Stacked BCDU-net with semantic CMR synthesis: application to Myocardial Pathology Segmentation challenge

Carlos Martín-Isla⋆[1], Maryam Asadi-Aghbolaghi⋆[2], Polyxeni Gkontra[3], Victor M. Campello[1], Sergio Escalera[1,3], and Karim Lekadir[1]

[1] Departament de Matemàtiques & Informàtica, Universitat de Barcelona, Spain
carlos.martinisla@ub.edu
[2] Institute for Research in Fundamental Sciences (IPM), Iran
[3] Computer Vision Center, Univeritat Autònoma de Barcelona, Spain

Abstract. Accurate segmentation of pathological tissue, such as scar tissue and edema, from cardiac magnetic resonance images (CMR) is fundamental to the assessment of the severity of myocardial infarction and myocardial viability. There are many accurate solutions for automatic segmentation of cardiac structures from CMR. On the contrary, a solution has not as yet been found for the automatic segmentation of myocardial pathological regions due to their challenging nature. As part of the Myocardial Pathology Segmentation combining multi-sequence CMR (MyoPS) challenge, we propose a fully automatic pipeline for segmenting pathological tissue using registered multi-sequence CMR images sequences (LGE, bSSFP and T2). The proposed approach involves a two-staged process. First, in order to reduce task complexity, a two-stacked BCDU-net is proposed to a) detect a small ROI based on accurate myocardium segmentation and b) perform inside-ROI multi-modal pathological region segmentation. Second, in order to regularize the proposed stacked architecture and deal with the under-represented data problem, we propose a synthetic data augmentation pipeline that generates anatomically meaningful samples. The outputs of the proposed stacked BCDU-NET with semantic CMR synthesis are post-processed based on anatomical constrains to refine output segmentation masks. Results from 25 different patients demonstrate that the proposed model improves 1-stage equivalent architectures and benefits from the addition of synthetic anatomically meaningful samples. A final ensemble of 15 trained models show a challenge Dice test score of 0.665±0.143 and 0.698±0.128 for scar and scar+edema, respectively.

Keywords: Cardiac Magnetic Resonance · Myocardial Pathology Segmentation · Deep Learning · BCDU-Net · LGE · bSSFP · T2.

⋆ Both authors contributed equally to this work.

1 Introduction

Myocardial viability assessment is key in the diagnosis of patients suffering from myocardial infarction and ischemic heart disease, among others. Cardiovascular magnetic resonance (CMR) is a well-established imaging technique that provides anatomical and functional information of the heart. Multiple sequences with different properties can be acquired, registered and combined to obtain a complete viability assessment. Late gadolinium enhancement magnetic resonance imaging (LGE-MRI) is widely used to assess presence, location and extent of regional scar or fibrotic tissue in the myocardium. T2-weighted CMR images are able to identify edema and acute or recent myocardial ischemic injury, and have been employed to distinguish acute coronary syndrome (ACS) from non-ACS as well as acute from chronic myocardial infarction. On the other hand, balanced - Steady State Free Precession (bSSFP) cine sequence presents clear boundaries for the cardiac anatomical regions, often unclear in the first two modalities due the presence of pathological regions.

LGE and T2-weighted are well-established techniques to many CMR examinations, but there are challenges in their quantification and interpretation due to a variety of factors. First, image analysis depends on image quality which can be affected by CMR acquisition protocol. Suboptimal parameters such as inversion time (TI), repetition time (TR), echo time (TE) need to be correctly identified in order to maximize the difference in intensity curves between pathological and non pathological regions, but also to minimize inter-subject acquisitions variability. Additionally, timing after contrast administration in LGE is important to allow sufficient wash-out of the contrast agent. On top of that, the variability in morphology and texture of infarcted, edemic areas and the combination of both leads to a difficult automation of the process. For this reason, manual and automated techniques with no user interaction for infarct borders detection often results in significant within-patient variability [1–4].

In order to explore the complementary nature of existing modalities for the purpose of myocardial pathology segmentation, the MyoPS challenge is proposed. It includes a challenging data distribution of 45 multi-modality subjects with the goal of doing an accurate automatic infarcted and edemic regions segmentation.

In this work, we propose a challenge solution based on a stacked BCDU-NET late fusion architecture including localisation and segmentation stages. Additionally, we tackle the insuffcent training size by means of state-of-the-art generative adversarial models [5,6]. To do so, we propose an image synthesis strategy based on Semantic Image Synthesis with Spatially-Adaptive Normalization [7]. The results demonstrate that the proposed model improves 1-stage equivalent architectures and benefits from the addition of synthetic anatomically meaningful samples.

