

Transformers-Based Neural Network for Cardiac Infarction Segmentation in Delayed-Enhancement MRI

1st Erwan Lecesne

CNR, UMR 5525, VetAgro Sup, Grenoble INP, TIMC

University of Grenoble

Grenoble, France

Inserm, LTSI - UMR 1099

University of Rennes

Rennes, France

erwan.lecesne@univ-rennes1.fr

2nd Antoine Simon

Inserm, LTSI - UMR 1099

University of Rennes

Rennes, France

0000-0001-6023-6427

3rd Mireille Garreau

Inserm, LTSI - UMR 1099

University of Rennes

Rennes, France

0000-0001-5728-1403

4th Gilles Barone-Rochette

Clinic of Cardiology, Cardiovascular and Thoracic Department

University of Grenoble Alpes

Grenoble, France

0000-0002-1146-0542

5th Céline Fouard

CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC

name of organization (of Aff.)

Grenoble, France

0000-0002-3500-9916

Abstract—Accurately and robustly segmenting myocardial infarction (MI) is crucial for clinical diagnosis of cardiac diseases, treatment and planning. In this study, we propose a novel deep learning model specifically designed for automatic segmentation of MI in Late Gadolinium Enhancement cardiac MRI (LGE-MRI). LGE-MRI is widely used in clinical practice to quantify MI and plays a vital role in treatment decisions. However, due to the presence of high anisotropy and inhomogeneities in LGE-MRI, accurately segmenting the infarcted tissue poses significant challenges.

Our approach introduces the use of U-Net transformers for MI segmentation. By leveraging the power of transformer-based architectures, our model achieves competitive results. We evaluated our method on the 2020 MICCAI EMIDEC challenge dataset and obtained a dice score of 91.33% for myocardium segmentation and 74.41% for infarction segmentation. These results demonstrate the effectiveness of our approach, showcasing its superiority over existing state-of-the-art methods in the EMIDEC challenge.

Index Terms—Late gadolinium magnetic resonance imaging, Left ventricle scar, Segmentation, Transformers network

I. INTRODUCTION

Late Gadolinium Enhancement Magnetic Resonance Imaging (LGE-MRI) is commonly employed to detect and assess infarcted areas following a myocardial infarction (MI). Its goal is to provide a contrast between the infarct tissues and healthy myocardium, enabling precise characterization of the scar's location and size. The ratio of infarcted tissue to healthy tissue is a crucial factor in determining appropriate treatment

during routine clinical practice [1]. Accurately delineating the infarcted area is thus essential for optimal clinical diagnosis and treatment planning. While cardiologists have expertise in manually segmenting the infarction area in MRI-LGE, this process is laborious, time-consuming, and prone to high inter-observer variability [2]. Therefore, developing a fully automatic infarction segmentation approach is necessary to improve clinical routine.

However, segmentation of LGE-MRI presents significant difficulties due to the high intensity and heart shape variations between images. Additionally, the poor inter-slice spatial resolution makes it difficult to leverage volumetric information, limiting the effectiveness of conventional methods [3]. These methods often rely on semi-automatic thresholding techniques such as n -Standard Deviations (n -SD) or Full Width at Half Maximum (FWHM), requiring manual region-of-interest (ROI) selection [4] [5].

Methods based on machine learning, such as conditional random fields (CRF) [3] [6], have demonstrated promising results in fully-automatic segmentation of LGE-MRI. However, these methods often require significant engineering efforts or prior knowledge to achieve high accuracy. On the other hand, deep learning (DL) approaches can automatically discover complex features from the data, making them more flexible and establishing themselves as the new efficient solution [7].

In 2020, the MICCAI EMIDEC challenge aimed to promote research on cardiac structure segmentation in patients with myocardial infarction, providing 100 LGE-MRI images along with ground truth segmentations. The winning method by Zhang [8] utilized a cascaded 2D-3D U-Net architecture [9]. In

The authors acknowledge grant support from the French Agence Nationale de la Recherche (Investissements d'Avenir program, Labex CAMI, ANR-11-LABX-0004).

a 5-fold cross-validation, their approach achieved an average Dice score of 94.40% for the myocardium and 72.08% for the infarction, showing the difficulty of this task.

