

Magnetic Resonance Materials in Physics, Biology and Medicine

Development and testing of a deep learning-based strategy for scar segmentation on CMR-LGE images

--Manuscript Draft--

Manuscript Number:	
Full Title:	Development and testing of a deep learning-based strategy for scar segmentation on CMR-LGE images
Article Type:	Original Paper
Funding Information:	
Abstract:	<p>Object: The aim of this paper is to investigate the use of fully-convolutional neural networks (FCNNs) to segment scar tissue in the left ventricle from cardiac magnetic resonance with late gadolinium enhancement (CMR-LGE) images.</p> <p>Methods: A successful FCNN in the literature (the ENet) was modified and trained to provide scar-tissue segmentation. Two segmentation protocols (Protocol 1 and Protocol 2) were investigated, the latter limiting the scar-segmentation search area to the left ventricular myocardial tissue region. CMR-LGE from 30 patients with ischemic-heart disease were retrospectively analyzed, for a total of 250 images, presenting high variability in terms of scar dimension and location. Segmentation results were assessed against manual scar-tissue tracing using one-patient-out cross validation.</p> <p>Results: Protocol 2 outperformed Protocol 1 significantly (p-value < 0.05), with median sensitivity and Dice similarity coefficient equal to 88.07% (inter-quartile range (IQR) = 18.84% and 71.25% (IQR = 31.82%), respectively.</p> <p>Discussion: Both segmentation protocols were able to detect scar tissues in the CMR-LGE images but higher performance was achieved when limiting the search area to the myocardial region. The findings of this paper represent an encouraging starting point for the use of FCNNs for the automatic segmentation of nonviable scar tissue from CMR-LGE images.</p>
Corresponding Author:	Enrico Gianluca Caiani, Ph.D. Politecnico di Milano MILANO, Lombardy ITALY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Politecnico di Milano
Corresponding Author's Secondary Institution:	
First Author:	Sara Moccia, PhD
First Author Secondary Information:	
Order of Authors:	<p>Sara Moccia, PhD</p> <p>Riccardo Banali, MS</p> <p>Chiara Martini, MD</p> <p>Giuseppe Muscogiuri, MS</p> <p>Gianluca Pontone, PhD</p> <p>Mauro Pepi, MS</p> <p>Enrico Gianluca Caiani, Ph.D.</p>
Order of Authors Secondary Information:	
Author Comments:	



Dipartimento di Elettronica, Informazione e Bioingegneria

Politecnico di Milano

20133 Milano (Italia)
Piazza Leonardo da Vinci, 32

August 6th, 2018

Prof. David G. Norris
Editor-in-Chief
Magnetic Resonance Materials in Physics, Biology and Medicine

Dear Prof. Norris,

here enclosed you can find the manuscript titled "DEVELOPMENT AND TESTING OF A DEEP LEARNING-BASED STRATEGY FOR SCAR SEGMENTATION ON CMR-LGE IMAGES", which we would like to submit for publication in the ***Magnetic Resonance Materials in Physics, Biology and Medicine Journal***.

The undersigned author warrants that the article is not under consideration by another journal, and has not been previously published.

I sign for and accept the responsibility for releasing this material on behalf of any and all co-authors.

Sincerely,

Enrico G Caiani, PhD

Prof. Enrico G Caiani, PhD
Electronics, Information and Bioengineering Department
Politecnico di Milano
Milan (IT)
enrico.caiani@polimi.it

[Click here to view linked References](#)

Noname manuscript No. (will be inserted by the editor)
--

Development and testing of a deep learning-based strategy for scar segmentation on CMR-LGE images

Sara Moccia · Riccardo Banali · Chiara Martini · Giuseppe Muscogiuri · Gianluca Pontone · Mauro Pepi · Enrico Gianluca Caiani

Received: date / Accepted: date

Abstract Object: The aim of this paper is to investigate the use of fully-convolutional neural networks (FCNNs) to segment scar tissue in the left ventricle from cardiac magnetic resonance with late gadolinium enhancement (CMR-LGE) images. **Methods:** A successful FCNN in the literature (the ENet) was modified and trained to provide scar-tissue segmentation. Two segmentation protocols (**Protocol 1** and **Protocol 2**) were investigated, the latter limiting the scar-segmentation search area to the left ventricular myocardial tissue region. CMR-LGE from 30 patients with ischemic-heart disease were retrospectively analyzed, for a total of 250 images, presenting high variability in terms of scar dimension and location. Segmentation results were assessed against manual scar-tissue tracing using one-patient-out cross validation. **Results: Protocol 2** outperformed **Protocol 1** significantly (p-value < 0.05), with median sensitivity and Dice similarity coefficient equal to 88.07% (inter-quartile range (IQR) = 18.84%) and 71.25% (IQR = 31.82%), respectively. **Discussion:** Both segmentation protocols were able to detect scar tissues in the CMR-LGE images but higher performance was achieved when limiting the search area to the myocardial region. The findings of this paper represent an encouraging starting point for the use of FCNNs for the automatic segmentation of nonviable scar tissue from CMR-LGE images.

S. Moccia
Department of Information Engineering, Università Politecnica delle Marche, Ancona (Italy)
Department of Advanced Robotics, Istituto Italiano di Tecnologia, Genoa (Italy)

R. Banali and E.G. Caiani
Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan (Italy)

C. Martini
Diagnostic Department, Azienda Ospedaliera-Universitaria di Parma, Parma (Italy)

G. Muscogiuri, G. Pontone and M. Pepi
Clinical Cardiology Unit and Department of Cardiovascular Imaging, Centro Cardiologico Monzino IRCCS, Milan (Italy)

E-mail: enrico.caiani@polimi.it

Keywords Scar segmentation · Deep learning · Fully-convolutional neural networks · CMR-LGE images

1 Introduction

The presence of nonviable scar tissue in left ventricle (LV) has a crucial prognostic and therapeutic role. Indeed, it enables assessing LV remodeling, as well as patient’s cardiac dysfunction or mortality [1, 2, 3].

