

Stacked and Parallel U-Nets with Multi-Output for Myocardial Pathology Segmentation

Zhou Zhao*, Nicolas Boutry^[0000-0001-6278-4638], and Élodie Puybureau^[0000-0002-2748-6624]

EPITA Research and Development Laboratory (LRDE), Le Kremlin-Bicêtre, France

*Corresponding author

zz@lrde.epita.fr, nicolas.boutry@lrde.epita.fr,
elodie.puybureau@lrde.epita.fr

Abstract. In the field of medical imaging, many different image modalities contain different information, helping practitioners to make diagnostic, follow-up, etc. To better analyze images, mixing multi-modalities information has become a trend. This paper provides one cascaded UNet framework and uses three different modalities (the late gadolinium enhancement (LGE) CMR sequence, the balanced- Steady State Free Precession (bSSFP) cine sequence and the T2-weighted CMR) to complete the segmentation of the myocardium, scar and edema in the context of the MICCAI 2020 myocardial pathology segmentation combining multi-sequence CMR Challenge dataset (MyoPS 2020). We evaluate the proposed method with 5-fold-cross-validation on the MyoPS 2020 dataset.

Keywords: Deep Learning · Myocardial Pathology · Segmentation · UNet.

1 Introduction

The assessment of myocardial viability is essential for diagnosis and follow-up of patients suffering from myocardial infarction (MI) [17, 16]. However, many different images modalities in the field of medical imaging are available and are complementary. Late gadolinium enhancement (LGE) cardiac magnetic resonance (CMR) sequence which visualizes MI, T2-weighted CMR (imaging the acute injury and ischemic regions) and balanced-Steady State Free Precession (bSSFP) cine sequence (which captures cardiac motions and presents clear boundaries) are examples of such imaging modalities. Therefore, making a better use of the information in these different modalities has become a research focus. In recent years, many semi-automated and automated methods have been proposed for multi-modal medical image segmentation using deep learning-based methods, such as convolutional neural networks (CNNs) [8] and fully convolutional networks (FCNs) [9] especially the U-Net architecture [11]. For example, Guo [3, 4] proposed a conceptual image fusion architecture for supervised biomedical image analysis. They designed and implemented an image segmentation system based on deep CNNs to contour the lesions of soft tissue sarcomas using multimodal images by fusing the information derived from different modalities.

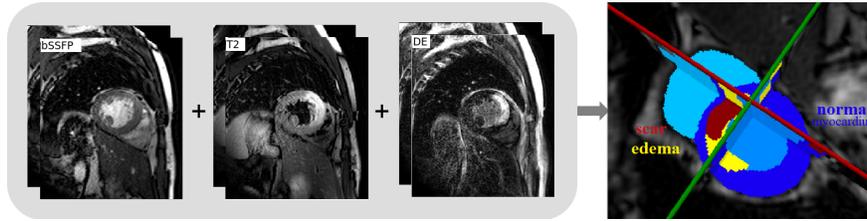


Fig. 1: Myocardial pathology, the picture is from MyoPS2020 challenge ¹.

Although we can use multi-modal information to improve the myocardial pathology segmentation, class imbalance remains a problem to tackle. Network overfitting is common in the field of medical imaging because of the relatively small size of handled datasets. Data augmentation is classically used in the pre-processing stage to overcome this limitation, and weighted loss functions are designed. For example, Zhao et al. [15, 10] used data augmentation by rotating and flipping the heart segmentations to reduce the impact of overfitting. Zhao et al. [14] proposed an automated data augmentation method for synthesizing labeled medical images, which provided significant improvements over state-of-the-art methods for one-shot biomedical image segmentation. Sudre et al. [13] proposed the generalized dice to solve the problem of highly unbalanced segmentations. Abraham et al. [1] proposed a generalized focal loss function based on the Tversky index to address the issue of data imbalance in medical image segmentation. Examples of data augmentation methods to overcome this issue can be found in [2, 12, 6, 5, 7]. However, datasets obtained through data augmentation are strongly correlated with the original datasets. Therefore, the proportion of negative samples remains significantly larger than the proportion of positive samples after data augmentation. Thus, data augmentation does not reduce the risk of overfitting. For the proposed improved loss function can effectively reduce the issues of class imbalance, it does not fundamentally address the problems caused by the lack of datasets.

