Efficient 3D deep learning for myocardial diseases segmentation

Khawla Brahim[1,2,3], Abdul Qayyum[2], Alain Lalande[2], Arnaud Boucher[2], Anis Sakly[3], and Fabrice Meriaudeau[2]

[1] National Engineering School of Sousse, University of Sousse, Tunisia
[2] ImViA EA 7535 laboratory, University of Burgundy, France
{Khawla.Brahim , Abdul.Qayyum , Alain.Lalande , Arnaud.Boucher , Fabrice.Meriaudeau}@u-bourgogne.fr
[3] LASEE laboratory, National Engineering School of Monastir, University of Monastir, Tunisia
sakly_anis@yahoo.fr

Abstract. Automated myocardial segmentation from late gadolinium enhancement magnetic resonance images (LGE-MRI) is a critical step in the diagnosis of cardiac pathologies such as ischemia and myocardial infarction. This paper proposes a deep learning framework for improved myocardial diseases segmentation. In the first step we propose an encoder-decoder segmentation network that generates myocardium and cavity segmentations from the whole volume then followed by a 3D U-Net based on Shape prior identifies myocardial infarction and MVO segmentations from the encoder-decoder prediction. The proposed network, achieves good segmentation performance, as computed by average dice ratio over all predicted substructures, respectively : 'Myocardium': 96.29%, 'Infarctus': 76.56%, 'No-reflow': 93.12% on our validation EMIDEC dataset consisting of LGE-MRI volumes of 16 patients extracted from the training data.

Keywords: LGE-MRI · Myocardial Infarction · Deep Learning.

1 Introduction

According to the World Health Organization (WHO) [1], Myocardial Infarction (MI) is one of the main cause of death globally. It essentially develops when oxygen-rich blood flow to the myocardium is suddenly interrupted [2]. However, when revascularization fails, permanent microvascular obstruction phenomenon (MVO, also known as No-reflow) can occur in scar regions. Efficient quantification of infarcts and MVO is essential for diagnosis and therapy planning.

Myocardial Scar Segmentation aims to accurately recognizing myocardial scars areas. Previous prevalent scar segmentation works were often performed using thresholding-based methods, such as the n-standard deviations (n-SD) [3], the full-width at half-maximum (FWHM) [4] and the region growing [5], which are responsive to the regional intensity variation. However, these algorithms frequently require a prior knowledge of the expert myocardial location determined

by its epicardial and endocardial annotations, to delineate the areas of interest for segmentation [6].

Manual MI delineation is time-consuming and prone to inter and intra-observer variations. Hence, there is a need for accurate and automatic segmentation models to ease the work load of medical experts. Deep learning-based methods for medical image segmentation play an important role in cardiac function analysis and follow up of different diseases due to their feature extraction effectiveness. Deep architectural betterment has been a target of several scientists for diverse works. U-Net based networks [7] have often been used. 2D U-Net network demonstrated impressive performances for valuable segmentation of myocardium structure on the ACDC 2018 challenge [8]. Fahmy et al. [9], used the U-Net based method to delineate the myocardium and the scars from LGE images obtained from subjects with hypertrophic cardiomyopathy (HCM). Fatemeh Zabihollahy et al. proposed 2D U-Net for powerful segmentation of myocardial regions from 3D LGE-MRI [10]. Applying 3D Fully Convolutional Networks (FCN) which integrate 3D context across different slices improve estimating disease diagnosis [11]. Interestingly, frequent 3D FCN works achieve promising performances in segmenting cardiovascular volumes with robust 3D consistency [12,13]. Xu et al. [14] provided an RNN method for infarction assessment which exploits motion patterns to accurately segment MI region from cine MR image sequences.

2 Material and Method

2.1 Datasets and Pre-processing

The dataset [15] is supplied by the EMIDEC segmentation challenge and consists of 150 volumes. The ground truth annotation includes five labels: background (0), cavity(1), normal myocardium (2), myocardial infarction (3) and no-reflow (4). The labeled scans are split into training (80 patients), validation (20 patients) and test (50 patients) subsets. We first cropped original volumes to a normalized set. Supplementary empty slices are added to maintain size fixed, resulting in Nifti images of shape $96 \times 96 \times 16$ for all present subjects.

2.2 3D Proposed Model

Given V = $\{V_1, V_2, ..., V_n\}$ a set of 3D LGE-MRI input volumes, our approach is trained end-to-end on each of them, and learns to predict 3D segmentation for the whole volume. To reach this goal, the proposed method consists of two major steps: we train the first proposed network, to learn the myocardium regions. Then, the extracted attributes from the 3D pre-trained Autoencoder model are transferred to 3D U-Net to segment myocardial diseases and increase the model performance. Details of each part are elucidated in the following paragraphs.

