Improving Myocardial Scar Segmentation with End-to-End and Cascaded CNN Using Hybrid Loss and Multi-Modality CMR Imaging

Farheen Ramzan[1,2], Chen (Cherise) Chen[1,2], Richard H Clayton[1,2]

[1] Insigneo Institute for *in-silico*, Medicine, University of Sheffield, Sheffield, UK
[2] Department of Computer Science, University of Sheffield, Sheffield, UK

Abstract

Late gadolinium enhancement (LGE) CMR imaging is commonly used to locate and quantify ventricular scars for treatment planning after myocardial infarction. Effective treatment such as ablation can mitigate ventricular tachycardia risk. A deep learning based segmentation approach can provide an automated and efficient way to segment and assess the severity of ventricular scars. In this paper, we investigate various end-to-end and cascaded deep learning pipelines based on a self-configuring convolutional neural network for myocardial scar segmentation. We build a standardized framework to systematically a) study whether the cascaded pipeline (which first learns to capture cardiac anatomical structural information and then uses this as shape prior knowledge to refine the segmentation of abnormal scar and edema regions) works better than the end-to-end one-stage approach; b) analyze whether adding complementary CMR sequences including T2 and bSSFP can improve the segmentation accuracy; c) search for the best combination of existing loss functions to tackle class imbalance, the problem behind the task. Our results suggest that the end-to-end model works better than the cascaded one, which can achieve an average Dice score of 0.7257 when trained with a hybrid loss (based on the standard Dice and focalTvesky loss functions) for the scar segmentation task on the public MyoPS 2020 test dataset, outperforming the winner algorithm of the MyoPS challenge.

1. Introduction

Myocardial Infarction (MI) is caused by a sudden reduction or complete cessation of blood supply to a part of the myocardium and is the most common cause of ventricular scar that provides a substrate for potentially life-threatening ventricular tachycardia (VT) [1]. Cardiac magnetic resonance (CMR) imaging, including late gadolinium enhancement (LGE), T2-weighted, and balanced steady-state free precession (bSSFP) cine sequence, is used to capture the effects of MI in-vivo [2].

In clinical practice, the segmentation of myocardial pathology from CMR images is required for treatment planning of VT. However, manual segmentation is inefficient and subject to inter-observer variability. Deep Learning (DL) approaches are being broadly adopted in biomedical image analysis, for instance, U-nets [3] have been widely used for segmentation tasks of varying complexities. Consequently, several automated approaches for scar segmentation have emerged [4]. These approaches require manual configuration of the network architecture leading to sub-optimal performance. Thus, the accurate segmentation of scars remains challenging due to the heterogeneous nature and small size of scars, and indistinguishable contours owing to low soft tissue contrast in CMR imaging.

Studies [2] indicate that LGE can visualize MI, T2-weighted can capture acute injury and ischemic regions, and bSSFP cine sequence can detect clear ventricular boundaries. Therefore, to capture complex information about scar shape, it can be hypothesized that incorporating information from multiple CMR sequences effectively can help improve the segmentation performance.

Our primary goals are: (i) to build a standardized pipeline for myocardial scar segmentation based on the commonly used self-configuring 2D convolutional neural network (CNN), i.e., nn-Unet [5]; (ii) to assess the performance differences between end-to-end and cascaded pipelines to study the impact of incorporating prior knowledge of anatomical structures and the spatial location of myocardial scars, in together with varied combinations of three CMR modalities as model input; and (iii) to identify the most effective loss function from established options to enhance segmentation accuracy in the challenging areas of abnormal scars and edema, which are affected by class imbalance and poor imaging contrasts. This study seeks to pave the way for future research into developing computational models for myocardial scar characterization to guide clinicians in radio-frequency catheter ablation to abolosh VT [1].