Therefore, in this paper, in order to segment myocardial pathology (see Fig. 1), we begin with a segmentation of the anatomical tissue (left ventricle (LV), right ventricle (RV), whole heart (WH), myocardium (myo)) around it, and then let the network learn a relationship between these segmentation results to obtain the myocardial pathology. Compared with direct segmentation of myocardial pathology, the effect of class imbalance can be reduced by the segmentation of surrounding anatomical tissues, because it helps the network to focus on the small lesions regarding to the surrounding tissues.

2 Methodology

2.1 Overview of Network Architecture

We propose a hybrid network (see Fig. 2) using 5 UNet [11] to segment myocardial pathology. Our network is composed of three UNet named **UNet1** and two named **UNet2**. The main difference between **UNet1** and **UNet2** is number of

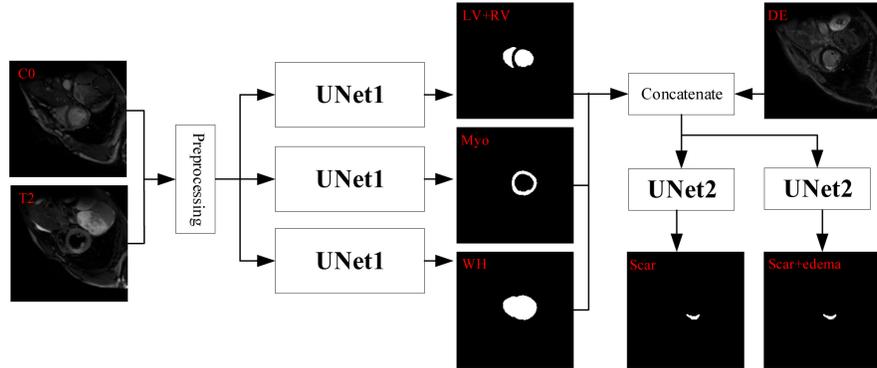


Fig. 2: Global overview of the proposed method.

Table 1: The structural configuration of UNet.

Layers	Input size		Operation	Kernel	Stride	Regul.	Output size	
	UNet1	UNet2					UNet1	UNet2
Input image	(240,240,2)	(240,240,4)	-	-	-	-	(240,240,2)	(240,240,4)
C1	(240,240,2)	(240,240,4)	[Conv2d+relu]*2	3	1	L2	(240,240,64)	(240,240,8)
C2	(240,240,64)	(240,240,8)	Maxpooling2d	2	-	-	(120,120,64)	(120,120,8)
C3	(120,120,64)	(120,120,8)	[Conv2d+relu]*2	3	1	L2	(120,120,128)	(120,120,16)
C4	(120,120,128)	(120,120,16)	Maxpooling2d	2	-	-	(60,60,128)	(60,60,16)
C5	(60,60,128)	(60,60,16)	[Conv2d+relu]*2	3	1	L2	(60,60,256)	(60,60,32)
C6	(60,60,256)	(60,60,32)	Maxpooling2d	2	-	-	(30,30,256)	(30,30,32)
C7	(30,30,256)	(30,30,32)	[Conv2d+relu]*2+Dropout	3	1	L2	(30,30,512)	(30,30,64)
C8	(30,30,512)	(30,30,64)	Maxpooling2d	2	-	-	(15,15,512)	(15,15,64)
C9	(15,15,512)	(15,15,64)	[Conv2d+relu]*2+Dropout	3	1	L2	(15,15,1024)	(15,15,128)
O1	(240,240,2)	(240,240,2)	Sigmoid	-	-	-	(240,240,1)	(240,240,1)

filters as shown in Table. 1: the number of filters of **UNet1** is [64 128 256 512 256 128 64] and the number of filters of **UNet2** is [8 16 32 64 32 16 8]. Their framework is same. It consists of the classical two parts of the UNet network as shown in Fig. 3: a down-sampling part and an up-sampling part, and shortcut connections between the two parts to fuse high-level features and low-level features. **UNet1** is used to segment the anatomical tissue around myocardial pathology and obtain three segmentation results: LV+RV, Myo, and WH. **UNet2** is used to segment myocardial pathology by learning the relationships between the surrounding anatomical tissue and the pathological ones. Since the lesions are very small and unbalanced, we reduce the number of filters of **UNet2** in order to reduce the impact of overfitting.

