al.*, "Automatic cardiac disease assessment on CINE-MRI via time-series segmentation and domain specific features," in *Proc. Int. Workshop Statist. Atlases Comput. Models Heart*, 2017, pp. 120–129.

[21] I. B. Ayed *et al.*, "Max-flow segmentation of the left ventricle by recovering subject-specific distributions via a bound of the bhattacharyya measure," *Med. Image Anal.*, vol. 16, no. 1, pp. 87–100, 2012.

Authorized licensed use limited to: University of Electronic Science and Tech of China. Downloaded on February 06,2025 at 14:58:24 UTC from IEEE Xplore. Restrictions apply.

[22] M. A. Zuluaga *et al.*, "Automatic right ventricle segmentation using multi-label fusion in cardiac MRI," in *Proc. MICCAI*, 2012.

[23] Y. Ou *et al.*, "Multi-atlas segmentation of the right ventricle in cardiac MRI," in *Proc. MICCAI*, 2012.

[24] J. Ringenberg *et al.*, "Fast, accurate, and fully automatic segmentation of the right ventricle in short-axis cardiac MRI," *Comput. Med. Imag. Graph*, vol. 38, no. 3, pp. 190–201, 2014.

[25] S. Queirós *et al.*, "Fast automatic myocardial segmentation in 4d cine cmr datasets," *Med. Image Anal.*, vol. 18, no. 7, pp. 1115–1131, 2014.

[26] J. Woo, P. J. Slomka, C.-C. J. Kuo, and B.-W. Hong, "Multiphase segmentation using an implicit dual shape prior: Application to detection of left ventricle in cardiac mri," *Comput. Vis. Image Understanding*, vol. 117, no. 9, pp. 1084–1094, 2013.

[27] Y. Wu, Y. Wang, and Y. Jia, "Segmentation of the left ventricle in cardiac cine mri using a shape-constrained snake model," *Comput. Vis. Image Understanding*, vol. 117, no. 9, pp. 990–1003, 2013.

[28] W. Bai, W. Shi, C. Ledig, and D. Rueckert, "Multi-atlas segmentation with augmented features for cardiac mr images," *Med. Image Anal.*, vol. 19, no. 1, pp. 98–109, 2015.

[29] S. C. Mitchell *et al.*, "Multistage hybrid active appearance model matching: Segmentation of left and right ventricles in cardiac mr images," *IEEE Trans. Med. Imag.*, vol. 20, no. 5, pp. 415–423, May 2001.

[30] T. A. Ngo *et al.*, "Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance," *Med. Image Anal.*, vol. 35, pp. 159–171, 2017.

[31] M. R. Avendi *et al.*, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI," *Med. Image Anal.*, vol. 30, pp. 108–119, 2016.

[32] C. Rupprecht *et al.*, "Deep active contours," 2016, arXiv:1607.05074.

[33] H. Yang *et al.*, "Deep fusion net for multi-atlas segmentation: Application to cardiac MR images," in *Proc. MICCAI*, 2016, pp. 521–528.

[34] J. Long E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[35] R. P. Poudel *et al.*, "Recurrent fully convolutional neural networks for multi-slice mri cardiac segmentation," in *Proc. Int. Workshop Reconstruction Anal. Moving Body Organs*, 2016, pp. 83–94.

[36] O. Oktay *et al.*, "Multi-input cardiac image super-resolution using convolutional neural networks," in *Proc. 19th Int. Conf. Med. Image Comput. Comput. Assisted Interventions*, 2016, pp. 246–254.

[37] B. Gu, Y. Shan, V. S. Sheng, Y. Zheng, and S. Li, "Sparse regression with output correlation for cardiac ejection fraction estimation," *Inform. Sci.*, vol. 423, pp. 303–312, 2018.

[38] W. Xue, A. Islam, M. Bhaduri, and S. Li, "Direct multitype cardiac indices estimation via joint representation and regression learning," *IEEE Trans. Med. Imag.*, vol. 36, no. 10, pp. 2057–2067, Oct. 2017.

[39] W. Xue, G. Brahm, S. Pandey, S. Leung, and S. Li, "Full left ventricle quantification via deep multitask relationships learning," *Med. Image Anal.*, vol. 43, pp. 54–65, 2018.