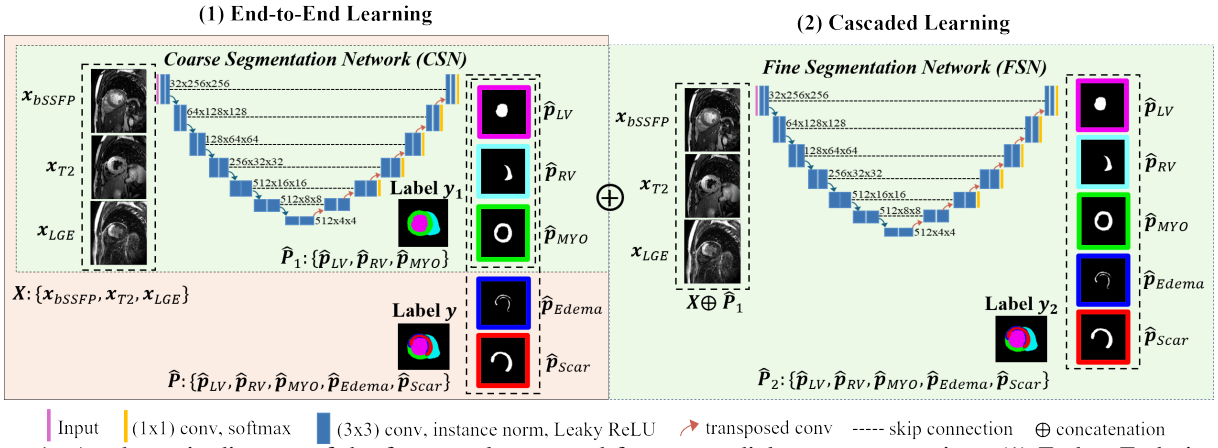

Figure 1: A schematic diagram of the framework proposed for myocardial scar segmentation. (1) End-to-End pipeline (orange region) takes $\mathbf{X} : \{\mathbf{x}_{bSSFP}, \mathbf{x}_{T2}, \mathbf{x}_{LGE}\}$ as input and predicts $\hat{\mathbf{P}} : \{\hat{\mathbf{p}}_{LV}, \hat{\mathbf{p}}_{RV}, \hat{\mathbf{p}}_{MYO}, \hat{\mathbf{p}}_{Edema}, \hat{\mathbf{p}}_{Scar}\}$ as output in one step. (2) Cascaded pipeline (green region), the first stage is a coarse segmentation network (CSN) that takes $\mathbf{X} : \{\mathbf{x}_{bSSFP}, \mathbf{x}_{T2}, \mathbf{x}_{LGE}\}$ as input and predicts $\hat{\mathbf{P}}_1 : \{\hat{\mathbf{p}}_{LV}, \hat{\mathbf{p}}_{RV}, \hat{\mathbf{p}}_{MYO}\}$ as output. The second stage is a fine segmentation network (FSN) that takes $\mathbf{X} \bigoplus \hat{\mathbf{P}}_1$ as input and predicts the final segmentation maps denoted by $\hat{\mathbf{P}}_2 : \{\hat{\mathbf{p}}_{LV}, \hat{\mathbf{p}}_{RV}, \hat{\mathbf{p}}_{MYO}, \hat{\mathbf{p}}_{Edema}, \hat{\mathbf{p}}_{Scar}\}$, where $\bigoplus$ denotes channel-wise concatenation.

2. Methods

Dataset: The MyoPS 2020 challenge dataset [6, 7] was used, which provides 45 CMR images acquired from the same patients in three different modalities including LGE-MRI, T2-weighted, and bSSFP. Gold standard labels for left ventricle (LV) blood pool, right ventricle (RV) blood pool, normal myocardium (MYO), scar, and edema have also been provided.

End-to-End Learning: Baseline Network. A 2D nnUNet [5] is adopted in an end-to-end manner, which takes three CMR imaging modalities $\mathbf{X} : \{\mathbf{x}_{bSSFP}, \mathbf{x}_{T2}, \mathbf{x}_{LGE}\}$ as input and produces cardiac substructures and myocardial pathology in one step. Where $\hat{\mathbf{P}} : \{\hat{\mathbf{p}}_{LV}, \hat{\mathbf{p}}_{RV}, \hat{\mathbf{p}}_{MYO}, \hat{\mathbf{p}}_{Edema}, \hat{\mathbf{p}}_{Scar}\}$ is the output of the end-to-end learning framework. The output of the baseline model with trainable parameters θ can be denoted by:

$$\hat{\mathbf{P}} = Baseline(\mathbf{X}; \theta)$$

Cascaded Learning: Coarse to Fine Segmentation (CNN^{C2F}). Alternatively, one can adopt a cascaded coarse-to-fine segmentation framework CNN^{C2F} for this task, which consists of two stages to exploit the hierarchical nature of the infarcted regions with respect to cardiac substructures. It first focuses on segmenting the complete MYO, covering healthy MYO, scar, and edema regions ($MYO(complete) = MYO(healthy) \cup Scar \cup Edema$), and then differentiating scar and edema regions ($Scar(complete) = Scar \cup Edema$) from the healthy ones. An overview of the pipeline is also presented in Figure 1. Here, each CNN utilizes nnUNet [5] as a base architecture. The nnUNet framework automatically configures itself to select pre-processing methods, network architecture, training scheme, and post-processing methods to train a network configuration in a five-fold cross-validation setting and determines the optimal model ensemble for each stage.