3 Experimental Results

Dataset Description. We evaluate our method on the myocardial pathology segmentation combining multi-sequence CMR ² dataset (MyoPS 2020). Its aim is to segment myocardial pathology, especially scar (infarcted) and edema regions.

² <http://www.sdspeople.fudan.edu.cn/zhuangxiahai/0/MyoPS20/index.html>

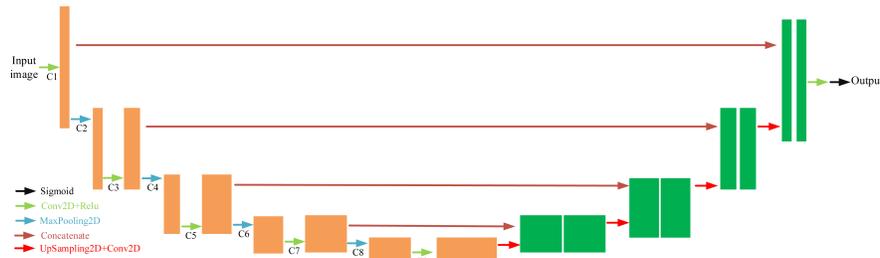


Fig. 3: Architecture of networks.

It contains 45 cases of multi-sequence CMR (25 cases for training and 20 cases for testing). Each case refers to a patient with three sequence CMR, i.e., LGE, T2 and bSSFP CMR. The slice spacings of multi-sequence CMR volume range from 11.999 mm/pixel to 23.000 mm/pixel, while in-plane resolution ranged from 0.729 mm/pixel to 0.762 mm/pixel. The average sizes: $482 \times 479 \times 4$ pixels.

Preprocessing and Postprocessing. We cropped each slice to 240×240 pixels and we do not use data augmentation. The pre-processing begins with a Gaussian normalization. For post-processing, we pad with zeros to get back a initial width and height of a slice.

Implementation and Experimental Setup. We implemented our experiments on Keras/TensorFlow using a NVidia Quadro P6000 GPU. We used five different loss functions for training the network and used sigmoid to get a probability distribution of the left and right ventricle, myocardium, whole heart, scar and edema, and scar, respectively (as shown in Fig. 2). Adam optimizer (batch-size = 1, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 0.001$, lr = 0.0001) and did not use learning rate decay. We trained the network during 300 epochs.

Training Step. First, we kept weight of **UNet2** unchanged, which means **UNet2** was not trained at the beginning, then we trained **UNet1**. After finished the train of **UNet1**, we kept weight of **UNet1** unchanged, then trained **UNet2**.

Evaluation Methods. One metric is used to evaluate our method: dice coefficient (DC) to evaluate the regions of myocardial pathology.

3.1 Segmentation Results

As shown in Table. 2, we evaluate the proposed method with 5-fold-cross-validation. We obtain a mean DC of 92.3% on WH, 84.9% on LV+RV, and 84.7% on Myo by **UNet1**. Without using data augmentation, based on the original dataset, we obtain a higher segmentation accuracy, which lays the foundation for the subsequent segmentation of myocardial pathology. Finally, we obtain a mean DC of 20.6% on edema, 51% on scar by **UNet2**. We used the trained network to predict the testset (20 cases) and received the evaluation of our prediction results from the MyoPS2020 organizer: the mean DC of 58.6% on scar and the mean DC of 63.9% on scar and edema.

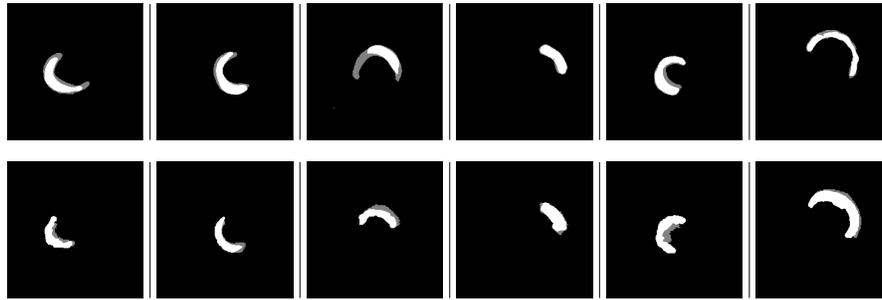
Table 2: Evaluation results on 5-fold-cross-validation.