[40] M. Khened *et al.*, "Densely connected fully convolutional network for short-axis cardiac cine MR image segmentation and heart diagnosis using random forest," in *Proc. Int. Workshop Statist. Atlases Comput. Models Heart*, 2017, pp. 140–151.

[41] G. Huang *et al.*, "Densely connected convolutional networks," in *Proc. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[42] C. Baumgartner *et al.*, "An exploration of 2d and 3d deep learning techniques for cardiac MR image segmentation," in *Proc. Int. Workshop Statist. Atlases Comput. Models Heart*, 2017, pp, 111–119.

[43] C. Zotti *et al.*, "Gridnet with automatic shape prior registration for automatic MRI cardiac segmentation," in *Proc. Int. Workshop Statist. Atlases Comput. Models Heart*, 2017, pp. 73–81.

[44] Y. Jang *et al.*, "Automatic segmentation of lv and rv in cardiac mri," in *Proc. Int. Workshop Statist. Atlases Comput. Models Heart*, 2017, pp. 161–169.

[45] R. Mehta and J. Sivaswamy, "M-net: A convolutional neural network for deep brain structure segmentation," in *Proc. Int. Symp. Biomed. Imag.*, 2017, pp. 437–440.

[46] B. Kastler, "Cardiovascular anatomy and atlas of MR normal anatomy," in *MRI Cardiovascular Malformations.* New York, NY, USA: Springer, 2011, ch. 2, pp. 17–39.

[47] Z. Luo, A. Mishra, A. Achkar, J. Eichel, S.-Z. Li, and P.-M. Jodoin, "Non-local deep features for salient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017. pp. 6593–6601.

[48] M. M. Rohé *et al.*, "SVF-NET: Learning deformable image registration using shape matching," in *Proc. Med. Image Comput. Comput. Assisted Intervention*, 2017, pp. 266–274.

[49] N. Srivastava *et al.*, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[50] W. R. Crum, O. Camara, and D.L.G. Hill, "Generalized overlap measures for evaluation and validation in medical image analysis," *IEEE Trans. Med. Imag.*, vol. 25, no. 11, pp. 1451–1461, Nov. 2006.

[51] A. Lalande *et al.*, "Evaluation of cardiac structure segmentation in cine magnetic resonance imaging," in *Multi-Modality Cardiac Imaging*. New York, NY, USA: Wiley, 2015, pp. 169–215.

[52] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, 2015.

[53] F. Chollet *et al.*, "Keras," 2015, [Online]. Available: https://github.com/fchollet/keras

[54] E. Whitley and J. Ball, "Statistics review 6: Nonparametric methods," *Crit. Care*, vol. 6, no. 6, pp. 509–513, 2002.

[55] O. Bernard *et al.*, "Standardized evaluation system for left ventricular segmentation algorithms in 3d echocardiography," *IEEE Trans. Med. Img.*, vol. 35, no. 4, pp. 967–977, Apr. 2016.

[56] D. P.Young and J. M. Ferryman, "Pets metrics: On-line performance evaluation service," in *Proc. IEEE Int. Workshop Perform. Eval. Tracking Syst*, 2005, pp. 317–324.

[57] M. Rohé *et al.*, "Automatic multi-atlas segmentation of myocardium with SVF-NET," in *Proc. Int. Workshop Statist. Atlases Comput. Models Heart*, 2017, pp. 170–177.

[58] J. M. Wolterink *et al.*, "Automatic segmentation and disease classification using cardiac cine MR images," in *Proc. Int. Workshop Statist. Atlases Comput. Models Heart*, 2017, pp. 101–110.

[59] J. Patravali *et al.*, "2d–3d fully convolutional neural networks for cardiac MR segmentation," in *Proc. Int. Workshop Statist. Atlases Comput. Models Heart*, 2017, pp. 130–139.

[60] G. Tziritas and E. Grinias, "Fast fully-automatic localization of left ventricle and myocardium in mri using mrf model optimization, substructures tracking and b-spline smoothing," in *Proc. Int. Workshop Statist. Atlases Comput. Models Heart*, 2017, pp. 91–100.

[61] X. Yang *et al.*, "Class-balanced deep neural network for automatic ventricular structure segmentation," in *Proc. Int. Workshop Statist. Atlases Comput. Models Heart*, 2017, pp. 152–160.

Authorized licensed use limited to: University of Electronic Science and Tech of China. Downloaded on February 06,2025 at 14:58:24 UTC from IEEE Xplore. Restrictions apply.