Coarse Segmentation Network (CSN): In the first stage, a Coarse Segmentation Network (CSN) learns to effectively capture information about the cardiac substructures. The 2D images incl. paired three different modalities $\mathbf{X} : \{\mathbf{x}_{bSSFP}, \mathbf{x}_{T2}, \mathbf{x}_{LGE}\}$ along with the ground truth y_1 are randomly drawn from the training set for supervised training. The output of CSN is coarse structural segmentation maps of LV, RV and MYO: $\hat{\mathbf{P}}_1 : \{\hat{\mathbf{p}}_{LV}, \hat{\mathbf{p}}_{RV}, \hat{\mathbf{p}}_{MYO}\}$. The output of the first stage with trainable parameters θ_1 can be denoted by:

$$\hat{\mathbf{P}}_1 = CSN(\mathbf{X}; \theta_1)$$

Fine Segmentation Network (FSN): The second stage is a Fine Segmentation Network (FSN) that incorporates the coarse maps generated by the first stage as prior knowledge to segment edema and scar regions. It takes $\mathbf{X} \bigoplus \hat{\mathbf{P}}_1$ (channel-wise concatenation of $\mathbf{X}$ and $\hat{\mathbf{P}}_1$) as input to predict the final segmentation maps $\hat{\mathbf{P}}_2 : \{\hat{\mathbf{p}}_{LV}, \hat{\mathbf{p}}_{RV}, \hat{\mathbf{p}}_{MYO}, \hat{\mathbf{p}}_{Edema}, \hat{\mathbf{p}}_{Scar}\}$. The output of FSN with trainable parameters θ_2 can be denoted by:

$$\hat{\mathbf{P}}_2 = FSN(\mathbf{X} \bigoplus \hat{\mathbf{P}}_1; \theta_2)$$

Loss Functions: During training, the segmentation loss of each stage is computed individually. We formalized some hybrid loss functions by combining existing loss functions such as Dice loss (L_{Dice}), cross-entropy loss (L_{CE}), focal loss (L_{Focal}), Tversky loss (L_{Tv}), boundary loss (L_{BD}), Hausdorff distance loss (L_{HD}) and intersection-over-union loss (L_{IOU}) [8]:

$$L_{DiceCE}(V_G, V_P) = L_{Dice} + L_{CE} \quad (1)$$

$$L_{DiceHD}(V_G, V_P) = L_{Dice} + L_{HD} \quad (2)$$

Table 1: Results obtained on the MyOPS test dataset with models trained using the end-to-end (baseline) and the cascaded pipeline strategies with varied CMR modalities as input, where *Phase*$_1$ and *Phase*$_2$ represent CSN and FSN of CNN^{C2F}, respectively. Here, we used the standard loss: L_{DiceCE} in our experiments.

Strategy	Model Type		CMR Modality (Input)			Cardiac Structure (Output)					Dice (Scar)	Dice (Edema + Scar)	Dice (Average)
			LGE	T2	bSSFP	LV	RV	MYO	Edema	Scar			
S_1	End-to-End		✓	✗	✗	✓	✓	✓	✓	✓	0.6425	0.5894	0.6160
S_2	End-to-End		✓	✓	✓	✓	✓	✓	✓	✓	**0.7247**	0.7095	**0.7171**
S_3	Cascaded	*Phase*$_1$	✗	✗	✓	✓	✗	✓	✗	✗	0.6051	0.6458	0.6255
		Phase$_2$	✓	✓	✗	✗	✗	✗	✓	✓			
S_4	Cascaded	*Phase*$_1$	✗	✗	✓	✓	✓	✓	✗	✗	0.6087	0.6445	0.6266
		Phase$_2$	✓	✓	✗	✓	✓	✓	✓	✓			
S_5	Cascaded	*Phase*$_1$	✓	✗	✗	✓	✓	✓	✗	✗	0.6196	0.6014	0.6105
		Phase$_2$	✓	✗	✗	✓	✓	✓	✓	✓			
S_6	Cascaded	*Phase*$_1$	✓	✓	✓	✓	✓	✓	✗	✗	**0.7076**	0.7215	**0.7146**
		Phase$_2$	✓	✓	✓	✓	✓	✓	✓	✓			
Zhai et al. [4] *MyOPS Winner*	Cascaded	*Phase*$_1$	✓	✓	✓	✓	✓	✓	✗	✗	0.6723	**0.7314**	0.7019
		Phase$_2$	✓	✓	✓	✓	✓	✓	✓	✓			

