# Segmentation of cardiac infarction in delayed-enhancement MRI using probability map and transformers-based neural networks

Erwan Lecesne, Antoine Simon, Mireille Garreau, Gilles Barone-Rochette, Céline Fouard

## Highlights

- Need for segmentation of healthy and infarcted tissues in delayed-enhancement MRI.
- Lack of robustness and contour accuracy of the state of the art methods.
- Use of probability map instead of binary segmentation for a better sensibility and robustness.
- Use of an adapted contours loss function.
- Outperforms the state of the art for infarction segmentation.

# Segmentation of cardiac infarction in delayed-enhancement MRI using probability map and transformers-based neural networks

Erwan Lecesne[a], Antoine Simon[b], Mireille Garreau[c], Gilles Barone-Rochette[d], Céline Fouard[e]

[a] *Univ Rennes, Inserm, LTSI - UMR 1099, , Rennes, 35000, , France*
[b] *Univ Rennes, Inserm, LTSI - UMR 1099, , Rennes, 35000, , France*
[c] *Univ Rennes, Inserm, LTSI - UMR 1099, , Rennes, 35000, , France*
[d] *Clinic of Cardiology, Cardiovascular and Thoracic Department, University Hospital of Grenoble, , Grenoble, 38000, , France*
[e] *Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, , Grenoble, 38000, , France*

**Abstract**

Background and Objective: Automatic segmentation of myocardial infarction is of great clinical interest for the quantitative evaluation of myocardial infarction (MI). Late Gadolinium Enhancement cardiac MRI (LGE-MRI) is commonly used in clinical practice to quantify MI, which is crucial for clinical diagnosis and treatment of cardiac diseases. However, the segmentation of infarcted tissue in LGE-MRI is highly challenging due to its high anisotropy and inhomogeneities.

Methods: The innovative aspect of our work lies in the utilization of a probability map of the healthy myocardium to guide the localization of infarction, as well as the combination of 2D U-Net and U-Net transformers to achieve the final segmentation. Instead of employing a binary segmentation map, we propose using a probability map of the normal myocardium, obtained through a dedicated 2D U-Net. To leverage spatial information, we employ a U-Net transformers network where we incorporate the probability map into the original image as an additional input. Then, To address the limitations of U-Net in segmenting accurately the contours, we introduce an adapted loss function.

Results: Our method has been evaluated on the 2020 MICCAI EMIDEC challenge dataset, yielding competitive results. Specifically, we achieved a

Dice score of 92.94% for the myocardium and 92.36% for the infarction. These outcomes highlight the competitiveness of our approach.

Conclusion: In the case of the infarction class, our proposed method outperforms state-of-the-art techniques across all metrics evaluated in the challenge, establishing its superior performance in infarction segmentation. This study further reinforces the importance of integrating a contour loss into the segmentation process.

*Keywords:* Late gadolinium magnetic resonance imaging, Left ventricle scar, Segmentation, Transformers network, U-Net

## 1. Introduction

Late Gadolinium Enhancement Magnetic Resonance Imaging (LGE-MRI) is widely used for infarcted area detection after a myocardial infarction (MI). It results to a good contrast of the infarct tissues, enabling to characterize the position and size of the scar in the myocardium. The infarct to healthy tissue ratio is an important indicator to choose an adapted treatment in clinical routine [1]. The delineation of the infarcted area is thus a key to an optimal clinical diagnosis and treatment planning. Cardiologists have good experience to delineate the infarction area in MRI-LGE, however a manual segmentation is tedious and time-consuming and may lead to a high inter-observer variability [2]. Therefore a way to meet the clinical needs is to develop a fully-automatic infarction segmentation.

However, LGE-MRI segmentation is a challenging task, due to contrast and heart shape variations between images. Furthermore, the inter-slices spatial resolution is poor, hence it is difficult to exploit a volumic information. This limits the efficiency of conventional methods [3] based on semi-automatic thresholding techniques like n-Standard Deviations [4] (n-SD), Full Width at Half Maximum (FWHM) [5] and expectation–maximization (EM) algorithm coupled with morphology filters (watershed) [6]. A manual region-of-interest (ROI) selection is required for these three techniques.

Machine learning based methods like conditional random fields (CRF) [3] [7], enable a fully-automatic segmentation of LGE-MRI with good results. However, they require an amount of engineering or prior knowledge to achieve a good level of accuracy. On the other hand, deep-learning (DL) based methods discovers automatically complex features from data. It makes DL methods more flexible to use and establishing themselves as the new standard [8].

Among the numerous existing architectures, the U-Net has emerged as a prominent and widely recognized architecture for medical image segmentation [9]. The fully convolutional network (FCN) architecture has a main limitation, it lies in its dependence on fully connected layers, which results in a loss of spatial information and ultimately leads to a decline in overall performance. In contrast, the U-Net architecture is specifically tailored to address the challenges encountered in medical image segmentation tasks and has been extensively used [10].

The primary drawback of FCN stems from its reliance on fully connected layers, which leads to a loss of spatial information and consequently results in a decline in overall performance.

The U-Net architecture consists of two branches connected by skip connections. The encoder branch extracts features and creates a compact representation of the information contained in the image. This compression is primarily achieved through successive convolutions. The decoder branch reconstructs an image from the compact representation with transposed convolutional layers. The skip connections between each stage connect the encoder and the decoder. It ensures the reuse of feature maps of the same dimension from previous layers through concatenation.

Upendra et al. [11] proposed a novel method that combines the U-Net architecture with Cine-MRI registration on LGE-MRI to accurately segment the key cardiac structures, including the myocardium, left ventricle, and right ventricle. The dataset used is the 2019 MS-CMRSeg challenge and the method obtained an average Dice score of 84.73% and 71.49% for the Cine-MRI and LGE-MRI respectively.