Scar-tissue presence is commonly identified using cardiac magnetic resonance with late gadolinium enhancement (CMR-LGE) [4]. CMR-LGE is performed within ten to twenty minutes after the intravenous administration of gadolinium, thus when gadolinium has already been washed out from healthy tissues. This results in hyperenhanced (HE) intensity areas where nonviable scar tissue is present [5].

In clinical practice, scar analysis from CMR-LGE images is performed qualitatively using the American heart association (AHA) 17-segment model [6], which provides conventional risk stratification [7]. In particular, the average transmural extent of HE areas is estimated within each segment (0%, 1-25%, 26-50%, 51-75%, 76-100%) to perform diagnosis. Guidelines also suggest to compare CMR-LGE images with cine and perfusion images (if the latter are obtained) to correctly categorize ischemia and viability. However, a quantitative approach to scar analysis would provide supplementary information to be exploited for diagnosis and follow-up evaluation [8].

Several methods have been proposed for quantitative scar analysis that rely on automatic or semi-automatic scar segmentation in the manually-traced LV region from CMR-LGE images. Methods include two widely used threshold-based semi-automatic algorithms [7]. These two algorithms are semi-automatic as a region of interest has to be manually identified within the myocardial region for calculating the threshold values. The two thresholds are defined as: (i) the intensity value n -standard deviations higher than the mean intensity of a user-defined region in the normal myocardium (nSD), (ii) the half value of maximum intensity of a user-defined HE region (full width at half maximum (FWHM)). In addition, other methods in the literature largely exploit pixel-intensity information for scar-tissue segmentation through thresholding (e.g. [9, 10, 11, 12, 13, 14, 15]). Popular approaches include also clustering techniques, such as Gaussian mixture model (GMM) [16], fuzzy c -means [8], and superpixel segmentation [17, 18], where a superpixel is defined as a group of connected pixels with similar gray-level intensity and texture [19]. Max-flow and graph-cut optimization were explored in [20, 21], while level-set modeling was used in [22].

Although these methodologies achieved encouraging segmentation performance, they suffer from variability in CMR-LGE images (e.g. in terms of noise and intensity level associated to HE areas), and/or require heavy operator intervention. This strongly hampers the translation of the developed methodologies into the actual clinical practice [22]. To tackle image variability and reduce operator intervention, the literature on medical-image segmentation is focusing more and more on deep-learning (DL) approaches based on convolutional neural networks (CNNs) [23]. A CNN is a neural network that consists of (i) a set of convolutional layers and (ii) one or more fully-connected layers. The convolutional layers allow feature extraction while the fully-connected layers perform feature classification. The weights

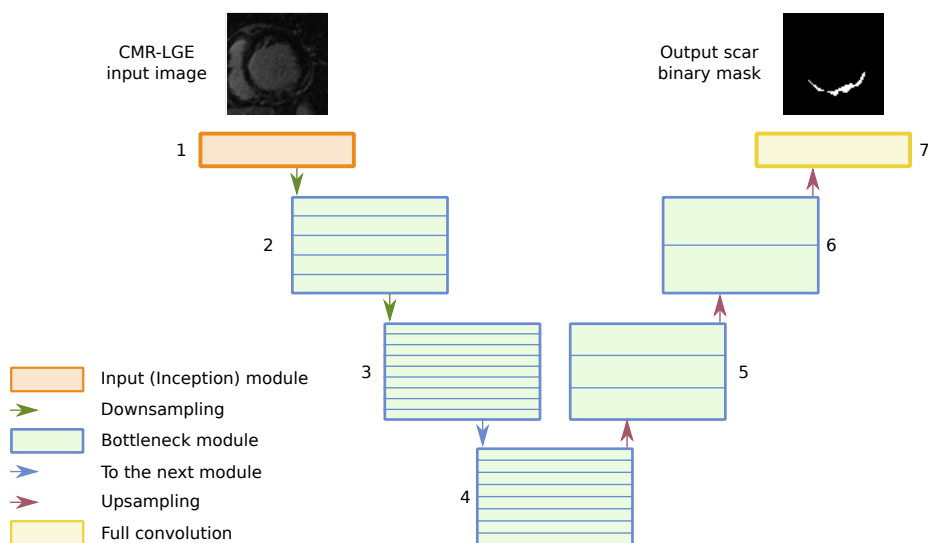


Fig. 1 Fully-convolutional neural network (FCNN) architecture. Numbers refer to the FCNN-module indexes.

of both CNN convolutional kernels and fully-connected connections are learned automatically during the CNN training process.

Despite the potentiality of DL for medical-image segmentation, few DL-based attempts at scar segmentation from CMR-LGE images can be found in the literature. These attempts include [24, 18], that exploit CNNs for superpixels-based feature extraction and sparse auto-encoders for superpixel classification (i.e., each superpixel is classified as scar tissue or healthy myocardial tissue). Additionally, in [25] CNN-based image features were extracted from image square patches, and the features from one patch were classified with fully-connected layers while the output class was assigned to the central pixel of the patch.

It is worth noting that these DL-based methodologies classify superpixels or image patches individually. To encode spatial-connection information while performing segmentation, advancements in DL in other fields (such as natural-image segmentation, and, more recently, medical image segmentation in other anatomical districts) have led to the introduction of fully-convolutional neural networks (FCNNs). In a FCNN, fully-connected layers are replaced by upsampling layers to provide directly fast and accurate image segmentation [26, 27].

Considering the performance achieved by FCNNs in other fields, the aim of this paper was to investigate the feasibility and accuracy of FCNNs for scar segmentation in CMR-LGE images.