2 Materials and methods

2.1 Dataset

A set of 45 cases of multi-sequence CMR are collected for the challenge. Each case refers to a patient with three CMR sequences, i.e., LGE, T2 and bSSFP CMR. All clinical data have got institutional ethic approval and have been anonymized. The data released have been pre-processed using the MvMM method [9, 10] to align the three-sequence CMR into a common space and to resample them into the same spatial resolution.

The provided gold standard labels of interest for the challenge are LV myocardial edema (label 1220) and LV myocardial scars (label 2221). Additional annotations of cardiac structures are provided: left ventricular (LV) blood pool (label 500), right ventricular blood pool (label 600) and LV normal myocardium (label 200). Thus, the evaluation of the test data will be focused on the myocardial pathology segmentation, i.e., scars and edema. The inter-observer variation of manual scar segmentation, in terms of Dice, was 0.5243±0.1578, which gives an insight of the difficulty of the task.

2.2 Proposed Method

An overview of the proposed automated segmentation method is presented in Figure 1. The approach consists of two stacked segmentation networks. In brief, after preprocessing, we employ a computationally efficient U-Net [12] on the bSSFP CMR to localize the rounded shape of myocardium which includes the LV normal myocardium, LV myocardial edema and scar tissue. Subsequently, the bSSFP, T2-weighted and LGE CMR are cropped using the bounding box of the localized myocardium. Histogram normalization is then applied on the cropped part of imgages. During the second stage, the cropped multi-sequence CMR is passed to a higher capacity model, the BCDU-Net [11], to segment the myocardium scar and edema. The output is finally post-processed based on anatomical constrains to refine output segmentation masks. The individual stages are explained in detail in the following sections.

Preprocessing Before the training process, all images were cropped so that they had a pixel size of 256 × 256. Furthermore, all images were normalised between 0 and 1 within the Region Of Interest (ROI) for each independent modality.

Localization Network The pathological tissue is located within LV blood pool and LV normal myocardium. Therefore, we first employ a network to localize the myocardial ROI, i.e. a binary segmentation, using cine-MRI as the input modality. Cine-MRI was chosen over the other modalities for this task because it is the most accurate for myocardial boundary detection due to its clear structure definition and lack of appearance of pathological regions. This task will reduce

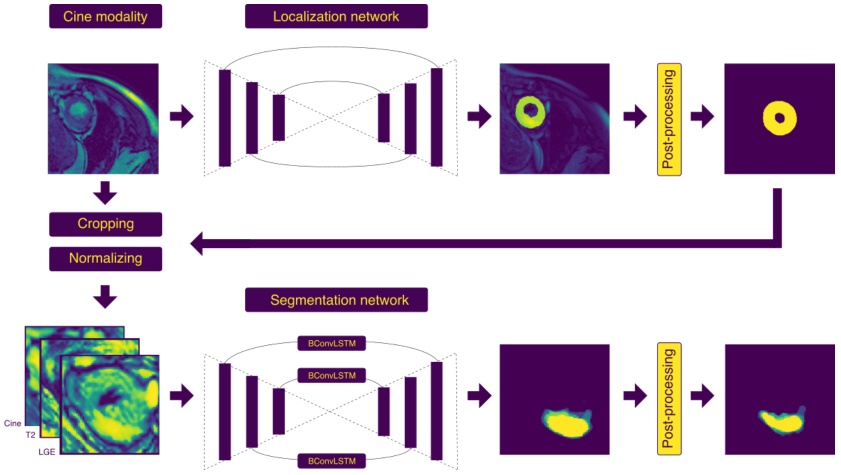

Fig. 1. Overview of the proposed stacked network.

the search space when dealing with scar and edema segmentation by the stacked network. To do that, the myocardium, edema, and scar labels are considered as the foreground, and the other labels (left ventricular blood pool, right ventricular blood pool) as the background. U-Net, [12], is a popular convolutional network architecture for fast and precise segmentation of images which is built upon the Fully Convolutional Network (FCN). The main advantages of this network is that is capable to work well with few training samples, and the network has the potential to make use of the global location and context information at the same time.

This symmetric network is separated in three parts of encoding (contracting), Bottleneck, and decoding (expanding) paths. The encoding path is composed of 4 blocks. In each block we have two 3×3 convolutional layers followed by one 2×2 Max Pooling function and ReLU. In each block, the number of feature maps are doubled, and the size of feature get half. The contracting path aims at progressively capturing context of the input image and increasing the dimension of feature representation block by block. These coarse contextual information are then transferred into the decoding path through skip connections. The output of the last block of the encoder is first passed to the bottleneck which is built by two 3×3 convolutional layers. At the end of bottleneck we have a high dimensional image representation with high semantic information.