Decourt et al. [12] proposed an innovative approach that leverages Generative Adversarial Networks (GANs) to generate precise left ventricle segmentations for pediatric MRI scans. This method offers the distinct advantage of achieving comparable performance with fewer data requirements compared to traditional U-Net approaches.

Cao et al. [13] developed the SwinUnet, a 2D U-Net architecture that exclusively utilizes transformers. This architecture demonstrates remarkable performance on the ACDC dataset, which comprises Cine-MRI scans. Specifically, it achieves a Dice score of 88.55% for the right ventricle, 85.62% for the myocardium, and 95.83% for the left ventricle. However, this network primarily emphasizes 2D slices, whereas in our application, we aspire to harness the potential of both the 2D and 3D dimensions while maintaining convolution as a valuable complement to the transformers.

Zabihollahy et al. [14], [15] introduced a novel approach known as the Cascaded Multi-Planar U-Net (CMPU-Net). This method involves training two subnetworks, each consisting of three U-Nets, utilizing 2D slices extracted from the axial, sagittal, and coronal views of the 3D LGE-MRI scans. This method achieves a Dice score of 85.14% for the myocardium and 88.61% for the infarction. However, in our specific application, the acquired images exhibit significant anisotropy along the Z-axis. This anisotropy results in poor resolution in the Z-axis, which imposes limitations on methods that solely rely on a 3D approach.

In 2020, a MICCAI challenge was proposed to stimulate research on segmentation of cardiac structures of patients with myocardial infarction. It included 100 LGE-MRI images with associated ground truth segmentations. The best ranked Zhang [16] of the 2020 MICCAI EMIDEC challenge dataset [17], introduced a method based on a cascaded 2D-3D U-Net [9] that achieved on a 5 folds cross-validation an average Dice score of 94.4% for the myocardium and 72.08% for the infarction. They were close to the best results which were obtained later by Brahim et al. [18], who did not take part in the challenge.The latter method introduce a shape prior framework composed of a 3D U-Net combined with a 3D auto-encoder with a Dice score of 95.10% and 76.14% for the myocardium and the infarction respectively.

These two methods are based on two steps, provide hard segmentation at the first step and exploit only convolutional networks. They provide rather low Dice scores for the infarction class compared to the myocardium class. This has a strong influence on the infarct to healthy tissue ratio. In order to improve infarction segmentation, in this paper we proposed a refined cascaded framework based on transformers and CNN trained and evaluated on the EMIDEC dataset, for a fully-automatic segmentation of the myocardium and the infarction.

## 2. Methods

In the EMIDEC challenge [19] the two best methods [16] were based on Fully Convolutional Neural Networks (FCNN). A new type of neural networks recently emerged for image segmentation task, the transformers networks [20] [21]. Combined with CNN this kind of networks outperform the standard U-Net in several problems [22]. Previous methods propose, as a first step, to segment the myocardium, but the presence of infarct tissue may lead to a poor detection. To overcome these problems, we propose first to

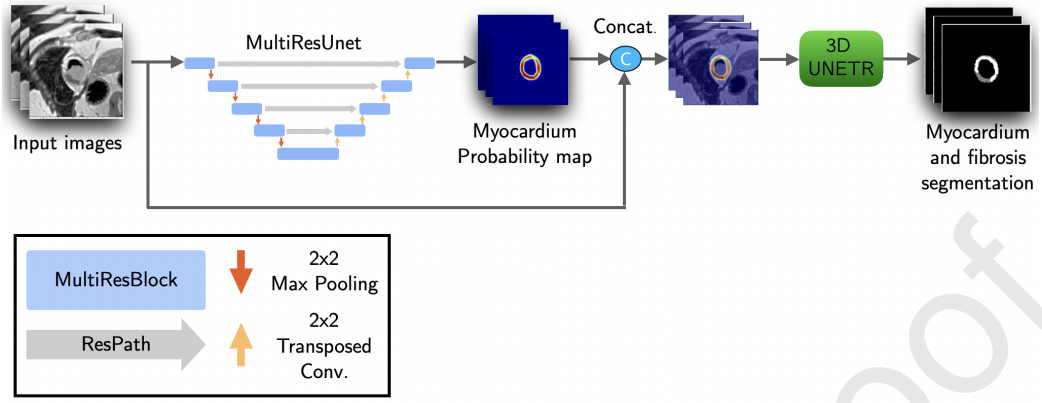



Figure 1: Overall pipeline of the proposed method. It begins with a 2D-UNet as the first stage, which generates a probability map of the normal myocardium. This map is then concatenated with the original image. The augmented image is fed into the UNETR, a U-net architecture enhanced with transformers.

extract a probability map of the myocardium healthy tissue, as this class is balanced and its location and shape vary little. We used a 2D U-Net with residual (MultiResUnet in Fig.1) to compute this probability map of the myocardium, in order to get the maximum information in 2D. This probability map is concatenated as a second channel to the original image to increase information as input of a second network. As this second network has to swipe through the whole image, it must take into account global information and not only the local one. We thus used transformers [20] which has a large receptive field with a positional encoding (Fig.2) of different features. In the considered case, the infarction is always contained in the myocardium. The objective is to leverage the structure of the probability map to achieve two goals. Firstly, it serves as a guidance mechanism for the second network (UNETR) in accurately segmenting the healthy myocardium. Secondly, by detecting areas with holes or gaps, it offers valuable cues for identifying infarction regions.