2 Methods

From the earlier attempts at using FCNNs for image segmentation [26], mainly relying on successful CNN architectures modified and tuned to include and train upsampling layers, several FCNN models have been proposed (e.g., [28, 29]). In this

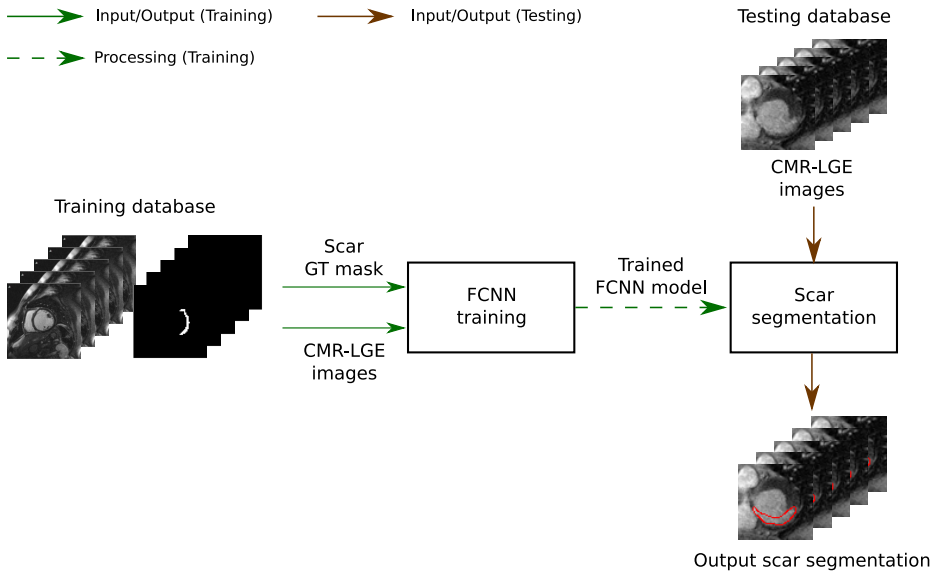


Fig. 2 Workflow of the first segmentation protocol (**Protocol 1**). GT: ground truth; CMR-LGE: cardiac magnetic resonance with late gadolinium enhancement; FCNN: Fully-convolutional neural network.

work, the efficient neural network (ENet), that was presented in [29] for natural-image segmentation, was exploited and modified. Indeed, while providing comparable accuracy to existing FCNN models, ENet was demonstrated to be faster to be trained and able in reducing significantly the number of required floating point operations per second (FLOPs). Moreover, ENet has been already found able to provide good results for ventricle segmentation in CMR images [30].

The ENet architecture consists of a sequence of 7 different stages (Fig. 1). The first stage, called initial stage, consists of an inception module [31], that concatenates the results of convolutional layers of different receptive field size (i.e., 5x5, 3x3 and 1x1) to allow a reach (multi-scale) feature representation. With respect to ENet, our FCNN initial stage had only 13 convolutional layers in parallel to a max-pooling layer, thus resulting in 14 feature maps after concatenation, instead of the original 16. Indeed, ENet was designed for RGB images, while our FCNN was fed with CMR-LGE images with one videointensity channel only. Convolution was performed with 3x3 kernels with stride 2 and max pooling with non-overlapping 2x2 windows.

ENet stages from 2 to 4 act as encoders for feature extraction and consist of a series of bottleneck modules. As in the original paper [31], in the implemented FCNN configuration each bottleneck module had a main branch and a lateral branch, the latter consisting of a series of three convolutional layers. The output of each of the three convolutional layers was activated with the parametric rectified linear unit (PReLU) [32]. Prior to activation, batch normalization was performed as regularization technique. The main branch consisted of a max pooling layer followed by padding to match the dimensions of the convolutional-filter output. The output maps from the two branches were summed up and activated with

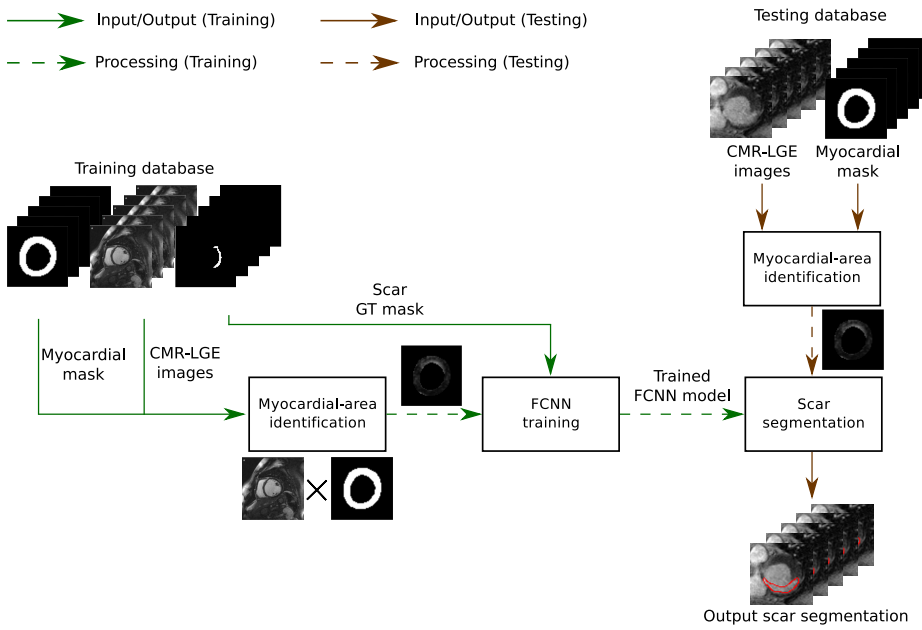


Fig. 3 Workflow of the second segmentation protocol (**Protocol 2**). GT: ground truth; CMR-LGE: cardiac magnetic resonance with late gadolinium enhancement; FCNN: Fully-convolutional neural network. Myocardial masks are obtained with manual delineation of myocardial contours.

PReLU. As shown in Fig. 1, stages 2, 3 and 4 were made of 5, 9 and 8 bottleneck modules, respectively.

Also stages 5 and 6 of ENet consist of a series of bottleneck modules (3 and 2, respectively), but these stages act as decoders performing upsampling. As in [32], here max unpooling and spatial convolution were present in the lateral and main branches of the bottleneck modules, respectively.