Zhang’s method consisted of a two-step process, with the initial step concentrating on generating a hard segmentation using convolutional networks exclusively. The resulting Dice scores for the infarction class were lower when compared to the myocardium class. This discrepancy significantly impacted the infarct to healthy tissue ratio. To overcome this limitation and improve infarction segmentation, we propose the use of transformer-based networks [10] [11] in this study. Our approach harnesses the power of transformers, which have been trained and evaluated on the EMIDEC dataset, to accomplish fully-automatic segmentation of both the myocardium and the infarction.

II. METHODS

About the EMIDEC challenge [12], the top-performing method [8]) is based on Fully Convolutional Neural Networks (FCNN). Transformers networks [10] [11] has emerged for image segmentation tasks as new networks. When combined with CNNs, these networks have demonstrated superior performance compared to the standard U-Net architecture in various problem domains [13].

Infarction, being a highly imbalanced class, presents challenges due to the variability in shape, location, and contrast within the myocardium across different image slices. Previous methods have proposed segmenting the myocardium as a first step, but this approach may lead to poor detection of the infarct tissue since error in the first step are propagated to the final result. To address these issues, we propose the use of transformers, which have a larger receptive field and incorporate positional encoding. As illustrated in the Fig.1), the image is partitioned into patches, and the relative positions of each patch to others are encoded and subsequently concatenated with the original information. In the considered case, the infarction is always contained within the myocardium.

A. 3D Unet with transformers

Our goal is, from 3D images of LGE-MRI, to segment the myocardium and the infarct tissue.

In summary, UNETR [13] combines the U-Net structure with transformers in its encoder as illustrated Fig. 2. It divides the input image into patches, applies positional embeddings, and uses a stack of transformer blocks for encoding. The decoder branch, similar to U-Net [9], combines features from the encoder and uses deconvolutional layers for the final output generation. In this network, the input image is a 3D volume represented by a tensor of shape $v \in \mathbb{R}^{h \times w \times d \times c}$, where h , w , d are the height, width, and depth dimensions respectively, and c represents the number of input channels.

To process the input image, it is divided into non-overlapping 3D patches with dimensions (P_H, P_W, P_D) . These patches are then flattened into 1D sequences, resulting in a sequence length of $N = \frac{H \times W \times D \times C}{P_H \times P_W \times P_D}$, where H , W , D denote the respective dimensions of the input image.

To maintain the spatial information of the patches, a 1D learnable vector of positional embeddings is added to the sequence. These positional embeddings encode the relative positions of the patches within the image. For instance, the infarct tissue is consistently connected to the myocardium.

The encoder consists of a stack of twelve transformer blocks. Each transformer block comprises multi-head self-attention (MSA) and a multilayer perceptron (MLP) [14]. The MSA mechanism allows the network to capture dependencies between different elements of the sequence, while the MLP introduces non-linear transformations to further process the encoded features.

Similar to the U-Net architecture, the features extracted by the encoder are concatenated with the decoder branch. The decoder consists of deconvolutional layers, which help reconstruct the spatial information and generate the final output.

B. Implementation of the network

The models were implemented using PyTorch and trained on an NVIDIA A100 Tensor Core GPU with 40GB of memory.

As for the UNETR model, training was conducted with a batch size of 8. The AdamW optimizer, introduced by Loshchilov and Hutter [16], was employed with an initial learning rate of 0.0003. The input patches for UNETR had a size of $(8 \times 8 \times 1)$.

To assess the models’ performance, we employed five-fold cross-validation. Each testing fold consisted of 67% pathological cases and 33% healthy cases.

A simple 2D preprocessing pipeline was applied to all images. This included cropping and padding each image to a size of $128 \times 128 \times 16$, followed by a median filter and histogram equalization.

To augment the training data, we used the TorchIO framework [17]. It provided random geometric transformations such as rotation, scaling, translation, flipping, elastic deformation, as well as random intensity transformations such as blur and gamma adjustments.