3D Myocardium Segmentation The proposed model is deigned based on the concept of encoder-decoder with skip connections. In encoding or analysis path, the proposed Inception-Res block has been introduced with convolutional block attention module (CBAM) and in decoding or synthesis path, the proposed EDP (expansion, depthwise and projection layer block) module has been presented after 2D upsampling layer. The attention module has been used in skip connection that caters information at every block from encoder to decoder side. The number of channels is doubled at each Inception-Res block and input size of feature maps are reduced by half using depthwise convolution layer in analysis path. We have used progressive feature extraction approach at encoder side, the number of Inception-Res block are increased progressively at each stage of the decoder side. The first encoder block used one Inception-Res block, similarly, second, third, and fourth used 2,3,4 number of Inception-Res block respectively. Similarly, in synthesis path, the size of feature maps increases after 2D up-sampling layer and original size of training images will return at output in final layer.

We proposed a modified inception module in encoder side of our proposed model. In inception residual block, the features maps are aggregated from various branches using kernels of different sizes that make the network wider and having the capacity to learn more features. The residual connections provide easy learning with reference to the input feature maps, instead of learning an unreferenced function [16]. The proposed model is shown in Fig. 1.

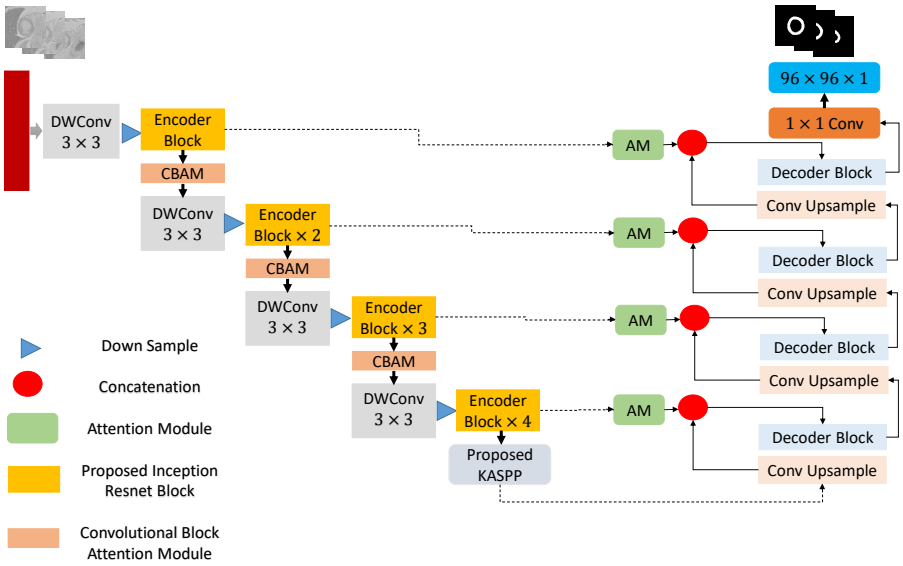
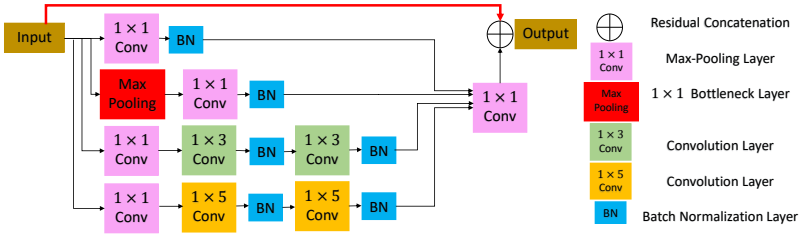


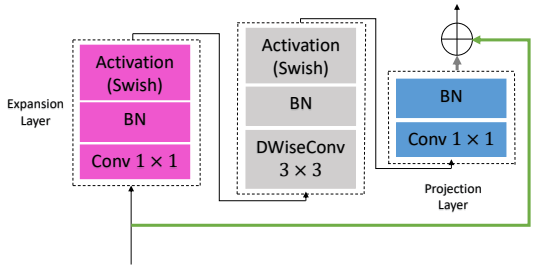
Fig. 1: Proposed model based on Inception-ResNet and EDP blocks.

In the decoder side, we have proposed EDP (expansion, depthwise and projection layer block). Similarly, as in encoder, the expansion layer increases the number of feature maps and in projection layer decreases the feature maps with some regularization layer such as batch normalization and activation. The complete layer structure for the decoder is shown in Fig. 2b.