The last stage of the proposed FCNN consisted of a bare full convolution. The convolution kernel had height and width equal to the size of the CMR-LGE images, and 2 channels, as the problem addressed in this paper is a binary segmentation problem.

2.1 Segmentation protocols

After modifying the ENet architecture to deal with the scar-segmentation task, two segmentation protocols were investigated. The aim of the first protocol (**Protocol 1**) was to explore the potential of the presented FCNN to directly provide scar segmentation from CMR-LGE images. Thus, during training, the FCNN was fed with CMR-LGE images and the relative scar ground-truth (GT) masks. Scar-GT creation is explained in Sec. 3. The workflow of **Protocol 1** is shown in Fig. 2.

The second protocol (**Protocol 2**) aimed to explore the potential of the FCNN in segmenting scar in a pre-defined LV myocardial region. With **Protocol 2**, the aim was to investigate if delimiting the search area for scar segmentation, as

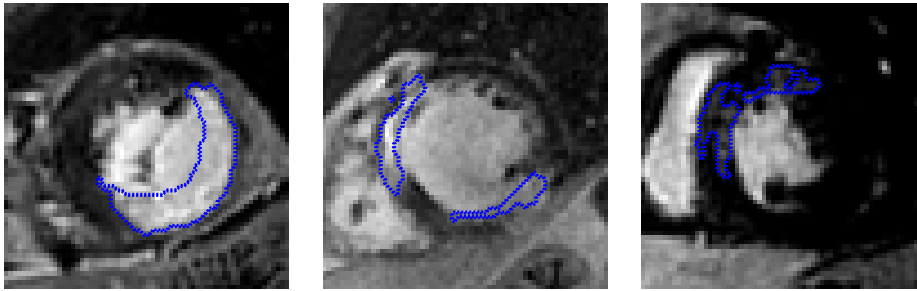


Fig. 4 An example of the masks characterizing the scar presence as drawn by an expert cardiologist for three patients. Scar location and dimension vary from slice to slice and from patient to patient.

currently done in the semi-automated quantification methods described in Sec. 1, could provide more accurate segmentation results. Thus, **Protocol 2** required a priori to manually identify the myocardial boundaries to delimit the search area. To this goal, prior to feeding the FCNN, CMR-LGE images were multiplied by the relevant binary myocardial masks (obtained through manual tracing of LV contours as explained in Sec. 3). The workflow of **Protocol 2** is shown in Fig. 3.

3 Experimental setup

The CMR-LGE images analyzed in this study refer to 30 different patients (26 men and 4 women) acquired at the *Centro Cardiologico Monzino* hospital in Milan (Italy), for a total of 250 short-axis images. These patients were retrospectively selected from the hospital database, with inclusion criteria a diagnosis of ischemic heart disease with a consequent presence of nonviable scar tissue in LV myocardium. Image size was 256x256 pixels and all the images were used for the analysis (i.e., no slice selection was performed).

For training and testing purposes, scar GT was obtained with manual tracing of scar contours by an expert cardiologist using Circle Cardiovascular Imaging v.5.6¹. LV-myocardium contours for **Protocol 2** were obtained in the same way. Examples of CMR-LGE images and relevant ground truth resulting from scar manual tracing are shown in Fig. 4.

Data pre-processing was performed prior to FCNN training and testing. In particular, CMR-LGE images were cropped to reduce the processing area, as commonly suggested in the literature [22]. Image cropping was fully automatic. First, LV diameter and center were retrieved using the circle Hough transform [33] from each CMR-LGE slice. Squared cropping was then performed by centering the crop area with the LV center and setting crop side length equal to double the LV diameter. To standardize the cropped-image size, as LV size varied from patient to patient, all the images were resized to 64x64 pixels, i.e. the minimum crop size found. The 64x64 images were processed by subtracting the intensity mean value from each image and normalizing by the intensity standard deviation.

¹ <https://www.circlecvi.com/>

To test the proposed segmentation approaches, the CMR-LGE image dataset was divided into two sets: the former was used for training and validation and the latter for testing purpose only. Considering the relatively limited number of CMR-LGE images available (even though comparable with similar work in the literature [22]), leave-one-patient-out cross-validation was used for robust performance evaluation. Thus, images from one patient were classified using the FCNN trained with all the images from the remaining 29 patients. This procedure was repeated for all the 30 patients.

Data augmentation was performed on the training set, by applying vertically flipping, horizontally flipping, and a combination of both, 90° rotation, 90° rotation and vertically flipping, 90° rotation and horizontally flipping, and 90° rotation with both vertically and horizontally flipping, for a total of 7 transformations.

3.1 FCNN training

Mini-batch gradient descent was used for FCNN training (for both the segmentation protocols), using a batch size equal to 4. Mini-batch gradient descent was chosen as a compromise between gradient descent and stochastic gradient descent, to provide fast training convergence while limiting the memory usage [34].

Cross entropy was used as loss function. The adaptive moment estimation (ADAM) [35] that adapts the learning rate by regularizing the gradient descent using both gradient amplitude and momentum, was used as training optimizer. To (upper) bound the learning rate during training, an exponentially-decaying learning-rate bounding function was defined, using an initial learning rate equal to 5e-5. This resulted to be useful especially during the last training epochs to further reduce the loss, as commonly recognized in the DL literature [36]. As a regularization technique, a weight decay equal to 2e-5 for the FCNN convolution layers was imposed. The FCNNs for both the two protocols were trained on 100 epochs. The best model among epochs according to the Dice similarity coefficient (*DSC*) [37] was then chosen. All the training parameters were established with a trial-and-error procedure.

FCNN training and testing were implemented using TensorFlow². All tests were performed using NVIDIA[®] GeForce[®] GTX 1050 (4 GB GDDR5 dedicated) on a Intel[®] Core[®] i7-7700HQ (2.8 GHz, 6 MB cache, 4 cores) computer with 16 GB DDR4-2400 SDRAM. FCNN training took ~30 hours for each of the two protocols.