$$L_{FocalTv}(V_G, V_P) = L_{Focal} + L_{Tv} \quad (3)$$

$$L_{DiceFocalTv}(V_G, V_P) = L_{Dice} + L_{Focal} + L_{Tv} \quad (4)$$

$$L_{IOUCE}(V_G, V_P) = L_{IOU} + L_{CE} \quad (5)$$

$$L_{H_1}(V_G, V_P) = \alpha(L_{Dice} + L_{CE}) + (1-\alpha)L_{BD} \quad (6)$$

$$L_{H_2}(V_G, V_P) = \alpha(L_{Dice} + L_{Focal}) + (1-\alpha)L_{BD} \quad (7)$$

$$L_{H_3}(V_G, V_P) = \alpha(L_{FocalTv} + L_{Focal}) + (1-\alpha)L_{BD} \quad (8)$$

$$L_{H_4}(V_G, V_P) = \alpha(L_{Dice} + L_{FocalTv}) + (1-\alpha)L_{BD} \quad (9)$$

Where V_P and V_G denote the predicted and ground truth segmentation, respectively and α continuously decays with the increase in the number of epochs so that in the early stage of the training, the model pays more attention to the compound loss (linear combination of other losses) and in the later stage of training, it pays more attention to the boundary loss.

3. Results and Discussion

Implementation Details: All the networks were trained for $epoch_{max} = 1000$ epochs with a batch size of 5 and a patch size of 256×256. The stochastic gradient descent (SGD) method was used as an optimizer with an initial learning rate of 0.01 reduced continuously using the 'polyLR' scheme: $1-(epoch/epoch_{max})^{0.9}$. Images were normalized by applying z-score normalization and random data augmentation was applied on the fly by using various transforms.

Quantitative Evaluation: We evaluated our models on the MyOPS test dataset which includes 20 test samples in three CMR modalities and an evaluation tool that provides Dice scores for scar and edema+scar. We designed different strategies for training end-to-end and cascaded networks. These strategies differ from each other based on the type and number of modalities used as input and output classes targeted by the network, results are given in Table 1. Note that for these experiments, we utilized L_{DiceCE} as a default loss function for both stages of the cascaded network. We observed that the end-to-end model (S_2) obtained the highest Dice score of 0.7247 for the scar segmentation task as well as the best overall average Dice score of 0.7171. While the cascaded model (S_6) achieved the best Dice score of 0.7215 for edema+scar segmentation.

Table 2: Results obtained with MyOPS test dataset by training end-to-end (S_2) and cascaded (S_6) networks using different loss functions.

Loss Function	Dice (Scar)		Dice (Edema + Scar)	
	S_2	S_6	S_2	S_6
L_{DiceCE}	0.7247	0.7076	**0.7095**	0.7215
L_{DiceHD}	0.7221	0.7111	0.7061	**0.7295**
$L_{FocalTv}$	0.7182	**0.7117**	0.7021	0.7225
$L_{DiceFocalTv}$	**0.7257**	0.7073	0.7082	0.7236
L_{IOUCE}	0.7236	0.7115	0.7090	0.7238
L_{H_1}	0.7148	0.7084	0.6992	0.7243
L_{H_2}	0.7205	0.7062	0.7055	0.7224
L_{H_3}	0.7231	**0.7140**	0.7094	0.7265
L_{H_4}	**0.7254**	0.7101	**0.7118**	**0.7272**

* Top two best Dice scores per model (**1**st and **2**nd) are highlighted.
* Gray region represents the overall best segmentation performance.
* We used L_{DiceCE} for the first stage of the cascaded model (S_6).

We designed further experiments with the two best strategies to employ several loss functions at the training stage as described in section 2, and results are presented in Table 2. Note that for the cascaded model (S_6), we used L_{DiceCE} for the first stage while employing different loss functions in the second stage. We observed that the accuracy for the scar segmentation task improved to 0.7257 when our end-to-end model was trained with $L_{DiceFocalTv}$. Similarly, the accuracy of the edema+scar segmentation task was improved to 0.7295 when the cascaded model was trained with L_{DiceHD}.

Our experiment results reveal that the end-to-end model could perform better than the cascaded model, which obtained the best average and scar segmentation accuracy. Prior work [4] majorly focuses on cascaded methods. However, these methods have not considered the fact that the segmentation error introduced in the first step could propagate to the next phase making it hard to obtain precise segmentation of the scar and edema regions. Therefore, better approaches to utilize complementary shape prior information are required.