The decoding path is composed of four blocks. Each block starts with performing a deconvolution (up-sampling) over the output of previous layer. The corresponding feature maps in the encoding path are then copied to this layer, and are then concatenated with the output of deconvolutional layer. These features are then go through one 3×3 convolutional layers. In each block of the decoder, the size of the feature maps gradually increases and the number of feature maps gradually decreases. The target of decoder in U-Net is to enable precise localisation by using transposed convolutions and recovering the size of the segmentation. Since that data is imbalanced and most of the pixels have background label, we use the weighted binary cross entropy loss to train the network.

In our U-net implementation, for efficiency purposes, the number of classes is used as the number of feature maps in the deconvolutions of the decoding path, as shown in [8,14]. It is also worth mentioning that we do not need a very accurate segmentation result here, since we just crop the smallest bounding box around the myocardium with a small margin of 10 pixels.

Normalisation The output of the localisation network provides the approximate location of the myocardial region. Therefore, by considering the fact that the myocardial infarcted and edemic regions are within such ROI, we can ignore unwanted background information by finding the smallest bounding box with a small margin around the myocardium. Moreover, an histogram equalisation is applied by modality, avoiding the effect of unuseful background pixels in the pixel histogram redistribution.

Segmentation We exploit the BCDU-Net [11] to segment the myocardial scar and edema from the normalized myocardium of the three input modalities. The BCDU-Net is an extension of U-Net by including bidirectional convolutional LSTM (BConvLSTM) [13] in the skip connection and reusing feature maps with densely convolutions. The output features of the deconvolutional layer contain more semantic information while the features extracted by the corresponding encoding layer have higher resolution. To combine these two kinds of features, the authors replaced the simple concatenation of the skip connection with nonlinear functions, i.e. BConvLSTM in the BCDU-Net which resulted in more precise segmentation output.

Moreover, the idea of densely connected convolutions is utilized in the bottleneck of the BCDU-Net. By having a sequence of convolutional layers, the network may learn redundant features, therefore, in the bottleneck of the BCDU-Net, features which are learned in each block are passed forward to the next block. The dense blocks help the method to enhance information flow and learn a diverse set of features based on the collective knowledge gained by previous layers. Furthermore, the convergence speed of the network is accelerated by employing Batch Normalization (BN) after the up-convolution filters.

Like U-Net, the encoding path of the BCD-Net includes four steps. Each step consists of two 3×3 convolutional filters followed by a 2×2 max pooling function and ReLU. The depth of feature maps are doubled at each step and the size of each feature map get half. There are two states of BConvLSTM in the skip connection of the BCDU-Net. The second state receives the output of the previous deconvolutional function and the input data of the first one its corresponding feature maps in the encoding path. The output of the second BConvLSTM is then passed to the two 3×3 convolutional filters. Like original

U-Net, the decoding path doubles the size of each feature map and halves the number of feature channels layer by layer to reach the original size of the input image after the final layer. To train the network, we use Dice score-based loss.

We propose to combine the three input modalities with a late fusion approach. In other words, the network is trained separately for the three modalities and before the last convolutional layer after the last deconvolutional layer, the three networks are merged.

Implementation Details All trainings were performed on a NVIDIA 1080 GPU with a batch size of 8. The Adam optimization function with learning rate equal to $1e-4$ was used to train both networks. Each network is trained with 50 as the number of epochs. The input size was 256 × 256 for both localization and segmentation networks.

2.3 Data augmentation strategy

Online augmentation A series of common augmentation techniques were applied to each batched image independently. For the first stacked u-net, these augmentations included random rotations between -15^{o} and 15^{o} and random scaling and offsets of a maximum of 30 pixels. For the second stacked u-net the offset augmentation is avoided due to the fact that images were already center-cropped.

Offline augmentation The rationale behind the proposed image synthesis is the insufficient training sample size. Low number of images, variability in modality acquisitions, in location and extent of pathological regions can cause loss of generalisation in CNN-based segmentation algorithms. Thus, in an effort to increase the number of annotated multi-sequence images, semantic image synthesis from annotated mask to multi-sequence CMR is performed in such way that new multi-modality images can be generated from altered versions of real annotations. To achieve this, the Semantic Image Synthesis with Spatially-Adaptive Normalization (SPADE) method [7] was implemented using the PyTorch library provided at this link†. Previous methods [6] directly feed the semantic layout as input to the deep network, which is then processed through stacks of convolution, normalization, and nonlinearity layers. In [7], is shown that this is suboptimal as the normalization layers tend to wash away semantic information, desired for accurate pathology tissue and cardiac structure generation. To address the issue, SPADE uses the input semantic annotation for modulating the activations in normalization layers through a spatially-adaptive, learned transformation. A general overview of the SPADE multi-modality generative model is represented in Figure 2.