3.2 Evaluation

Inspired by similar work in the literature for scar segmentation (e.g. [24, 18, 25]), the segmentation outcomes, obtained with both **Protocol 1** and **Protocol 2**, were quantitatively evaluated with respect to the GT in terms of pixel classification accuracy (*Acc*), sensitivity (*Se*), and specificity (*Sp*):

$$Acc = \frac{TP + TN}{n} \quad (1)$$

² <https://www.tensorflow.org/>

Table 1 Median (inter-quartile range) performance measures obtained for the first (**Protocol 1**) and second (**Protocol 2**) segmentation protocol. *Acc* = accuracy, *Sp* = specificity, *Se* = sensitivity, *DSC* = Dice similarity coefficient.

	<i>Acc</i>	<i>Sp</i>	<i>Se</i>	<i>DSC</i>
Protocol 1	95.79% (3.55%)	97.31% (3.01%)	68.77% (34.83%)	54.00% (41.03%)
Protocol 2	96.83% (3.26%)	97.89% (2.93%)	88.07% (17.84%)	71.25% (31.82%)

Table 2 Median normalized contingency table for **Protocol 1**.

		Ground-truth segmentation	
		Scar tissue	Background
FCNN-based segmentation	Scar tissue	68.77%	2.69%
	Background	31.23%	97.31%

Table 3 Median normalized contingency table for **Protocol 2**.

		Ground-truth segmentation	
		Scar tissue	Background
FCNN-based segmentation	Scar tissue	88.07%	2.11%
	Background	11.93%	97.89%

$$Se = \frac{TP}{TP + FN} \quad (2)$$

$$Sp = \frac{TN}{TN + FP} \quad (3)$$

where *TP* and *TN* are number of scar and background pixels that were correctly identified, respectively. *FP* and *FN* are the number of background pixels classified as scar tissue and the number of scar pixels classified as background, respectively. The *DSC*, representing an overlap measure, was also computed as:

$$DSC = \frac{2TP}{FP + FN + 2TP} \quad (4)$$

The Wilcoxon signed-rank test (significance level (α) = 0.05) was used to assess whether significant differences existed in *DSC* among the segmentation results obtained with **Protocol 1** and **Protocol 2**.

4 Results

Despite all patients were previously diagnosed with myocardial fibrosis in the LV, scar tissue was present only in 215 slices out of 250 (86% of the slices). The scar area in each slice ranged between 20 and 1259 pixels (pixel resolution: 1.49 x 1.49 mm). The dataset granted high intra- and inter-variability in both scar size and location in the LV, as can be seen from sample scar masks in Fig. 4.

Table 1 shows the performance measures obtained with **Protocol 1** and **Protocol 2**. The normalized contingency tables are shown in Table 2 (**Protocol 1**) and Table 3 (**Protocol 2**). With **Protocol 1**, median *Se* and *DSC* were 68.77% (IQR = 34.83%) and 54.00% (IQR = 41.03%), respectively. **Protocol 2** outperformed **Protocol 1** significantly (p-value < 0.05), with median *Se* and *DSC* equal to 88.07% (IQR = 18.84%) and 71.25% (IQR = 31.82%), respectively.

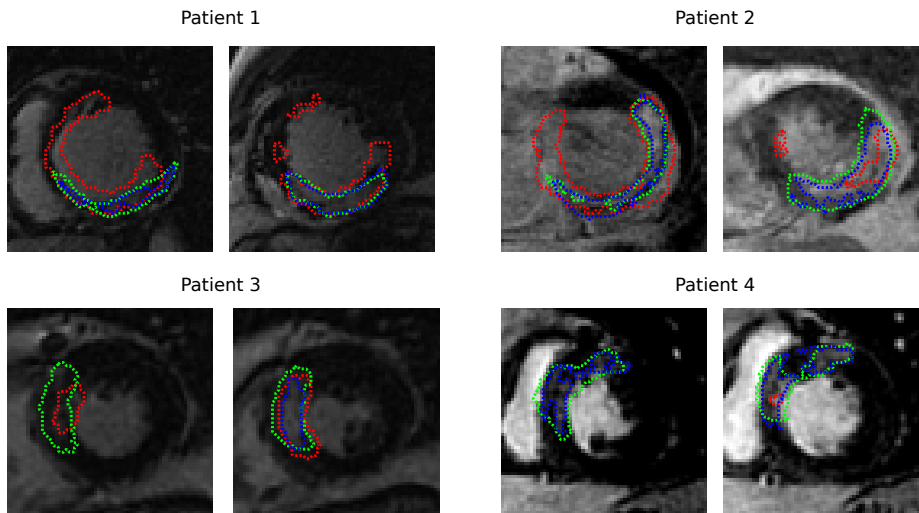


Fig. 5 Sample segmentation results obtained with **Protocol 1** (red contour) and **Protocol 2** (green contour) for four patient. The blue contour refers to ground-truth segmentation. Each row refers to a different patient.

Protocol 1 and **Protocol 2** failed in detecting the presence of scar tissue in 17 slices from 7 patients and 2 slices from 1 patient, respectively. **Protocol 1** and **Protocol 2** detected scar tissue when it was not present in 21 slices from 10 patients and 28 slices from 11 patients, respectively.

Some examples of segmentation outcome obtained with **Protocol 1** and **Protocol 2** for four patients, highlighting the observed scenarios, are shown in Fig. 5. In CMR-LGE slices from Patient 1 and Patient 2, the tendency of **Protocol 1** (red line) in overestimating GT scar contours (blue line) can be observed. In Patient 3, both **Protocol 1** and **Protocol 2** detected scar tissue while this was not evidenced by the expert cardiologist in three out of eight slices. A relevant example is shown in Fig. 5 bottom left. For Patient 4, **Protocol 1** failed (or barely succeeded) in detecting the presence of scar tissue in four slices out of nine. Two examples are shown in Fig. 5 bottom right.