The Fig. 2a shows the proposed modified Inception-Res block. As compared to the original Inception-Res architecture, batch normalization (BN) layer has been introduced after each convolutional layer except for bottleneck layers and as a second modification we are using 1×1 and 3×3 kernel, and also introduced 5×5 kernel branch as inspired by the DeepLab [17]. Batch normalization layer produced smooth training and can avoid gradient vanishing while retaining convolutional layers. The feature maps are aggregated by convolving with three kernels, namely 1×1, 3×3 and 5×5. The 3×3 and 5×5 kernels are further reduced into 1×3, 3×1, 1×5 and 5×1 to minimize the number of parameters.



(a) Proposed Inception-ResNet Block.



(b) Decoder Block.

Fig. 2: Proposed Inception-ResNet Block and EDP (expansion, depth-wise and projection) Decoder Block.

Assuming that x_l is the output of the l^{th} layer, $c_{(n\times n)}(.)$ is a $n \times n$ kernel convolutional layer, $c_b(.)$ represents the batch-normalization layer and 1×1 Conv denotes the bottleneck layer; the output of each Inception-Res block module from the decoder path is given in Eq. 1.

$$x_{l+1} = c_{1\times1}(c_{1\times1}(x_l).c_b(c_{3\times3}(c_{1\times1}(x_l))).c_b(c_{3\times3}(c_b(c_{3\times3}(c_{1\times1}(x_l))))).c_b(c_{5\times5}(c_b(c_{5\times5}(c_{1\times1}(x_l))))) + x_l \tag{1}$$

Ye Huang et al. [18] presented a kernel-sharing atrous convolutional (KSAC) layer in atrous spatial pyramid pooling (ASPP) module. The 3×3 kernel is shared with atrous convolutional layers with different dilation rates. In this paper, we have extended KSAC based ASPP module and combined different features extracted from the down-sampling path with the various scale features (five scales) in the KSAC based ASPP shown in Fig. 3. The proposed KSAC based ASPP (later noted KASPP) module captures features from low level as well as features from different down-sampled layers to obtain texture and position information from encoder side feature maps.

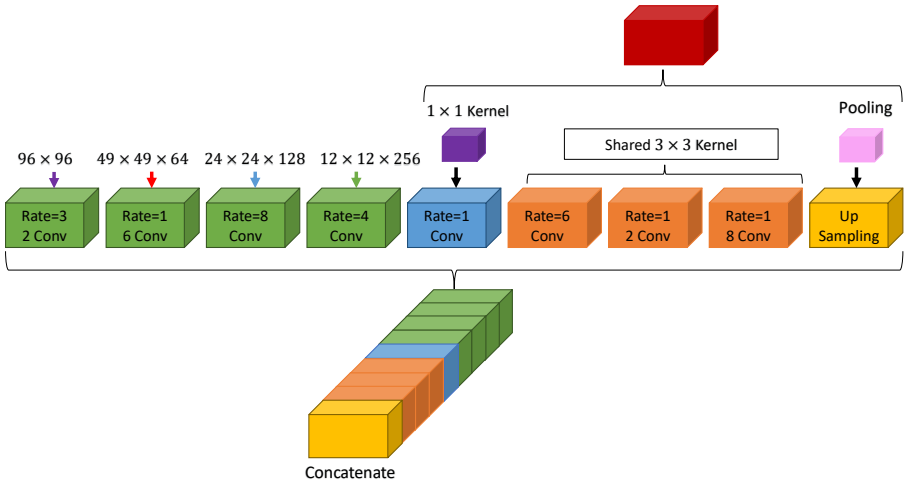


Fig. 3: The proposed K-atrous spatial pyramid pooling layer module.

3D Myocardial diseases Segmentation Experimental results proved that extracting volume patches of size $12 \times 12 \times 12$ pixels^3 from the training dataset attains the best results for the segmentation of diseased myocardial tissues.

The U-Net architecture is U-shaped model which firstly aims to catch more high-level features through an ensemble of convolutional and max pooling layers. Then the feature maps are up-sampled to recover the segmentation maps at the original spatial dimension. Therefore, the concatenation of feature maps of same resolution in the decoder path produces a promising medical segmentation.