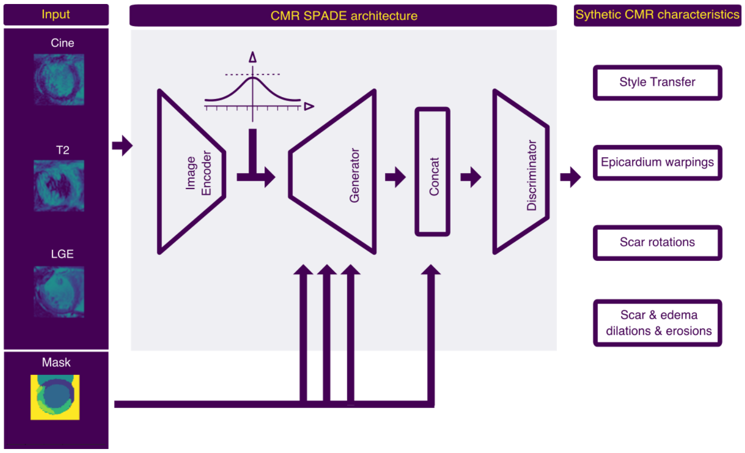

Fig. 2. Overview of the proposed SPADE generative model.

Two SPADE models were generated. For the training/validation subset, a model with 71 training images (17 subjects) was used and 31 validation images (8 subjects) were kept aside. For the final model, all the subjects were used to train an additional SPADE model.

Both models were trained during 45 epochs with a morphological augmentation consisting of warping epicardium contours between pairs of subjects. Both trainings took 24 hours on a NVIDIA 1080 GPU with a batch size of 2. The Adam optimizer was used with learning rate of 2x10$e-4$, with first and second moment decay rates of 0 and 0.9, respectively. The Variational Autoencoder (VAE) was generated with a latent dimension of 200.

Once the models were trained, a set of morphological operations were defined in order to generate different versions of real annotations. The resulting anatomical consistent annotations were used then to feed the SPADE models and generate synthetic multi-modality images with controlled characteristics:

Style transfer. By training the SPADE with a Variational Autoencoder (VAE), the style of the images can be transferred, generating a variety of images with different pathology appearances for the same morphology. The encoder and generator of our SPADE architecture form a VAE, in which the encoder tries to capture the style of the image, while the generator combines the encoded style and the segmentation mask information via the SPADEs to reconstruct the original image. The encoder also serves as a style guidance network at test time to capture the style of target images. For training the VAE, KL-Divergence loss term was used.

Every training image was used to generate a set of latent representations of size 200. The latter were used alone -with random linear combinations and scaling factors- or in conjunction with the methods described below in order to produce the final synthetic multi-modality images. The effect of this technique is shown in Figure 3, where an original image in first row is transferred to two additional pseudo-random styles, rows 2 and 3.

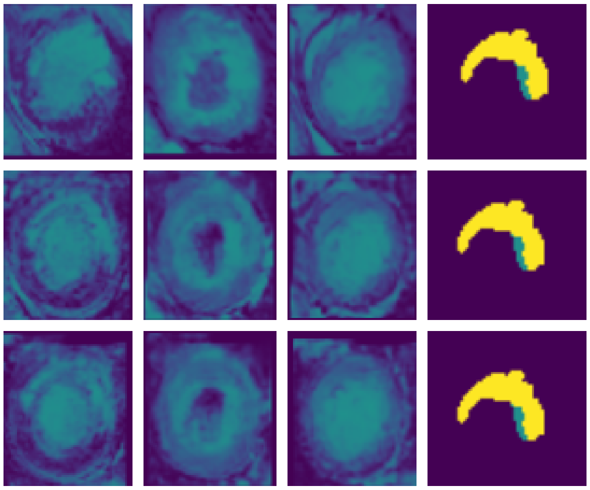

Fig. 3. Style modifications.

Epicardium warpings. As shown in Figure 4, a set of 8 equidistant landmarks were placed in the epicardial contour of the source and target annotations. Epicardial contours were then warped between pairs of training subjects by means of piecewise affine transformations.

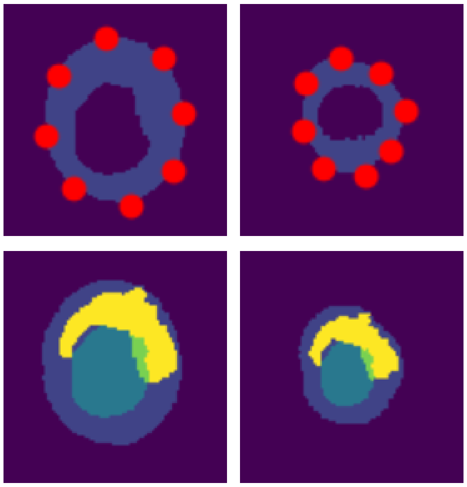

Fig. 4. Epicardial contour warping between a pair of subjects.