5 Discussion

In this paper, the feasibility and accuracy of FCNNs for scar segmentation in CMR-LGE images were assessed. Accordingly, ENet was properly modified to be applicable to this kind of images, and evaluation of its performance in two parallel segmentation protocols was achieved. In this pilot study, as only 30 patients were included, data augmentation techniques allowed increasing the total number of available images up to 2000, and leave-one-patient-out cross-validation was the method of choice to guarantee proper analysis. Computational training time (about 30 hours) could be considered acceptable, considering this approach and the use of not optimized computer architecture. Both segmentation protocols were in general able to detect scar tissue in the CMR-LGE images. Nonetheless, the performance achieved in **Protocol 1** was lower than the one achieved in **Protocol 2**.

1 This result was expected, as scar-segmentation algorithms in the literature (both
 2 traditional and based on DL) commonly require a priori knowledge of the LV my-
 3 ocardial position, defined by its myocardial borders, to define the regions of search
 4 for segmentation (e.g. [7, 8, 17, 18]). Indeed, in the CMR-LGE images, several
 5 structures are present surrounding the LV. Such structures have similar intensity
 6 and texture with respect to the scar tissue, making the task of **Protocol 1** more
 7 challenging than the one of **Protocol 2**.

8 The results obtained with **Protocol 2** (median $DSC = 71.25\%$ (IQR =
 9 31.82%)) were in line with those (median $DSC = 64.00\%$ (IQR = 20%)) reported
 10 in [22] for seven semi-automatic approaches. The dataset in [22] was built with
 11 15 subjects, for a total of 124 CMR-LGE training images (from 5 subjects) and
 12 208 CMR-LGE testing images (from 10 subjects). A direct comparison was not
 13 possible as, to the best of the authors' knowledge, the GT masks for the dataset
 14 presented in [22] were not provided for the testing images. Nonetheless, with re-
 15 spect to the state of the art approaches, both **Protocol 1** and **Protocol 2** were
 16 fully automatic and parameter free. This is an undoubted advantage compared
 17 to threshold approaches (such as nST and FWHM) that require both user in-
 18 teraction for manual delineation of regions of interest in the LV myocardium for
 19 threshold computation and parameter tuning for setting the thresholds (e.g. num-
 20 ber of standard deviations for nST). Similarly, clustering techniques such as GMM
 21 need to define the number of GMM classes, which is not always trivial [38, 39].
 22 With respect to other DL-based methodologies, such as [24, 18, 25], our approach
 23 directly provided the segmentation mask without requiring (i) pre-processing to
 24 extract and (ii) post processing to merge superpixels or patches from the LV my-
 25 ocardial region. This was achieved by exploiting a fully-convolutional architecture
 26 instead of an architecture based on CNNs with fully-connected layers for classifi-
 27 cation tasks. This is widely recognized in the literature to simultaneously simplify,
 28 toughen and speed up both learning and segmentation [26].

29 A first limitation of this study, as also reported in [8], is related to the lack of
 30 a true GT for algorithm training and testing. In fact, having a GT for the scar-
 31 segmentation task is not trivial. Histological validation in animal models that has
 32 been considered as GT in other studies (e.g. [40, 12]) resulted to be inadequate
 33 for humans [41]. However, expert tracing of contours of the object of interest is
 34 widely considered as acceptable strategy to provide a reference for comparison.
 35 A second issue is related to the fact that our evaluation protocol was based on
 36 a limited number of CMR-LGE images. Even if such number was comparable
 37 with other approaches in the state of the art [22] (332 CMR-LGE images from
 38 15 different subjects), a larger training dataset would allow encoding variability
 39 in image characteristics and scar size and position, to increase the segmentation
 40 performance. However, this initial work has to be intended as a proof of concept
 41 for the described methodology, and we are currently working with our clinical
 42 partners on expanding our training dataset.

43 As future extension of this work, instead of focusing on CMR-LGE images, 3D
 44 FCNNs will be investigated to deal with the 3D information encoded in CMR-LGE
 45 data. In fact, 3D FCNNs were recently shown to provide encouraging results for
 46 magnetic-resonance volumes of prostate [42].

47 In conclusion, the proposed strategy for automated scar segmentation from
 48 CMR-LGE images based on FCNN showed a good performance, in particular
 49 once the process was guided by limiting the processing area to the myocardium
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65

1 only. These results are promising for application of deep learning techniques to
2 this kind of medical imaging, and constitute the basis for future research involving
3 larger training datasets.
4

5 6 7 **Author contributions**

8 SM: Study conception and design/Analysis and interpretation of data/Drafting of
9 manuscript/Critical revision. RB: Acquisition of data/Analysis and interpretation
10 of data/Critical revision. CM: Acquisition of data/Analysis and interpretation of
11 data. GM: Acquisition of data/Analysis and interpretation of data/Critical revision.
12 GP: Acquisition of data/Analysis and interpretation of data/Critical revision.
13 MP: Acquisition of data/Analysis and interpretation of data/Critical revision.
14 EGC: Study conception and design/Analysis and interpretation of data/Drafting
15 of manuscript/Critical revision
16

17 18 19 **Compliance with ethical standards**

20
21 The authors declare that they have no conflict of interest.
22
23

24 **Ethical approval**

25
26 All procedures performed in studies involving human participants were in accordance
27 with the ethical standards of the institutional and/or national research
28 committee and with the 1964 Helsinki Declaration and its later amendments or
29 comparable ethical standards. For this type of study, formal consent is not required.
30
31

32 33 34 **Informed consent**

35
36 Informed consent was obtained from all individual participants included in the
37 study.
38
39

40 **References**

- 41
42 1. Alexandre J, Saloux E, Dugué AE, Lebon A, Lemaitre A, Roule V, Labombarda F, Provost N, Gomes S, Scanu P, et al (2013) Scar extent evaluated
43 by late gadolinium enhancement CMR: A powerful predictor of long term appropriate ICD therapy in patients with coronary artery disease. *Journal of Cardiovascular Magnetic Resonance* 15(1):12
44
45 2. Kelle S, Roes SD, Klein C, Kokocinski T, de Roos A, Fleck E, Bax JJ, Nagel E (2009) Prognostic value of myocardial infarct size and contractile reserve using magnetic resonance imaging. *Journal of the American College of Cardiology* 54(19):1770–1777
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