Moreover, to improve the accuracy on the edges, we propose to add a contour loss [23] based on the Root Means Square Error (RMSE) between the contours map of the ground truth and the prediction.

### *2.1. First stage : Extraction of a probability map with the MultiResUnet*

The MultiResUnet [24] is an enhanced version of the U-Net. The encoder features map are passed through a sequence of convolutional layers with

residual connections (ResPath) before concatenation to the decoder branch. As a result, the number of filters in the convolutional layers is doubled. We used a sigmoid function for the output layer, in order to keep the infarction information represented by a hole in the probability map (e.g decreasing values caused by the presence of the infarction), instead of a hard segmentation which may exclude infarcted areas.

These modifications enabled to retrieve more spatial information in different context size in the image. We have chosen for the loss function a combination of binary cross-entropy (CE) and Dice score as it proved its efficiency [16]. It can be summarized as follows :

$$\mathcal{L}_{2D} = \lambda_1 \mathcal{L}_{CE} + \lambda_2 \mathcal{L}_{Dice} \tag{1}$$

$$\mathcal{L}_{Dice} = 1 - \frac{2 * \sum_i^I y_{true}(i) * y_{pred}(i) + \epsilon}{\sum_i^I y_{true}^2(i) + y_{pred}^2(i) + \epsilon} \tag{2}$$

$$\mathcal{L}_{CE} = -\frac{1}{I} \sum_i^I y_{true}(i) \log(y_{pred}(i)) + (1 - y_{true}(i)) \log(1 - y_{pred}(i)) \tag{3}$$

Where $I$ is the number of pixels and $y_{pred}$, $y_{true}$ the prediction and ground truth respectively; the weights $\lambda_1 = \lambda_2 = 1$.

### *2.2. Second stage: 3D U-Net with transfomers*

The U-Net with transformers (UNETR) [22] is a U-shape based network using transformers for the encoder (Fig. 2). The input image is 3D volume $v \in \mathbb{R}^{h \times w \times d \times c}$. with a resolution $(H, W, D)$ and $C$ the input channels, with the probability map as second channel, is divided into 3D non over-lapping patches with a resolution $(P_H, P_W, P_D)$. Then these patches are flattened to obtained a 1D sequence with a length $N = (H \times W \times D \times C)/(P_H \times P_W \times P_D)$. To preserve the spatial information of the extracted patches, a 1D learnable vector of positional embedding is added to the sequence, in order to encode the relative position of each patches. Then, a stack of twelve transformer blocks composed of multi-head self-attention [20] (MSA) and multilayer perceptron [25] (MLP) is used as encoder. Analogous to the U-Net [9], the multiple features from the encoder are concatenated with the decoder branch, which is composed of deconvolutional layers.

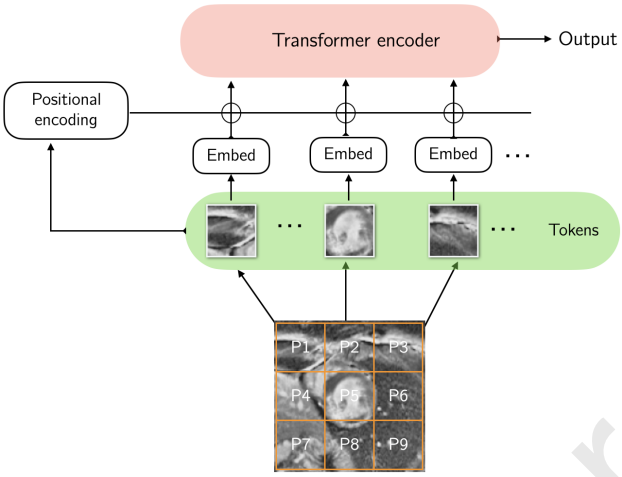



Figure 2: In the 3D U-Net with transformers, the embedding process begins by subdividing the image into patches, with each patch having a corresponding relative positional encoding. Inspired from [21]

### *2.3. Loss function*

The most commonly used loss function is the combination of cross-entropy and Dice (combo-loss). However a precise location and delineation of the infarction is crucial to medical applications, and the combo-loss may produce irregular contours. We thus added a contour quality to our loss function. The loss function is a combination of Cross-Entropy (CE), Dice, and a contour loss to increase the accuracy at the boundaries:

$$\mathcal{L}_{3D} = \lambda_1 \mathcal{L}_{CE} + \lambda_2 \mathcal{L}_{Dice} + \lambda_3 \mathcal{L}_{contour} \tag{4}$$

with the weights $\lambda_1 = \lambda_2 = \lambda_3 = 1$. The contour loss was defined by the Root Mean Square Error (RMSE) between the extracted edges of the prediction and the ground truth. Let's consider $\Omega_{pred}$ and $\Omega_{true}$ as the ground truth contour and the predicted contour. The contour loss is defined as :

$$\mathcal{L}_{contour} = (\frac{1}{J}\sum_{j}^{J}\frac{1}{I}\sum_{i}^{I}(\Omega_{pred}(i,j) - \Omega_{true}(i,j))^2)^{\frac{1}{2}} \tag{5}$$

Where $J$ is the number of classes and $I$ the number of voxels belonging to the contour.

The contours are extracted using morphological filters. Let's consider an image $I$, the binary morphological erosion $\ominus$ and $B$ the structural element, the contours $\Omega$ is defined as :

$$\Omega = I - (I \ominus B) \tag{6}$$

With the contours defined as such, it is possible to vary the thickness of the contours $\Omega$ by changing the size of the structural element $B$, which makes the contour loss eq. (5) more or less punitive.