Scar and edema rotations. As shown in Figure 7, scar, edema and myocardium labels were combined in a binary mask. The epicardium was then converted to a circular shape, rotated and reconverted to the original shape taking profit of the same technique used in the *Epicardium warpings* section. This set of transformations was then also applied to the original labels, generating a rotated version of the scar and edema within the myocardium. To ensure that the generated segmentations were not too far from the distribution seen by the SPADE generator while covering the label space, the rotation was fixed to four possible values of [-30°, -20°,20°,30°].

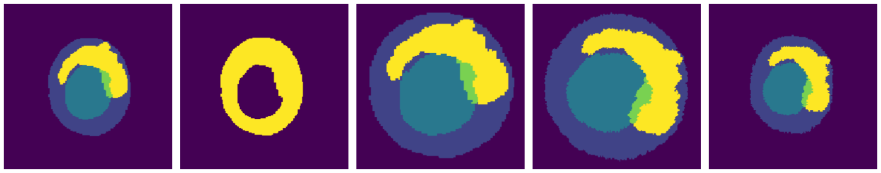

Fig. 5. Morphological operations involved in the scar rotation process.

Scar and edema dilations and erosions. A set of random complementary dilations and erosions with a random kernel radius from 1 to 3 pixels were applied to the training annotations. By fixing one of them for the scar label and applying the opposite one for the edema label, we avoid an empty gap between both. Random deletion of edemic labels is also included in this stage. In Figure 6 shows the effect of an eroded scar and dilated edema.

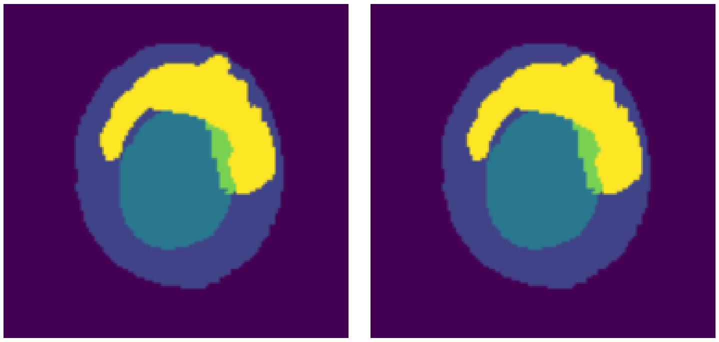

Fig. 6. Morphological operations involved in the scar and edema dilation and erosion process.

Offline datasets A group of datasets is generated by means of the augmentation strategies described above. More precisely, for each of the transformable labels, i.e. non-empty annotations, the original images are used up to three times to keep the training size relatively small. This methodology leads to the creation of a set of four datasets, one per type of augmentation, i.e. style transfer alone, pathology rotations, epicardial warping and pathology dilation/erosion. It should be noted that the resulting datasets contain the same amount of real and synthetic data. Additionally, for all datasets, random style transfers are applied after the annotation manipulation in the synthesis stage. In total, each dataset contains 415 images. A fifth dataset is generated by combining all individual four datasets. This dataset consists of 1660 images and is used to train and validate the models. The same procedure is repeated for the final ensemblea using the SPADE trained over all the training data. This leads to datasets of 597 and 2388 images, for the partial augmentations and the addition, respectively.

2.4 Post-processing

The myocardium, scar and edema-scar segmentations produced from the stacked networks were morphologically processed to satisfy certain anatomical constraints. In short axis CMR, the shape of the myocardium closely resembles that of a ring throughout the apex-base slices. Therefore, slices for which the automatically segmented myocardium is a partial ring must be detected and corrected. To this end, the skeleton of the myocardium was calculated for each slice. Subsequently, spur skeleton branches, i.e. branches consisting of pixels with only one neighboring pixel, were iteratively pruned. For non-complete rings, iterative pruning results in the removal of the entire skeleton. In such cases, the missing arc of the partial ring was completed by adding a circular ring whose thickness is equal to the maximum thickness of the detected myocardium. To construct the ring, the centroid of the convex hull of the detected myocardial region was used as its center. The thickness of the myocardium was given by the distance of the skeleton points to the closest non-myocardial pixel and the maximum among all points was considered. The corrected myocardium was subsequently used to refine the scar segmentation, while an additional step was necessary in the case of the edema-scar region. More precisely, edema can be noticed in the myocardium, but also in the LV blood pool close to the border with the myocardium. Therefore, an extended myocardial mask was created, which contained neighboring LV regions where edema could be localized. In order to achieve this, an artificial ring was constructed by using the myocardium skeleton and the distance of every pixel to it. Pixels belonging to the myocardium or the region enclosed by it were considered to belong to the extended myocardial mask if they were within a distance smaller than a threshold from the skeleton points. This threshold is defined as the maximum myocardium thickness plus a small margin of 6 pixels to account for errors in the myocardium segmentation.