Patient	101-105	106-110	111-115	116-120	121-125	Average	Test datasets
Edema	0.284	0.153	0.189	0.122	0.280	0.206	–
Scar	0.473	0.496	0.515	0.464	0.602	0.510	0.586
Myo	0.844	0.852	0.811	0.859	0.869	0.847	–
LV+RV	0.818	0.854	0.812	0.897	0.864	0.849	–
WH	0.925	0.937	0.876	0.918	0.959	0.923	–

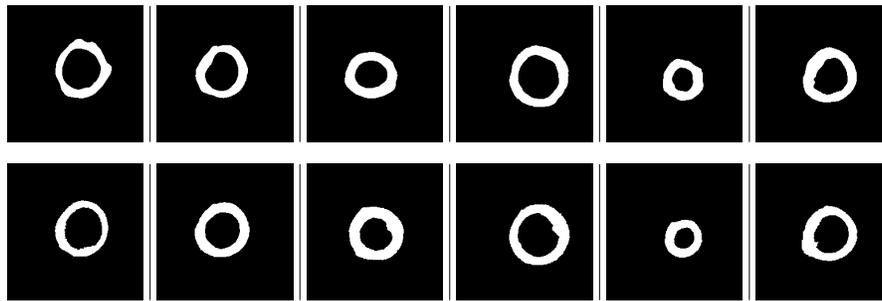
As shown in Fig. 4, for the segmentation results of whole heart, left and right ventricle, and myocardium, as the number of positive samples continues to decrease, the segmentation accuracy is also decreasing, and false segmentation is mainly concentrated at the boundary, which is mainly because ambiguities often appear near the boundaries of the target domains due to tissue similarities. For the segmentation results of edema and scar, the poorly segmentation result is not only on the boundary, but also in regions. In the original dataset, edema does not exist in many slices, which further leads to a reduction in the effective dataset for edema, therefore, the segmentation network is very difficult to segment edema.

4 Conclusion

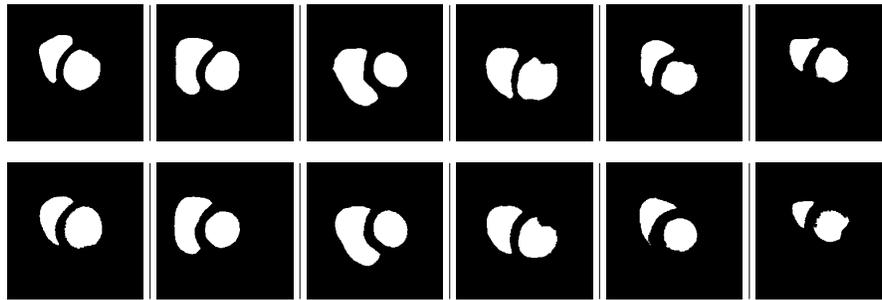
In this paper, we propose a way of reverse thinking, not to segment the myocardial pathology directly, but to learn a relationship between the surrounding normal tissue and it by designing one stacked and parallel UNets with multi-output framework. We evaluate the proposed method with 5-fold-cross-validation on the MICCAI 2020 myocardial pathology segmentation combining multi-sequence CMR Challenge dataset (MyoPS 2020) and achieve a mean DC of 20.6%, 51% on edema and scar, respectively. The computation time of the entire pipeline is less than 3 seconds for an entire 3D volume, making it usable for clinical practice. However, the segmentation accuracy of myocardial pathology is affected by the segmentation accuracy of surrounding normal tissues. Therefore, in our future work, we will continue to study the relationship between the surrounding normal tissue and myocardial pathology and improve the segmentation accuracy of surrounding normal tissues.