## 3. Experiments

### *3.1. Dataset and evaluation metrics*

To evaluate this method, we used the EMIDEC [17], [19] training set. Composed of 100 LGE-MRI exams, including 67 pathological cases and 33 healthy cases. Manual delineations are provided, with label 0 for the background, 1 for the left ventricular cavity, 2 for the healthy myocardium, 3 for the myocardial infarction and 4 for the no-reflow which is characterized by an hypo-signal caused by reduced blood flow [26].

In this paper, our primary focus is on the segmentation task of myocardial infarction in hyper-signal. It is important to note that we are not specifically addressing the no-reflow areas in hypo-signal. Clinically, the no-reflow phenomenon is considered as a part of the infarction; however, previous studies have shown that the no-reflow phenomenon tends to diminish after a few months [27], leading to the development of fibrosis in hyper-signal. In our application, we classify the no-reflow as part of the myocardium, although this may not align with clinical practice. This does not affect the performance of our application, which focuses on the assessment of fibrosis at the time the image is acquired.

The evaluation metrics used are based on the code provided in the challenge [28]. It is composed of Dice score, volume difference, Hausdorff distance and the volume difference ratio according to the volume of myocardium (only for the infarcted tissue). The Dice score is computed in 3D and the Hausdorff distance is defined as the maximum surface distance between the objects.

### *3.2. Implementation details*

All the models were implemented in Pytorch [29] and trained on a GPU NVIDIA [30] A100 Tensor Core 40GB.

The MultiRes-Unet was trained with a batch size of 16 and used the Adam optimizer [31] with initial learning rate of 0.001. The UNETR was trained with a batch size of 8 and used the AdamW optimizer [32] with initial learning rate of 0.0003. It uses a patch size of $(8 \times 8 \times 1)$. We employed five-fold cross validation with each testing fold composed of 67% of pathological cases and 33% of healthy cases.

A simple 2D preprocessing composed of cropping and padding by $128 \times 128 \times 16$, a median filter and histogram equalization were used for all images. In addition, we used data augmentation made with the framework TorchIO [33]. It is composed of random geometric transformations such as rotation, scaling, translation, flip, elastic deformation and random intensity transformations such as blur and gamma. A morphology filter (connected components filtering) was applied on the prediction to get the final segmentation.

### *3.3. Ablation study*

To measure experimentally the contribution of the contour loss eq. (5) and the probability map, we have conducted an ablation study as follows:

1. No contour loss (only Dice score and cross-entropy)
2. No probability map and no contour loss

The implementation details and evaluation metrics are unchanged.

## 4. Results

### *4.1. Whole method*

The Table 1 gathers the results of the different metrics obtained by our method. In order to test the robustness of the method, we used a k-fold validation. The best overall results are for the fold 2. In term of Dice score for the infarcted tissue, the fold 3 obtained the best results with 96.76%. Moreover the standard deviation is relatively small compared to the average, with 3.2% for the Dice score of the infarction class and and 0.85% for the myocardium. This demonstrates the robustness of the method. The proposed method achieves an average Dice score of 92.36% for the myocardium class and 92.94% for the infarction class. For the infarction, all metrics results are better than the state-of-the-art [18] (Table 2).

The Fig. 3 illustrates two cases with a Dice score close to the average Dice score ($\approx 92\%$) for the myocardium and infarcted tissue. For both cases the

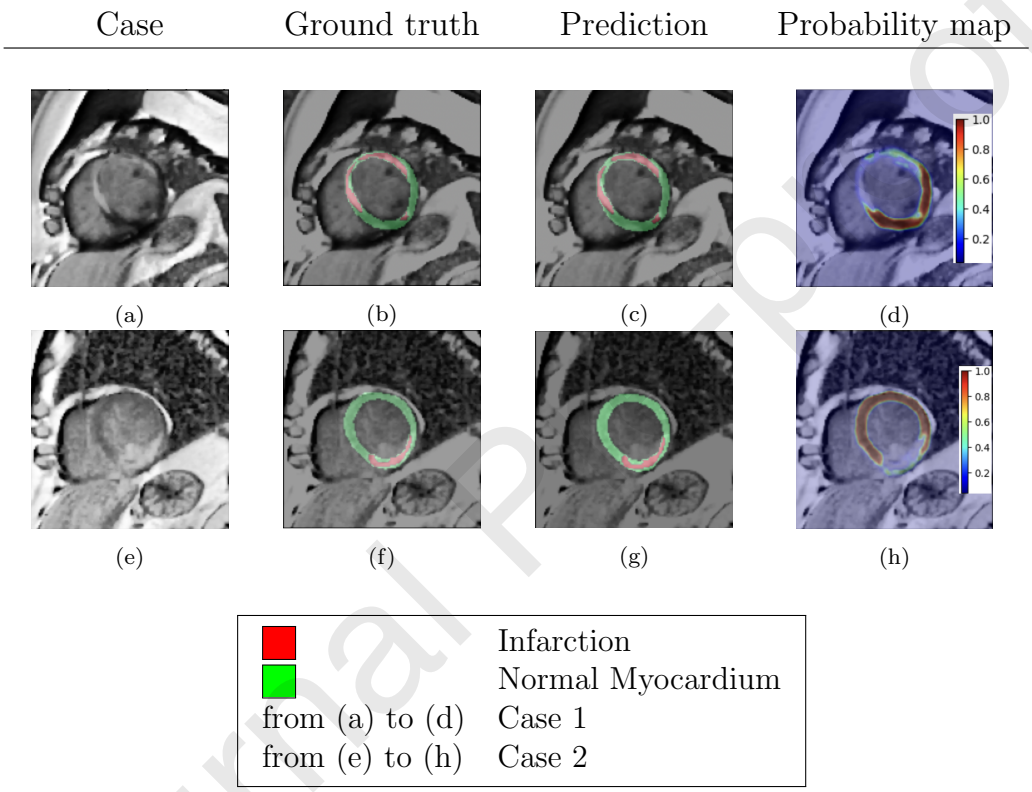



Figure 3: Examples of two cases corresponding to Dice scores close to the average Dice score ($\approx 92\%$) with the associated myocardium probability map, compared to the ground truth