As a first step in the process of refining the scar tissue, 3D components smaller than 100 voxels were considered to be artifacts and were, therefore, excluded from the segmentation mask. Despite good localization of the scar region by the network, we observed a tendency to underestimate the scar region and to produce multiple disconnected components instead of one continuous region. To tackle this issue, the components were connected by using their convex hull in cases where the output of the network consisted of more than one connected components. The area of the convex hull inside an eroded version of the extended myocardium was eliminated. For the erosion, a disk element with radius equal to 20% of the maximum myocardium radius was used. Furthermore, morphological closing of the image with a disk object of radius equal to 90% of the myocardium maximum thickness was performed to enlarge the component's border without losing the form of the original shape boundary in cases where only one component was observed. Lastly, areas outside the corrected myocardium and the joined edema-scar mask regions were excluded from the final scar segmentation.

In the case of the refinement of the joined edema-scar mask, 3D components of size smaller than 300 voxels were considered as artifacts. In addition, regions of edema-scar outside the extended myocardial area were excluded from the final segmentation by performing element-wise multiplication of the artificial extended myocardium region mask with the edema-scar segmentation.

3 Results

3.1 Protocol and Metrics of the challenge

In order to train our models and generate the ablation study, the training set is divided in two partitions. From the original 25 subjects, 8 of them are kept aside for validation, with the aim of preserving a large pool of subjects in the validation stage. The decision is motivated by the variability in image quality and the presence of difficult cases that may lead to a sub-optimal model selection. Moreover, this allows us to have a sufficient validation size to evaluate the post-processing algorithm. For the same reason, we avoided to preserve a test partition that leads to a conflict between validation and testing results and generates additional uncertainty when selecting the best method. After model generation, selection, evaluation and post-processing, 3D Dice scores are computed to select the final models taking into consideration the post-processing gains. For all the experiments, 2D Dice score is used as objective loss function, except for the localisation U-net, where the selected loss is binary weighted cross-entropy.

3.2 Ablation study

We performed a detailed ablation study in order to quantify the effect of every component of the proposed methodology individually. The results in terms of 2D Dice score (mean ± standard deviation), which is the accuracy evaluation metric used in the loss function of this work, are summarized in Table 1. In brief, our first experiment involved segmenting the scar and scar+segmentation using solely the original data without performing inter-stage normalization or offline augmentation. This resulted in a Dice score equal to 0.202 ± 0.286 and 0.170 ± 0.253 for scar and scar+edema, respectively. The low accuracy demonstrates the extremely challenging nature of the task and the need for incorporating a ROI-based normalization between stages and novel augmentation strategies. To test our assumption, we added the inter-stage cropping and normalization step to enhance the contrast between scar and edema and the rest of the tissue within the myocardial ROI where the pathological tissue is expected to localized. The mean dice score increased by 24.70% for scar and 33.80% for scar+edema.

We then compared the improvement offered by any of the four types of offline augmentation, i.e. style transfer alone, pathology rotations, epicardial warping and pathology dilation/erosion. Style transfer produced an improvement in terms of Dice by 9% and 14.4% for scar and scar+edema, respectively. The effects of epicardium warping and scar and edema rotation, were lower than that of style-transfer, but yet non-negligable. More precisely, the mean dice increased by 4.1% for scar and 7.8% for scar+edema in the case of epicardium warping. Similarly, when scar and edema rotation were applied the offered improvement was 1.7% for scar and 4.6% for scar+edema. Interestingly, scar and edema dilation and erosion did not provide any significant improvement in the scar tissue, but offered a 10.4% mean improvement in Dice for the scar+edema region. Subsequently, we combined the four types of data-augmentation. We observed a Dice score