3. Usta F, Gueaieb W, White JA, McKeen C, Ukwatta E (2018) Comparison of myocardial scar geometries generated from 2D and 3D LGE MRI. In: Medical Imaging 2018, International Society for Optics and Photonics, vol 10578, p 105780K
4. Dikici E, O'Donnell T, Setser R, White RD (2004) Quantification of delayed enhancement MR images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp 250–257
5. Kim RJ, Wu E, Rafael A, Chen EL, Parker MA, Simonetti O, Klocke FJ, Bonow RO, Judd RM (2000) The use of contrast-enhanced magnetic resonance imaging to identify reversible myocardial dysfunction. *New England Journal of Medicine* 343(20):1445–1453
6. Mewton N, Revel D, Bonnefoy E, Ovize M, Croisille P (2011) Comparison of visual scoring and quantitative planimetry methods for estimation of global infarct size on delayed enhanced cardiac MRI and validation with myocardial enzymes. *European Journal of Radiology* 78(1):87–92
7. Schulz-Menger J, Bluemke DA, Bremerich J, Flamm SD, Fogel MA, Friedrich MG, Kim RJ, von Knobelsdorff-Brenkenhoff F, Kramer CM, Pennell DJ, et al (2013) Standardized image interpretation and post processing in cardiovascular magnetic resonance: Society for cardiovascular magnetic resonance (SCMR) board of trustees task force on standardized post processing. *Journal of Cardiovascular Magnetic Resonance* 15(1):35
8. Carminati MC, Boniotti C, Fusini L, Andreini D, Pontone G, Pepi M, Caiani EG (2016) Comparison of image processing techniques for nonviable tissue quantification in late gadolinium enhancement cardiac magnetic resonance images. *Journal of Thoracic Imaging* 31(3):168–176
9. Hsu LY, Natanzon A, Kellman P, Hirsch GA, Aletras AH, Arai AE (2006) Quantitative myocardial infarction on delayed enhancement MRI. Part I: Animal validation of an automated feature analysis and combined thresholding infarct sizing algorithm. *Journal of Magnetic Resonance Imaging* 23(3):298–308
10. Hennemuth A, Seeger A, Friman O, Miller S, Klumpp B, Oeltze S, Peitgen HO (2008) A comprehensive approach to the analysis of contrast enhanced cardiac MR images. *IEEE Transactions on Medical Imaging* 27(11):1592–1610
11. Pop M, Ghugre NR, Ramanan V, Morikawa L, Stanisz G, Dick AJ, Wright GA (2013) Quantification of fibrosis in infarcted swine hearts by ex vivo late gadolinium-enhancement and diffusion-weighted mri methods. *Physics in Medicine & Biology* 58(15):5009
12. Fieno DS, Kim RJ, Chen EL, Lomasney JW, Klocke FJ, Judd RM (2000) Contrast-enhanced magnetic resonance imaging of myocardium at risk: Distinction between reversible and irreversible injury throughout infarct healing. *Journal of the American College of Cardiology* 36(6):1985–1991
13. Gerber BL, Garot J, Bluemke DA, Wu KC, Lima JA (2002) Accuracy of contrast-enhanced magnetic resonance imaging in predicting improvement of regional myocardial function in patients after acute myocardial infarction. *Circulation* 106(9):1083–1089
14. Setser RM, Bexell DG, O'Donnell TP, Stillman AE, Lieber ML, Schoenhagen P, White RD (2003) Quantitative assessment of myocardial scar in delayed enhancement magnetic resonance imaging. *Journal of Magnetic Resonance Imaging* 18(4):434–441