Finally, to obtain the final segmentation, a morphology filter specifically designed for connected components filtering was applied to the model’s predictions. This model was trained exclusively on EMIDEC dataset including 100 LGE-MRI with ground truth segmentations (myocardium and fibrosis). The results were evaluated with a set of metrics, with the first one being the Dice score. This score serves as a measure of similarity between two sets of data, typically represented as binary arrays. The Dice score is defined as

$$DC = \frac{2|A \cap B|}{|A| + |B|} \quad (1)$$

The second metric employed is the Hausdorff distance (HSD), which quantifies the spatial separation between two subsets within a metric space. The Hausdorff distance is calculated as:

$$HSD(A, B) = \max\{Dist(A, B), Dist(B, A)\} \quad (2)$$

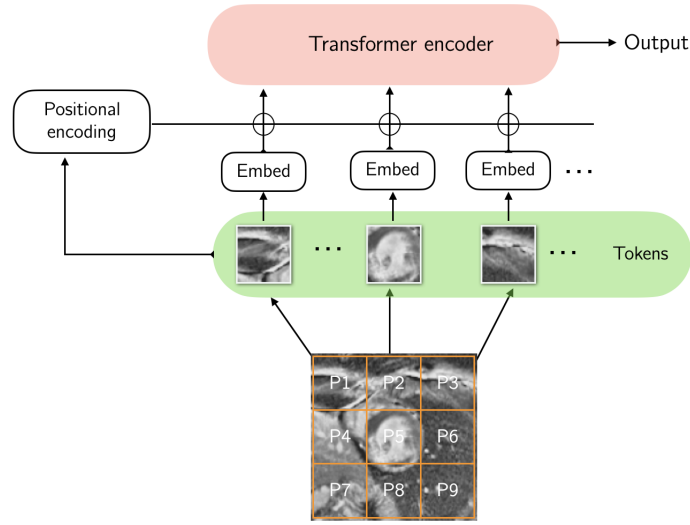


Fig. 1: Patches embedding and positional encoding for the transformers networks. Inspired from [11]

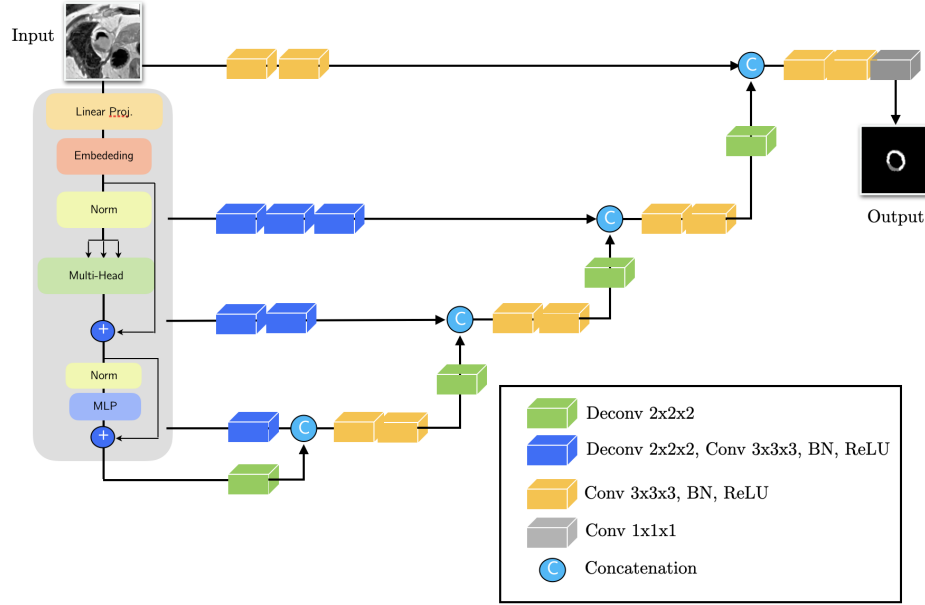


Fig. 2: UNETR architecture. Inspired from [13]

where :

$$Dist(A, B) = \sup_{x \in A} \inf_{y \in B} d(x, y). \quad (3)$$

And the volume difference ratio according to volume of myocardium (only for the infarcted tissue).

III. RESULTS AND DISCUSSION

The results of our method, including various metrics, are summarized in Table I. To evaluate the robustness of our approach, we conducted a k-fold validation. On average, our proposed method achieved a dice score of 91.33% for the myocardium class and 74.41% for the infarction class (see Table I). For the infarction our method outperforms Zhang's method in terms of all the metrics. This is specially the case

for the HSD distance. For the myocardium the Dice was better for Zhang's method (94.4 vs 91.3). The HSD is better with the proposed method, showing the ability of transformer to produce realistic results, globally consistent with the reference. The Fig. 3 illustrates three cases where the dice scores for the myocardium and infarcted tissue are above, close to, and below the average dice score. In the first two cases, the resulting segmentations (Fig. 3, case 1 and case 2) are close to the ground truth (Figure 3f). The segmented edges are smooth, and there is a clear connection between the infarction and myocardium, aligning with the biological reality.