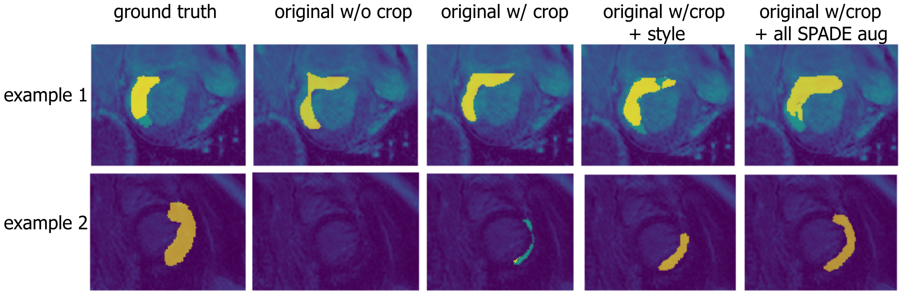

Fig. 7. Segmentation examples combining different sets of training data, showing the improvement of SPADE synthesis.

of 0.518 ± 0.286 and 0.617 ± 0.253 for scar and scar+edema, respectively. This indicates that for the case of pathological tissue segmentation the most effective augmentation type is style transfer, while morphological augmentations have a more limited effect. We speculate that this might be related to the highly irregular shape of the pathological tissue. However, these types of morphological augmentations might be important in other more regular structures. In this work, to account for possible variability found in the test sample non present in the training set, for the final model, we decided to use the combination of all augmentation types, presented as "All spade" in Table 1. Nonetheless, future work will focus on using the style transfer only for pathological tissue segmentation.

Lastly, we evaluated the improvement offered by applying post-processing on the outputs of the localization and segmentation networks. A visual example of the improvement can be seen in Figure 8. Post-processing produces a continuous scar region, while both edema and scar after post-processing are localized within the myocardial area and in the close vicinity of left ventricle, as physiologically expected.

Table 1. 2D Dice score (mean ± standard deviation) of the proposed method for scar and scar+edema for different data.

Data	Scar	Scar + Edema
Original data	0.202 ± 0.286	0.170 ± 0.253
Original data + cropping and normalizing	0.449 ± 0.261	0.508 ± 0.243
Style transfer	0.548 ± 0.250	0.640 ± 0.192
Epicardium warping	0.490 ± 0.260	0.586 ± 0.222
Scar and edema rotation	0.466 ± 0.241	0.554 ± 0.224
Scar and edema dilation and erosion	0.458 ± 0.299	0.600 ± 0.224
All spade	0.518 ± 0.286	0.617 ± 0.253

3.3 Challenge results

In order to obtain the final predictions, two ensembles are generated. For the first ensemble, a set of 5 models is generated with 10 consecutive training samples and 5 consecutive validation samples, with a roll factor of 5. For the second ensemble, a set of 15 models is generated with 22 consecutive training samples and 3 consecutive validation subjects, with a roll factor of 2, making the validation set to share one subject between consecutive models in the case of the 15 models ensemble.

The confidence maps of each one of the 5 models are averaged together. The final predictions of the 20 unseen test subjects provided by the challenge organization are defined as the maximum probability of each pixel belonging to each class, maximizing the expected results and reducing the variance. The same procedure was applied to the 15 models ensemble. After that, post-processing, as described in Section 2.4, is applied to further enhance the model's output. The effect of the ensemble size can be observed in Table 2. The bigger ensemble obtained better results due to the bigger training sizes. The effect of the low validation size was noticeable as a noisier validation curve, and attenuated by means of a greater regularisation power, with an overall improved accuracy. The quantitative effect of post-processing is also appreciated. The 15 models ensemble captured a greater number of non-trivial unconnected components. In combination with the convex hull process described in Section 2.4, for the 15 models ensemble the post-processing generated an improvement in accuracy of 2.9% for scar and 1.1% for scar+edema, respectively.

Table 2. 3D Dice score for the final testing set of 20 subjects.

Data	Scar	Scar + Edema
5 models ensemble	0.625±0.255	0.677±0.146
5 models ensemble + post-processing	0.635±0.281	0.692±0.143
15 models ensemble	0.636±0.243	0.687±0.131
15 models ensemble + post-processing	0.665±0.241	0.698±0.128

4 Discussion

This work proposes a novel approach to address automatic multi-sequence CMR pathology segmentation. The method is based on a two-staged process and leverages advanced state-of-the-art deep learning techniques. CMR pathology segmentation is a particularly challenging task even for the expert clinician due to the large variability in imaging quality and morphology of pathological regions. To tackle this limitation, we focus on reducing the task complexity. To this end, a localisation U-net is used to localize the myocardial ROI. Subsequently, the detected ROI is used to partially address the problem of intra- and inter-subject