Table 1: Quantitative 5-fold cross-validation results for the proposed method

| Class | Metrics | fold0 | fold1 | fold2 | fold3 | fold4 | Average | SD |
|---|---|---|---|---|---|---|---|---|
| Myocardium | Dice(%) | 90.74 | 92.26 | **93.07** | 92.77 | 92.95 | **92.36** | **0.85** |
| | VolDif($mm^3$) | 1176.77 | 953.74 | 659.05 | **525.07** | 764.52 | **815.83** | **228.5** |
| | HSD($mm$) | 4.9 | 4.06 | **2.91** | 3.57 | 3.07 | **3.7** | **0.72** |
| Infarction | Dice(%) | 91.45 | 92.63 | 95.93 | **96.76** | 87.91 | **92.94** | **3.2** |
| | VolDif($mm^3$) | 296.48 | 96.74 | **75.32** | 128.45 | 158.94 | **151.19** | **77.98** |
| | Ratio(%) | 2.09 | 0.48 | **0.37** | 0.66 | 1.22 | **0.96** | **0.63** |

Table 2: Comparison study. (Best values in bold font)

| Classe | Metrics | Methods | | |
|---|---|---|---|---|
| | | Zhang [16] | Brahim [18] | Ours |
| Myocardium | Dice(%) | 94.40 | **95.10** | 92.36 |
| | VolDif($mm^3$) | 6474.38 | **266.41** | 815.85 |
| | HSD($mm$) | 17.21 | 4.40 | **3.7** |
| Infarction | Dice(%) | 72.08 | 76.14 | **92.94** |
| | VolDif($mm^3$) | 4179.5 | 264.91 | **151.19** |
| | Ratio(%) | 3.41 | 5.32 | **0.96** |

resulting segmentation (Fig. 3g) is very close to the ground truth (Fig. 3f). The edges are smooth (Fig. 4m, Fig. 4l), the infarction and the myocardium are connected which corresponds to the biological reality. As expected the hole in the probability map covers the infarcted region of the myocardium. Moreover we can see on the edges of the probability map that the values decrease sharply, which means that the first network (MultiResUnet) [24] has moderate confidence on the contours delineation. This justifies the use of a contour loss for the second network (UNETR) [22].

To study more precisely the impact of the contour loss and probability map, we present in the next two sections ablation studies.

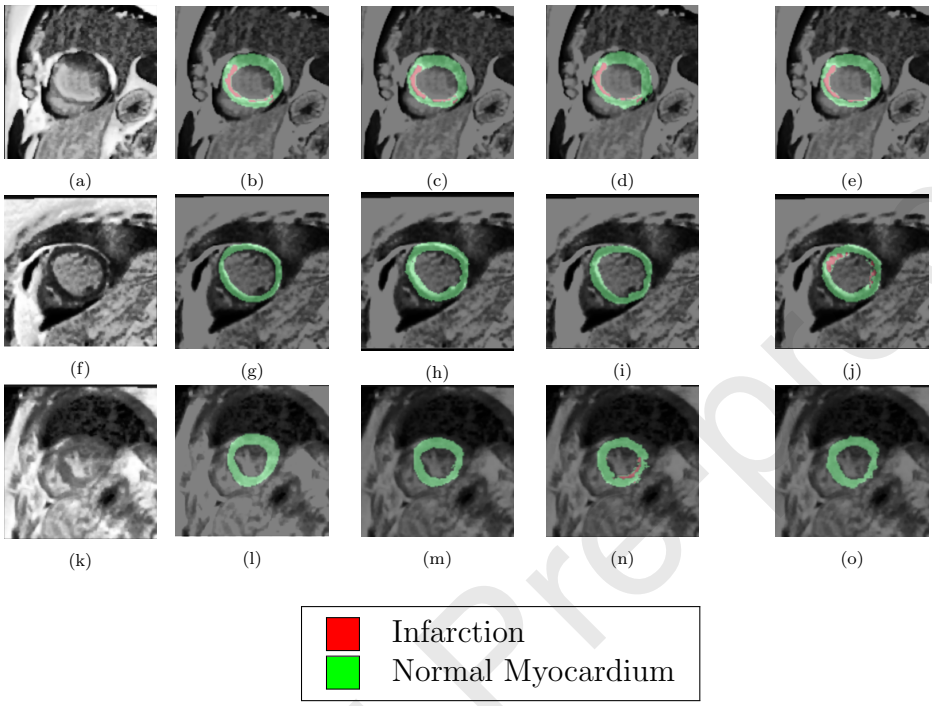



Figure 4: Qualitative results for the ablation study

Table 3: Quantitative results for the ablation study

| Classe | Metrics | No cont. loss | | No prob.map and no cont. loss | |
|---|---|---|---|---|---|
| | | Average | SD | Average | SD |
| Myocardium | Dice(%) | 91.05 | 1.7 | 91.33 | 0.96 |
| | VolDif($mm^3$) | 1021.17 | 272.23 | 795.34 | 201.51 |
| | HSD($mm$) | 4.98 | 1.31 | 5.22 | 0.75 |
| Infarction | Dice(%) | 80.07 | 6.35 | 74.41 | 4.99 |
| | VolDif($mm^3$) | 213.49 | 158.23 | 196.99 | 125.8 |
| | Ratio(%) | 1.38 | 1.15 | 1.26 | 0.9 |

*4.2. The role of the contour loss*

The Table 3 contains the quantitative results of the ablation study. For each metric previously seen, we have the average and standard deviation that are given for both studies (without the contour loss and without the probability and the contour loss). For the infarction, the Dice score is lower without the use of contour loss than with the proposed method (80.07% versus 92.94%). The volume difference is also higher without the contour loss (296.48 versus 151.19). For the myocardium the Hausdorff distance is also higher (4.98$mm$ versus 3.7$mm$). Without the contour loss, we can observe that the results contain false positive and also that the myocardium boundaries are less detected and less consistent.