Qualitative Evaluation: We performed a qualitative

evaluation by visually inspecting the predictions obtained with our models, a test case is shown in Figure 2. We observed that CMR images in the MyOPS test dataset are of quite low contrast and there is an extreme imbalance between the pixels representing myocardial pathology and other cardiac structures. Overall the results appear to be realistic, however, the predictions can only be compared to the LGE CMR in the absence of test ground truth labels. Hence, at the current stage, the predictions of our model need to be verified by a domain expert. However, our model can aid in accelerating the manual segmentation process and can reduce intra- and inter-observer variability to obtain correct segmentation for myocardial pathology.

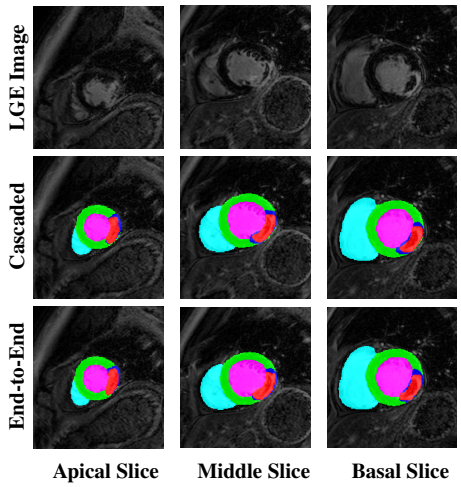

Figure 2: Visualizing predictions for three consecutive slices for a test CMR from MyOPS dataset: (row 1) LGE images, (row 2) and (row 3) predictions obtained with cascaded and end-to-end pipeline overlaid on LGE, respectively. Colored regions represent respective class LV, RV, MYO, edema and scar.

4. Conclusion

We have presented a standardized pipeline for myocardial scar segmentation based on a self-configuring 2D CNN, which could be trained either in an end-to-end or in a cascaded manner. We explored different model learning strategies by comparing different pipelines and varying input CMR modalities and output classes to affect the supervision process as well as exploring different loss functions and their combination for enhanced model supervision. Contrary to the previous studies, our end-to-end model achieved a higher accuracy for the scar segmentation task compared to the commonly used cascaded pipeline. Our study suggests that an end-to-end network taking into the three complementary imaging modalities as input and optimized with a composite loss incl. the commonly used Dice (focusing on the region mismatch) and the focalTvesky loss (focusing on the small structures under the class-imbalance setting) achieves the best performance on the scar segmentation. This study can be a step toward future research into the development of computational models for myocardial scar quantification to help clinicians in the patient-specific treatment planning of ventricular tachycardia.

Acknowledgments

Farheen Ramzan acknowledges support from a PhD scholarship funded by the Commonwealth Scholarship Commission in the United Kingdom. IT Services at The University of Sheffield provided High Performance Computing used in this study.

References

[1] Stevenson WG. Ventricular scars and ventricular tachycardia. Transactions of the American Clinical and Climatological Association 2009;120:403.

[2] Wu Y, Tang Z, Li B, Firmin D, Yang G. Recent advances in fibrosis and scar segmentation from cardiac mri: a state-of-the-art review and future perspectives. Frontiers in Physiology 2021;12:709230.

[3] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer, 2015; 234–241.

[4] Li L, Wu F, Wang S, Luo X, Martín-Isla C, Zhai S, Zhang J, Liu Y, Zhang Z, Ankenbrand MJ, et al. Myops: A benchmark of myocardial pathology segmentation combining three-sequence cardiac magnetic resonance images. Medical Image Analysis 2023;87:102808.

[5] Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 2021; 18(2):203–211.

[6] Zhuang X. Multivariate mixture model for myocardial segmentation combining multi-source images. IEEE Transactions on Pattern Analysis and Machine Intelligence 2018; 41(12):2933–2946.

[7] Qiu J, Li L, Wang S, Zhang K, Chen Y, Yang S, Zhuang X. Myops-net: Myocardial pathology segmentation with flexible combination of multi-sequence cmr images. Medical Image Analysis 2023;84:102694.

[8] Ma J, Chen J, Ng M, Huang R, Li Y, Li C, Yang X, Martel AL. Loss odyssey in medical image segmentation. Medical Image Analysis 2021;71:102035.

Address for correspondence:

Farheen Ramzan
Department of Computer Science, University of Sheffield
Regent Court, 211 Portobello, Sheffield S1 4DP, UK
framzan1@sheffield.ac.uk