As shown in Fig.4, our 3D network incorporates 3D U-Net with a Super Resolution (SR) module to constrain prior knowledge shape. 3D Autoencoder

focuses on how to accurately encode and reduce the original volume that can be rebuilt from the encoded representation. Hence a pre-trained 3D Autoencoder, is effective to regularize the generated result into a realistic shape. Pre-trained 3D Autoencoder is bound to the 3D U-Net and takes the segmented scan as input. A regularization term is established for restraining the segmentation result. The final loss function is defined in Eq. 2:

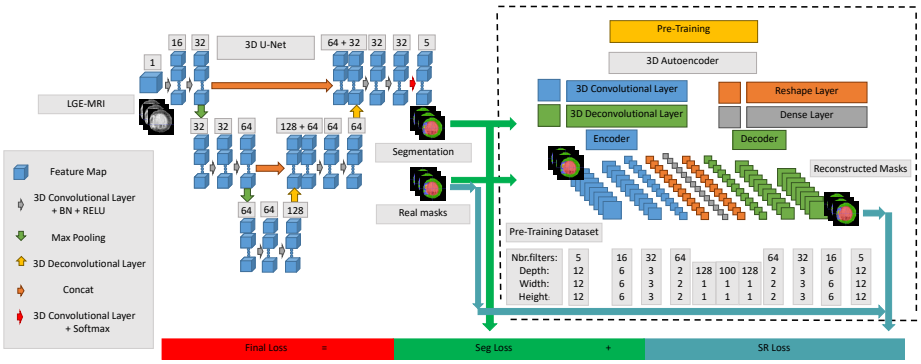


Fig. 4: Schematic representation of our approach for damaged myocardial segmentation.

$$L_{Final} = L_{Seg} + \lambda_{SR} \times L_{SR} \tag{2}$$

(Where L_{Seg} is the cross entropy loss function, λ_{SR} is the regularization term and L_{SR} is the L2 loss function which is determined from Frobenius norm Eq. 3. We choose $\lambda_{SR} = 10^{-2}$.)

$$L_{SR} = \sum_{i=1}^{n} ||RP_i - RG_i||_F^2 \tag{3}$$

(Where n is the number of training volumes, RG_i represents the reconstructed ground truth, RP_i denotes the reconstructed segmentation results and $||.||_F$ indicates the Frobenius norm of an $m \times n$ matrix.)

3 Results and Discussion

Each voxel was better determined through majority voting and morphological mathematics (erosion and dilatation) of class tissues acquired on patches. We used Dice Coefficient (DSC), Hausdorff Distance (HD), Volume Difference (AVD) and Absolute Volume Difference Rate according to volume of myocardium (AVDR) as evaluation metrics for all myocardial regions.

Table 1: Quantitative study for myocardial segmentation. Best values are represented in bold font.

Method	Metrics	Structures		
		Myocardium	Infarctus	No-reflow
3D U-Net	DSC %	95.71	74.98	68.61
	AVD mm^3	295.50	474.12	55.75
	HD mm	4.57	-	-
	AVDR %	-	9.06	0.97
3D U-Net + 3D Autoencoder + Post-processing	DSC %	**96.29**	**76.56**	**93.12**
	AVD mm^3	**270.00**	**234.1**	**26.69**
	HD mm	**3.77**	-	-
	AVDR %	-	**4.92**	**0.59**

We report in Table 1 the summary of comparative evaluation on validation set (including 16 whole volumes), showing the pertinence of 3D Autoencoder and Post-processing. Majority voting technique and morphological mathematics Post-processing are applied to increase sensitivity for quantifying scarred areas. Our proposed method successfully outperforms the baseline 3D U-Net model (average dice 88.66 % vs. 79.77 %).

To show the impact of 3D Autoencoder and Post-processing, Fig.5, displays exemplary visualization of the segmentation myocardial structures on two validation subjects. Gold standard and segmented volumes attained using our 3D network and 3D U-Net are presented in the matching row. These results demonstrate the performance of our developed framework in segmenting different structures of interest by aiming attention at relevant regions.

4 Conclusion

Automated myocardial tissue segmentation is paramount for diagnosis of cardiac diseases. In this paper, we present an end-to-end deep learning network. A modified proposed atrous convolutional layers, EDP and Inception-Res blocks are integrated to catch more high-level features and retain more finer and coarser information. The proposed myocardial segmentation adopts Pre-trained 3D Autoencoder with 3D U-Net, improving the segmentation efficiency. Our experimental results prove that this approach provided the best accuracy for myocardial segmentation and can find its pertinence in different medical imaging tasks based on deep learning.