However, the third case (Fig. 3, case 3) presents a challenging scenario. Despite the irregularity in the myocardium, the infarction is accurately located.

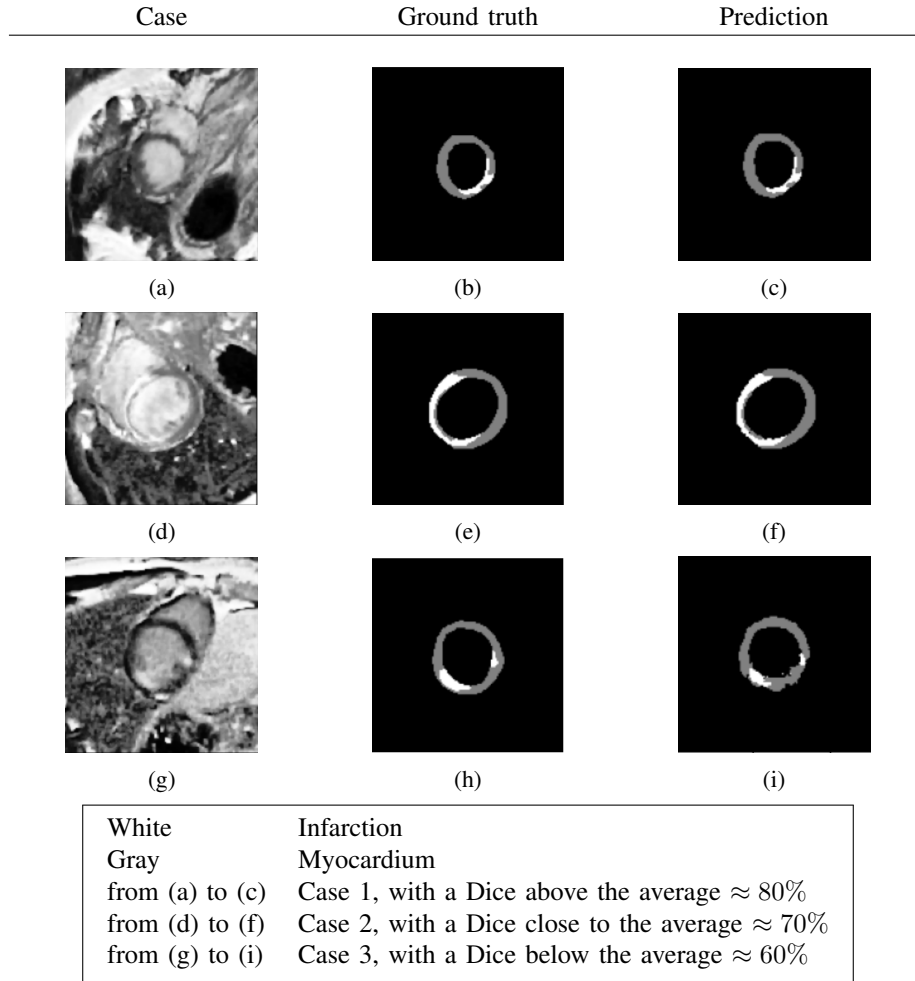


Fig. 3: Examples of three cases with dice scores above, close to, and below the average dice score (approximately 74%) when compared to the ground truth

TABLE I: Comparison study. (Best values in bold font)

Class	Metrics	Methods	
Myocardium		Zhang [8]	Ours
	Dice(%)	94.40	91.33
	HSD(mm)	17.21	5.22
Infarction	Dice(%)	72.08	74.41
	Ratio(%)	3.41	1.26

IV. CONCLUSION

In this paper, we propose the use of transformer networks for automatic segmentation of myocardial infarction from LGE-MRI images. The network is a 3D hybrid CNN with transformers, which enables spatial linking of feature maps, enhancing the network's performance.

Our approach demonstrates promising results in the detection and segmentation of myocardial infarction. Particularly, when compared to the method employed by the winner of the EMIDEC challenge, our method outperforms theirs across all metrics for the infarction. These findings affirm the efficacy and relevance of transformer networks in the domain of medical image segmentation. The futur works will focused on the evaluation of the method on the data acquired in differents clinical centers.