The figure 4 illustrates qualitative results for the ablation study and shows three examples. The contour loss that we added improves the Dice score and avoids false positive cases (Fig.4.m versus Fig.4.n). Thus the probability map makes it more robust for the infarction class and it also gives more consistent myocardium contours on difficult cases (Fig.4.m versus Fig.4.o).

*4.3. The role of the probability map*

The ablation study shows that by removing the contour loss and the probability map the standard deviation increases (Table 1 versus Table 3, column SD).

For the infarction, the Dice score is lower than the proposed method (Table 3: 74.41% versus 92.94%) and even lower than the method without the contour loss (Table 3: 74.41% versus 80.07%). For the myocardium the Hausdorff distance increased without the contour loss (4.98 mm) and without the probability map (5.22$mm$), compared to the proposed method (3.7$mm$).

Without the probability map and the contour loss the network (UNETR) is less sensitive to detect the infarction (Fig.4.c versus Fig.4.e). It results in less regular myocardium contours (Fig.4.m versus Fig.4.o) and moreover in false positives on difficult cases (Fig.4.h versus Fig.4.j). Moreover, the figure 5 shows the qualitative results on one difficult case close the apex, with a Dice score of 95.35% for the myocardium and 93.5% for the infarction. Our method has been able to detect the infarction whereas without the probability map and the contour loss, it failed to detect it.

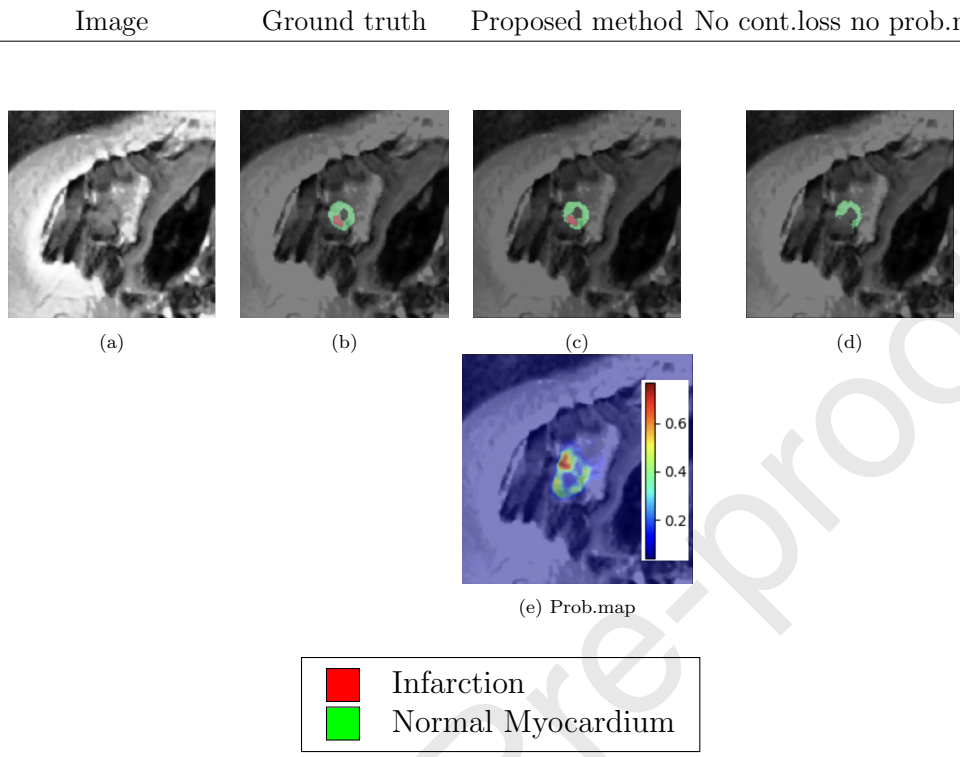



Figure 5: Qualitative results for a difficult case close to the apex

## 5. Discussion and Conclusion

In this paper, we introduce a novel approach for automated myocardial infarction segmentation from LGE-MRI using a hybrid cascaded neural network with transformers. The first network, a 2D CNN, generates a probability map of healthy myocardium tissues. This map is then concatenated with the original images, effectively increasing the available information for input into the second network.

This approach successfully addresses the challenge posed by a substantial slice thickness. However, it's important to note that at the edges of the probability map, there's a rapid decrease in values, indicating a moderate confidence in contour delineation by the first network. To address this limitation, we introduce a contour loss in the second network.

The second network is a 3D hybrid CNN with transformers, enabling the network to spatially link feature maps. Spatial relationships can be incorporated into CNNs through texture analysis techniques like the Gray-Level

Co-Occurrence Matrix (GLCM), as demonstrated in the approach by Attalah et al. [34]. However, it's important to note that these methods were primarily employed for detection and diagnosis purposes and not specifically for segmentation. The UNETR utilizes transformers as encoders for 3D images. However, a novel approach introduced by Abdelhamed et al. [35] known as Nested Hierarchical Transformer (NesT) employs basic local transformer modules on non-overlapping image blocks and hierarchically aggregates them but it's important to note that this method is currently applied exclusively to 2D images.