References

1. Mackay, J., & Mensah, G. A.: The atlas of heart disease and stroke. In World Health Organization. (2004).

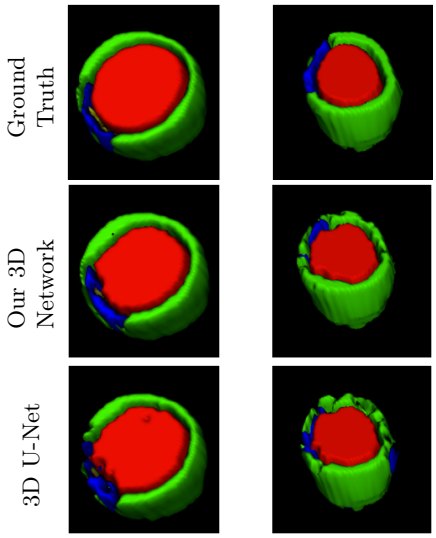


Fig. 5: Exemplary results of 3D segmentation on two patients on the EMIDEC Challenge dataset. The first, second and third rows represent ground truth annotation, our 3D proposed network and 3D U-Net segmentation tissue. Our approach reaches qualitatively better segmentations than 3D U-Net.

2. Surawicz, B., & Knilans, T.: Chou's Electrocardiography in Clinical Practice E-Book: Adult and Pediatric. In Elsevier Health Sciences. (2008).
3. Kim, R. J., et al.: Relationship of MRI delayed contrast enhancement to irreversible injury, infarct age, and contractile function. In Circulation, 100, pp. 1992–2002. (1999).
4. Amado, L. C., et al.: Accurate and objective infarct sizing by contrast-enhanced magnetic resonance imaging in a canine myocardial infarction model. In Journal of the American College of Cardiology, **44**(12), pp. 2383–2389. (2004).
5. Albà, X., i Ventura, R. M. F., Lekadir, K., & Frangi, A. F.: Healthy and scar myocardial tissue classification in DE-MRI. In International Workshop on Statistical Atlases and Computational Models of the Heart. Springer, Berlin, Heidelberg, pp. 62–70. (2012, October).
6. Carminati, M. C., Boniotti, C., Fusini, L., Andreini, D., Pontone, G., Pepi, M., & Caiani, E. G: Comparison of image processing techniques for nonviable tissue quantification in late gadolinium enhancement cardiac magnetic resonance images. In Journal of thoracic imaging, **31**(3), pp. 168–176. (2016).
7. Ronneberger, O., Fischer, P., & Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, Springer, pp. 234–241. (2015).
8. Bernard, O., et al.: Deep learning techniques for automatic MRI cardiac multistructures segmentation and diagnosis: Is the problem solved?. In IEEE Transactions on Medical Imaging. (2018).
9. Fahmy, A. S., et al.: Automated cardiac MR scar quantification in hypertrophic cardiomyopathy using deep convolutional neural networks. In JACC: Cardiovas-

cular Imaging, **11**(12), pp. 1917–1918. (2018).
10. Zabihollahy, F., White, J. A., & Ukwatta, E.: Fully automated segmentation of left ventricular myocardium from 3D late gadolinium enhancement magnetic resonance images using a U-net convolutional neural network-based model. In Medical Imaging 2019: Computer-Aided Diagnosis (10950, pp. 109503C), International Society for Optics and Photonics. (2019, March).
11. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In International conference on medical image computing and computer-assisted intervention, Springer, Cham. pp. 424–432. (2016, October).
12. Dou, Q., Yu, L., Chen, H., Jin, Y., Yang, X., Qin, J., & Heng, P. A.: 3D deeply supervised network for automated segmentation of volumetric medical images. In Medical image analysis, vol. 41, pp. 40–54. (2017).
13. Yu, L., Yang, X., Qin, J., & Heng, P. A.: 3D FractalNet: dense volumetric segmentation for cardiovascular MRI volumes. In Reconstruction, segmentation, and analysis of medical images, Springer, Cham, pp. 103–110. (2016).
14. Xu, C., et al.: Direct delineation of myocardial infarction without contrast agents using a joint motion feature learning architecture. In Medical image analysis, vol. 50, pp. 82–94. (2018).
15. Lalande, A. at al. Emidec: A Database Usable for the Automatic Evaluation of Myocardial Infarction from Delayed-Enhancement Cardiac MRI. In Data, vol. 5, pp. 89. https://doi.org/10.3390/data5040089. (2020).
16. Szegedy, C., et al.: Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9. (2015).
17. Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV), pp. 801–818. (2018).
18. Huang, Y., Wang, Q., Jia, W., & He, X.: See More Than Once–Kernel-Sharing Atrous Convolution for Semantic Segmentation. arXiv preprint arXiv:1908.09443. (2019).