ACKNOWLEDGMENT

The authors acknowledge grant support from the French Agence Nationale de la Recherche (Investissements d'Avenir program, Labex CAMI, ANR-11-LABX-0004).

REFERENCES

- [1] Kim, R., Wu, E., Rafael, A., Chen, E., Parker, M., Simonetti, O., Klocke, F., Bonow, R. & Judd, R. The use of contrast-enhanced magnetic resonance imaging to identify reversible myocardial dysfunction. *New England Journal Of Medicine*. **343**, 1445-1453 (2000)
- [2] Tao, Q., Piers, S., Lamb, H. & Geest, R. Automated left ventricle segmentation in late gadolinium-enhanced MRI for objective myocardial scar assessment. *Journal Of Magnetic Resonance Imaging*. **42**, 390-399 (2015)
- [3] Karim, R., Bhagirath, P., Claus, P., Housden, R., Chen, Z., Karimaghloo, Z., Sohn, H., Rodriguez, L., Vera, S., Albà, X. & Others Evaluation of state-of-the-art segmentation algorithms for left ventricle infarct from late Gadolinium enhancement MR images. *Medical Image Analysis*. **30** pp. 95-107 (2016)
- [4] Kim, R., Fieno, D., Parrish, T., Harris, K., Chen, E., Simonetti, O., Bundy, J., Finn, J., Klocke, F. & Judd, R. Relationship of MRI delayed contrast enhancement to irreversible injury, infarct age, and contractile function. *Circulation*. **100**, 1992-2002 (1999)
- [5] Amado, L., Gerber, B., Gupta, S., Rettmann, D., Szarf, G., Schock, R., Nasir, K., Kraitman, D. & Lima, J. Accurate and objective infarct sizing by contrast-enhanced magnetic resonance imaging in a canine myocardial infarction model. *Journal Of The American College Of Cardiology*. **44**, 2383-2389 (2004)
- [6] Karimaghloo, Z., Shah, M., Francis, S., Arnold, D., Collins, D. & Arbel, T. Automatic detection of gadolinium-enhancing multiple sclerosis lesions in brain MRI using conditional random fields. *IEEE Transactions On Medical Imaging*. **31**, 1181-1194 (2012)
- [7] Chen, C., Qin, C., Qiu, H., Tarroni, G., Duan, J., Bai, W. & Rueckert, D. Deep learning for cardiac image segmentation: a review. *Frontiers In Cardiovascular Medicine*. pp. 25 (2020)
- [8] Zhang, Y. Cascaded convolutional neural network for automatic myocardial infarction segmentation from delayed-enhancement cardiac MRI. *International Workshop On Statistical Atlases And Computational Models Of The Heart*. pp. 328-333 (2020)
- [9] Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. *International Conference On Medical Image Computing And Computer-assisted Intervention*. pp. 234-241 (2015)
- [10] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, Ł. & Polosukhin, I. Attention is all you need. *Advances In Neural Information Processing Systems*. **30** (2017)
- [11] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. & Others An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference On Learning Representations, ICLR 2021*. (2021)
- [12] Lalonde, A., Chen, Z., Decourselle, T., Qayyum, A., Pommier, T., Lorgis, L., Rosa, E., Cochet, A., Cottin, Y., Ginhac, D. & Others Emidec: a database usable for the automatic evaluation of myocardial infarction from delayed-enhancement cardiac MRI. *Data*. **5**, 89 (2020)
- [13] Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H. & Xu, D. Unetr: Transformers for 3d medical image segmentation. *Proceedings Of The IEEE/CVF Winter Conference On Applications Of Computer Vision*. pp. 574-584 (2022)
- [14] Ramchoun, H., Ghanou, Y., Ettaouil, M. & Janati Idrissi, M. Multilayer perceptron: Architecture optimization and training. (International Journal of Interactive Multimedia, 2016)
- [15] Ba, J. & Diederik, K. Adam: A Method for Stochastic Optimization. *Anon. International Conferenceon Learning Representations. SanDego: ICLR 2015*. (2015)
- [16] Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. *ArXiv Preprint ArXiv:1711.05101*. (2017)
- [17] Pérez-García, F., Sparks, R. & Ourselin, S. TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods And Programs In Biomedicine*. **208** pp. 106236 (2021)