Additionally, the inclusion of a contour loss enhances performance at the edges. The utilization of transformers in handling highly anisotropic images proves to be more efficient compared to convolution-based networks. This observation is attributed to transformers' ability to capture long-range dependencies and model spatial relationships, which is particularly advantageous for handling the challenges posed by highly anisotropic data.

The intended operation of the contour loss, resulting in smoother contours and a reduction in false positive detections. From a clinical perspective, precise contour definitions are essential for accurate analysis and interpretation.

Our proposed method outperforms state-of-the-art techniques across all evaluated metrics in the challenge [17], demonstrating superior performance in infarction segmentation. However, it's important to note that it does not incorporate the no-reflow class as part of the infarction, which is particularly challenging to segment. Furthermore, the method's robustness has not been tested on alternative datasets.

To further enhance the method, exploring the integration of Generative Adversarial Networks (GANs), as exemplified by the approach employed by Hendriks et al. [36], holds significant promise. GANs possess the capability to generate realistic synthetic samples, which can lead to improved data augmentation, particularly benefiting scenarios involving rare cases. This strategic augmentation approach has the potential to bolster the model's robustness and enhance its generalization capabilities by introducing greater diversity into the training data.

Additionally, to enhance the model's robustness and flexibility, we can consider incorporating different CMR-sequences, such as bSSFP CMR, which provides clear boundaries of the myocardium and left ventricle, into the network training, following the approach demonstrated by Qiu et al. [37].

Furthermore, in addressing class imbalance concerns, we may consider implementing a dedicated network for region of interest (ROI) extraction,

drawing inspiration from the work of Langarizadeh et al. [38].

## 6. Acknowledgements

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] R. J. Kim, E. Wu, A. Rafael, E.-L. Chen, M. A. Parker, O. Simonetti, F. J. Klocke, R. O. Bonow, R. M. Judd, The use of contrast-enhanced magnetic resonance imaging to identify reversible myocardial dysfunction, New England Journal of Medicine 343 (2000) 1445–1453.

[2] Q. Tao, S. R. Piers, H. J. Lamb, R. J. van der Geest, Automated left ventricle segmentation in late gadolinium-enhanced MRI for objective myocardial scar assessment, Journal of Magnetic Resonance Imaging 42 (2015) 390–399.

[3] R. Karim, P. Bhagirath, P. Claus, R. J. Housden, Z. Chen, Z. Karimaghaloo, H.-M. Sohn, L. L. Rodríguez, S. Vera, X. Albà, et al., Evaluation of state-of-the-art segmentation algorithms for left ventricle infarct from late gadolinium enhancement mr images, Medical image analysis 30 (2016) 95–107.

[4] R. J. Kim, D. S. Fieno, T. B. Parrish, K. Harris, E.-L. Chen, O. Simonetti, J. Bundy, J. P. Finn, F. J. Klocke, R. M. Judd, Relationship of MRI delayed contrast enhancement to irreversible injury, infarct age, and contractile function, Circulation 100 (1999) 1992–2002.

[5] L. C. Amado, B. L. Gerber, S. N. Gupta, D. W. Rettmann, G. Szarf, R. Schock, K. Nasir, D. L. Kraitchman, J. A. Lima, 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 (2004) 2383–2389.

[6] A. Hennemuth, A. Seeger, O. Friman, S. Miller, B. Klumpp, S. Oeltze-Jafra, H.-O. Peitgen, A comprehensive approach to the analysis of contrast enhanced cardiac mr images, IEEE transactions on medical imaging 27 (2008) 1592–610. doi:10.1109/TMI.2008.2006512.

[7] Z. Karimaghaloo, M. Shah, S. J. Francis, D. L. Arnold, D. L. Collins, T. Arbel, Automatic detection of gadolinium-enhancing multiple sclerosis lesions in brain MRI using conditional random fields, IEEE transactions on medical imaging 31 (2012) 1181–1194.

[8] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, D. Rueckert, Deep learning for cardiac image segmentation: a review, Frontiers in Cardiovascular Medicine (2020) 25.

[9] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.

[10] N. Siddique, S. Paheding, C. P. Elkin, V. Devabhaktuni, U-net and its variants for medical image segmentation: A review of theory and applications, Ieee Access 9 (2021) 82031–82057.

[11] R. R. Upendra, R. Simon, C. A. Linte, Joint deep learning framework for image registration and segmentation of late gadolinium enhanced MRI and cine cardiac MRI, in: Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling, volume 11598, SPIE, 2021, pp. 96–103.

[12] C. Decourt, L. Duong, Semi-supervised generative adversarial networks for the segmentation of the left ventricle in pediatric MRI, Computers in Biology and Medicine 123 (2020) 103884.

[13] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, M. Wang, Swin-unet: Unet-like pure transformer for medical image segmentation, in: European conference on computer vision, Springer, 2022, pp. 205–218.

[14] F. Zabihollahy, M. Rajchl, J. A. White, E. Ukwatta, Fully automated segmentation of left ventricular scar from 3d late gadolinium enhancement magnetic resonance imaging using a cascaded multi-planar u-net (cmpu-net), Medical physics 47 (2020) 1645–1655.

[15] F. Zabihollahy, S. Rajan, E. Ukwatta, Machine learning-based segmentation of left ventricular myocardial fibrosis from magnetic resonance imaging, Current Cardiology Reports 22 (2020) 1–8.