- 1 15. Lund GK, Stork A, Saeed M, Bansmann MP, Gerken JH, Muller V, Mester J,
2 Higgins CB, Adam G, Meinertz T (2004) Acute myocardial infarction: Eval-
3 uation with first-pass enhancement and delayed enhancement MR imaging
4 compared with ^{201}Tl SPECT imaging. *Radiology* 232(1):49–57
- 5 16. Hennemuth A, Friman O, Huellebrand M, Peitgen HO (2012) Mixture-model-
6 based segmentation of myocardial delayed enhancement MRI. In: *International*
7 *Workshop on Statistical Atlases and Computational Models of the Heart*,
8 Springer, pp 87–96
- 9 17. Grau V (2017) Automated LGE myocardial scar segmentation using
10 MaskSLIC supervoxels-replicating the clinical method. In: *Medical Image Un-*
11 *derstanding and Analysis*, Springer, vol 723, p 229
- 12 18. Yang G, Zhuang X, Khan H, Haldar S, Nyktari E, Li L, Wage R, Ye X,
13 Slabaugh G, Mohiaddin R, et al (2017) Fully automatic segmentation and ob-
14 jective assessment of atrial scars for longstanding persistent atrial fibrillation
15 patients using late gadolinium-enhanced MRI. arXiv preprint arXiv:170509529
- 16 19. Li Z, Chen J (2015) Superpixel segmentation using linear spectral clustering.
17 In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp 1356–
18 1363
- 19 20. Usta F, Gueaieb W, White JA, Ukwatta E (2018) 3d scar segmentation from
20 LGE-MRI using a continuous max-flow method. In: *Medical Imaging 2018:*
21 *Biomedical Applications in Molecular, Structural, and Functional Imaging*,
22 *International Society for Optics and Photonics*, vol 10578, p 105780U
- 23 21. Liu X, Shen Y, Zhao X, Zhang S (2017) Quantized segmentation of fibrotic
24 tissue of left atrial from delay-enhancement MRI images using level-set and
25 graph-cut. In: *IEEE International Conference on Machine Vision and Infor-*
26 *mation Technology*, IEEE, pp 23–27
- 27 22. Karim R, Bhagirath P, Claus P, Housden RJ, Chen Z, Karimaghloo Z, Sohn
28 HM, Rodríguez LL, Vera S, Albà X, et al (2016) Evaluation of state-of-the-
29 art segmentation algorithms for left ventricle infarct from late gadolinium
30 enhancement MR images. *Medical Image Analysis* 30:95–107
- 31 23. Maier-Hein L, Vedula SS, Speidel S, Navab N, Kikinis R, Park A, Eisenmann
32 M, Feussner H, Forestier G, Giannarou S, et al (2017) Surgical data science
33 for next-generation interventions. *Nature Biomedical Engineering* 1(9):691
- 34 24. Yang G, Zhuang X, Khan H, Haldar S, Nyktari E, Ye X, Slabaugh G, Wong
35 T, Mohiaddin R, Keegan J, et al (2017) Segmenting atrial fibrosis from
36 late gadolinium-enhanced cardiac MRI by deep-learned features with stacked
37 sparse auto-encoders. In: *Annual Conference on Medical Image Understanding*
38 *and Analysis*, Springer, pp 195–206
- 39 25. Zabihollahy F, White JA, Ukwatta E (2018) Myocardial scar segmentation
40 from magnetic resonance images using convolutional neural network. In: *Med-*
41 *ical Imaging 2018*, *International Society for Optics and Photonics*, vol 10575,
42 p 105752Z
- 43 26. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for se-
44 mantic segmentation. In: *IEEE Conference on Computer Vision and Pattern*
45 *Recognition*, IEEE, pp 3431–3440
- 46 27. Moccia S, De Momi E, El Hadji S, Mattos LS (2018) Blood vessel segmentation
47 algorithms – Review of methods, datasets and evaluation metrics. *Computer*
48 *Methods and Programs in Biomedicine* 158:71–91
- 49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1 28. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for
2 biomedical image segmentation. In: International Conference on Medical Im-
3 age Computing and Computer-Assisted Intervention, Springer, pp 234–241
- 4 29. Paszke A, Chaurasia A, Kim S, Culurciello E (2016) Enet: A deep neu-
5 ral network architecture for real-time semantic segmentation. arXiv preprint
6 arXiv:160602147
- 7 30. Lieman-Sifry J, Le M, Lau F, Sall S, Golden D (2017) Fastventricle: Cardiac
8 segmentation with Enet. In: International Conference on Functional Imaging
9 and Modeling of the Heart, Springer, pp 127–138
- 10 31. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the
11 inception architecture for computer vision. In: IEEE Conference on Computer
12 Vision and Pattern Recognition, pp 2818–2826
- 13 32. Xu B, Wang N, Chen T, Li M (2015) Empirical evaluation of rectified activa-
14 tions in convolutional network. arXiv preprint arXiv:150500853
- 15 33. Pedersen SJK (2007) Circular Hough transform. Aalborg University, Vision,
16 Graphics, and Interactive Systems 123:123
- 17 34. Ruder S (2016) An overview of gradient descent optimization algorithms. arXiv
18 preprint arXiv:160904747
- 19 35. Kinga D, Adam JB (2015) A method for stochastic optimization. In: Interna-
20 tional Conference on Learning Representations, vol 5
- 21 36. Schmidhuber J (2015) Deep learning in neural networks: An overview. *Neural*
22 *Networks* 61:85–117
- 23 37. Dice LR (1945) Measures of the amount of ecologic association between
24 species. *Ecology* 26(3):297–302
- 25 38. Sakamoto Y, Ishiguro M, Kitagawa G (1986) Akaike information criterion
26 statistics. Dordrecht, The Netherlands: D Reidel 81
- 27 39. Vrieze SI (2012) Model selection and psychological theory: a discussion of the
28 differences between the Akaike information criterion (AIC) and the Bayesian
29 information criterion (BIC). *Psychological Methods* 17(2):228
- 30 40. Amado LC, Gerber BL, Gupta SN, Rettmann DW, Szarf G, Schock R, Nasir
31 K, Kraitchman DL, Lima JA (2004) Accurate and objective infarct sizing by
32 contrast-enhanced magnetic resonance imaging in a canine myocardial infarct-
33 ion model. *Journal of the American College of Cardiology* 44(12):2383–2389
- 34 41. Flett AS, Hasleton J, Cook C, Hausenloy D, Quarta G, Ariti C, Muthurangu V,
35 Moon JC (2011) Evaluation of techniques for the quantification of myocardial
36 scar of differing etiology using cardiac magnetic resonance. *Journal of the*
37 *American College of Cardiology* 4(2):150–156
- 38 42. Milletari F, Navab N, Ahmadi SA (2016) V-net: Fully convolutional neural
39 networks for volumetric medical image segmentation. In: IEEE Fourth Inter-
40 national Conference on 3D Vision, IEEE, pp 565–571
- 41
- 42
- 43
- 44
- 45
- 46
- 47
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60
- 61
- 62
- 63
- 64
- 65


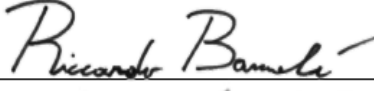

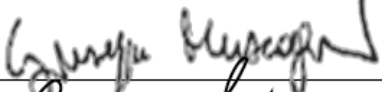
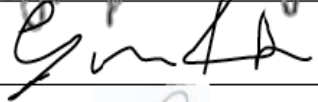


Author's Agreement

To Prof. David G. Norris
Editor-in-Chief Magnetic Resonance Materials in Physics, Biology and Medicine

Article Title: Development and testing of a deep learning-based strategy for scar segmentation on CMR-LGE images.

Authors: Sara Moccia, Riccardo Banali, Chiara Martini, Giuseppe Muscogiuri, Gianluca Pontone, Mauro Pepi, Enrico Gianluca Caiani

All authors having signed below confirm that they meet the criteria of authorship and that they have read and approved the manuscript.

Name	Signature	Date
Sara Moccia		27 July 2018
Riccardo Banali		30 July 2018
Chiara Martini		30 July 2018
Giuseppe Muscogiuri		29 July 2018
Gianluca Pontone		27 July 2018
Mauro Pepi		30 July 2018
Enrico Gianluca Caiani		6 August 2018