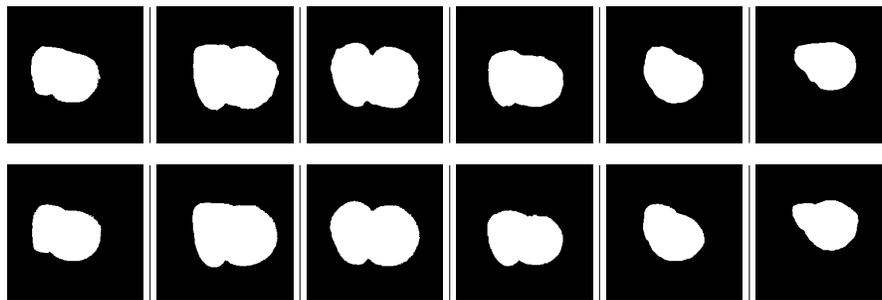
(b) Edema and scar. Scar is in white. Top = segmentation, bottom = Ground Truth



(d) Myocardium. Top = segmentation, bottom = Ground Truth



(f) Left and right ventricle. Top = segmentation, bottom = Ground Truth



(h) Whole heart. Top = segmentation, bottom = Ground Truth

Fig. 4: Qualitative segmentation results.

References

- [1] Abraham, N., Khan, N.M.: A novel focal tvsky loss function with improved attention u-net for lesion segmentation. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). pp. 683–687. IEEE (2019)
- [2] Eaton-Rosen, Z., Bragman, F., Ourselin, S., Cardoso, M.J.: Improving data augmentation for medical image segmentation (2018)
- [3] Guo, Z., Li, X., Huang, H., Guo, N., Li, Q.: Medical image segmentation based on multi-modal convolutional neural network: Study on image fusion schemes. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 903–907. IEEE (2018)
- [4] Guo, Z., Li, X., Huang, H., Guo, N., Li, Q.: Deep learning-based image segmentation on multimodal medical imaging. *IEEE Transactions on Radiation and Plasma Medical Sciences* **3**(2), 162–169 (2019)
- [5] Hashemi, S.R., Salehi, S.S.M., Erdogmus, D., Prabhu, S.P., Warfield, S.K., Gholipour, A.: Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: Application to multiple sclerosis lesion detection. *IEEE Access* **7**, 1721–1735 (2018)
- [6] Hussain, Z., Gimenez, F., Yi, D., Rubin, D.: Differential data augmentation techniques for medical imaging classification tasks. In: *AMIA Annual Symposium Proceedings*. vol. 2017, p. 979. American Medical Informatics Association (2017)
- [7] Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J., Ayed, I.B.: Boundary loss for highly unbalanced segmentation. In: *International conference on medical imaging with deep learning*. pp. 285–296 (2019)
- [8] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**(11), 2278–2324 (1998)
- [9] Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *Proc. of CVPR*. pp. 3431–3440 (2015)
- [10] Puybureau, É., et al.: Left atrial segmentation in a few seconds using fully convolutional network and transfer learning. In: *Proc. of the Intl. Workshop on STACOM*. LNCS, vol. 11395, pp. 339–347. Springer (2018)
- [11] Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: *Proc. of MICCAI*. LNCS, vol. 9351, pp. 234–241. Springer (2015)
- [12] Shin, H.C., Tenenholtz, N.A., Rogers, J.K., Schwarz, C.G., Senjem, M.L., Gunter, J.L., Andriole, K.P., Michalski, M.: Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: *International workshop on simulation and synthesis in medical imaging*. pp. 1–11. Springer (2018)
- [13] Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J.: 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*, pp. 240–248. Springer (2017)
- [14] Zhao, A., Balakrishnan, G., Durand, F., Gutttag, J.V., Dalca, A.V.: Data augmentation using learned transformations for one-shot medical image segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 8543–8553 (2019)

- [15] Zhao, Z., et al.: A two-stage temporal-like fully convolutional network framework for left ventricle segmentation and quantification on MR images. In: Proc. of the Intl. Workshop on STACOM. LNCS, vol. 12009, pp. 405–413. Springer (2019)
- [16] Zhuang, X.: Multivariate mixture model for cardiac segmentation from multi-sequence mri. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 581–588. Springer (2016)
- [17] Zhuang, X.: Multivariate mixture model for myocardial segmentation combining multi-source images. *IEEE transactions on pattern analysis and machine intelligence* **41**(12), 2933–2946 (2019)