[16] Y. Zhang, Cascaded convolutional neural network for automatic myocardial infarction segmentation from delayed-enhancement cardiac MRI, in: International Workshop on Statistical Atlases and Computational Models of the Heart, Springer, 2020, pp. 328–333.

[17] A. Lalande, Z. Chen, T. Decourselle, A. Qayyum, T. Pommier, L. Lorgis, E. de la Rosa, A. Cochet, Y. Cottin, D. Ginhac, et al., Emidec: a database usable for the automatic evaluation of myocardial infarction from delayed-enhancement cardiac MRI, Data 5 (2020) 89.

[18] K. Brahim, A. Qayyum, A. Lalande, A. Boucher, A. Sakly, F. Meriaudeau, A 3d network based shape prior for automatic myocardial disease segmentation in delayed-enhancement MRI, IRBM 42 (2021) 424–434.

[19] A. Lalande, Z. Chen, T. Pommier, T. Decourselle, A. Qayyum, M. Salomon, D. Ginhac, Y. Skandarani, A. Boucher, K. Brahim, et al., Deep learning methods for automatic evaluation of delayed enhancement-MRI. the results of the emidec challenge, Medical Image Analysis 79 (2022) 102428.

[20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).

[21] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, in: International Conference on Learning Representations, ICLR 2021, OpenReview.net, 2021.

[22] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, D. Xu, Unetr: Transformers for 3d medical image segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 574–584.

[23] X. Chen, B. M. Williams, S. R. Vallabhaneni, G. Czanner, R. Williams, Y. Zheng, Learning active contour models for medical image segmentation, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 11632–11640.

[24] N. Ibtehaz, M. S. Rahman, Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation, Neural Networks 121 (2020) 74–87.

[25] H. Ramchoun, Y. Ghanou, M. Ettaouil, M. A. Janati Idrissi, Multilayer perceptron: Architecture optimization and training (2016).

[26] V. Pineda, X. Merino, S. Gispert, P. Mahía, B. Garcia, R. Domínguez-Oronoz, No-reflow phenomenon in cardiac MRI: diagnosis and clinical implications, American Journal of Roentgenology 191 (2008) 73–79.

[27] R. Nijveld, A. Beek, A. Hirsch, M. Hofman, V. Uman, P. Algra, A. Van Rossum, 'no-reflow'after acute myocardial infarction: direct visualisation of microvascular obstruction by gadolinium-enhanced CMR, Netherlands Heart Journal 16 (2008) 179–181.

[28] A. Qayyum, Emidec evaluation metrics, 2020. URL: `https://github.com/EMIDEC-Challenge/Evaluation-metrics`, -.

[29] Pytorch, Pytorch documentation, 2022. URL: `https://pytorch.org/`, -.

[30] NVIDIA, Nvidia a100, 2022. URL: `https://www.nvidia.com/en-us/data-center/a100/`, -.

[31] J. Ba, K. Diederik, Adam: A method for stochastic optimization, in: Anon. International Conferenceon Learning Representations. SanDego: ICLR 2015, 2015.

[32] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, arXiv preprint arXiv:1711.05101 (2017).

[33] F. Pérez-García, R. Sparks, S. Ourselin, 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 (2021) 106236.

[34] O. Attallah, D. A. Ragab, Auto-myin: Automatic diagnosis of myocardial infarction via multiple glcms, cnns, and svms, Biomedical Signal Processing and Control 80 (2023) 104273. URL: `https://www.sciencedirect.com/science/article/pii/S1746809422007273`. doi:`https://doi.org/10.1016/j.bspc.2022.104273`.

[35] M. K. Abdelhamed, F. Meriaudeau, Nest unet: pure transformer segmentation network with an application for automatic cardiac myocardial infarction evaluation, in: Medical Imaging 2023: Computer-Aided Diagnosis, volume 12465, SPIE, 2023, pp. 608–619.

[36] F. Lau, T. Hendriks, J. Lieman-Sifry, S. Sall, D. Golden, Scargan: chained generative adversarial networks to simulate pathological tissue on cardiovascular mr scans, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4, Springer, 2018, pp. 343–350.

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

[38] M. Langarizadeh, M. Jahanshahi, et al., Myocardial fibrosis delineation in late gadolinium enhancement images of hypertrophic cardiomyopathy patients using deep learning methods., Journal of Health Administration 25 (2022).

# Conflicts of Interest Statement

Manuscript title: **Segmentation of cardiac infarction in delayed-enhancement MRI using probability map and transformers-based neural network**

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

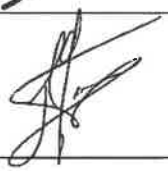
Author names:

**Erwan Lecesne**
**Antoine Simon**
**Mireille Garreau**
**Céline Fouard**
**Gilles Barone-Rochette**

The authors whose names are listed immediately below report the following details of affiliation or involvement in an organization or entity with a financial or non-financial interest in the subject matter or materials discussed in this manuscript. Please specify the nature of the conflict on a separate sheet of paper if the space below is inadequate.

Author names:

**This statement is signed by all the authors to indicate agreement that the above information is true and correct** (*a photocopy of this form may be used if there are more than 10 authors*):

| Author's name (typed) | Author's signature | Date |
|---|---|---|
| **Erwan Lecesne** |  | **16/01/2023** |
| **Antoine Simon** |  | 17/01/2023 |
| **Mireille Garreau** |  | 17/01/2023 |
| **Gilles Barone-Rochette** |  | 17/01/2023 |
| **Céline Fouard** |  | **16/01/2023** |
| | | |
| | | |
| | | |
| | | |
| | | |