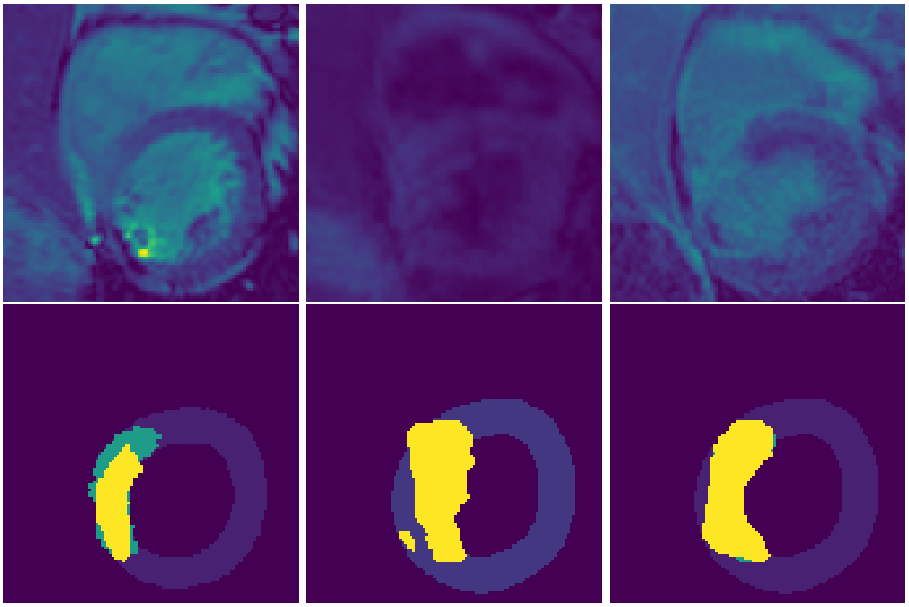

Fig. 8. Improvement offered by applying post-processing on the outputs of the localization and segmentation networks. On the top row, a slice from the bSSFP (left), T2-weighted (middle) and LGE (right) CMR are provided for one subject of the training dataset used as validation subject during training. On the bottom row, the corresponding manual segmentations for myocardium, scar and edema (left), the combined output of the two networks before (middle) and after (right) post-processing are provided. Post-processing permits to connect the two disconnected components produced by the network and constrain the segmentation within the myocardial area and neighboring LV area.

variability in signal intensity by using the bounding box of the ROI to crop the CMR images and perform a refined normalisation within the cropped region. The normalised CMR are then fed to a BCDU-net in order to perform the pathologic tissue segmentation. BCDU-net effectiveness has been previously demonstrated and is related to the bidirectional flow of the gradient. In addition, we address the problem of insufficient training examples by means of multi-modality semantic image synthesis using morphological and style transformations. This approach increases the variability of the training samples in terms of the location of the infarcted and edemic tissues within the myocardium, as well as, in terms of their appearance. The validation shows the effect of the stacked architecture with inter-stage normalisation, giving an insight about the importance of standarisation for multi-modality medical imaging acquisitions. Moreover, consistent results across the different semantic manipulations and their respective synthesis, indicate the potential of this set of transformations for enriching and improving generalization of multi-modality cardiac pathology segmentation algorithms. Future work includes the implementation of an end-to-end model as well as the exploration of the generated synthetic data in detail with the aim of enhancing interpretability and quality of the image synthesis methods.

5 Acknowledgements

This work was partly funded by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 825903 (euCanSHare project). This work has been partially supported by the Spanish project PID2019-105093GB-I00 (MINECO/FEDER, UE) and CERCA Programme/Generalitat de Catalunya.). This work is partially supported by ICREA under the ICREA Academia programme. KL is supported by the Ramon y Cajal Program of the the Spanish Ministry of Economy and Competitiveness under grant no. RYC-2015-17183.

References

1. Klem, I., Heiberg, E., Van Assche, L., Parker, M. A., Kim, H. W., Grizzard, J. D., ... Kim, R. J. (2017). Sources of variability in quantification of cardiovascular magnetic resonance infarct size-reproducibility among three core laboratories. Journal of Cardiovascular Magnetic Resonance, 19(1), 62.
2. Thiele, H., Kappl, M. J., Conradi, S., Niebauer, J., Hambrecht, R., Schuler, G. (2006). Reproducibility of chronic and acute infarct size measurement by delayed enhancement-magnetic resonance imaging. Journal of the American College of Cardiology, 47(8), 1641-1645.
3. Flett, A. S., Hasleton, J., Cook, C., Hausenloy, D., Quarta, G., Ariti, C., ... Moon, J. C. (2011). Evaluation of techniques for the quantification of myocardial scar of differing etiology using cardiac magnetic resonance. JACC: cardiovascular imaging, 4(2), 150-156.
4. Tao, Q., Piers, S. R., Lamb, H. J., van der Geest, R. J. (2015). Automated left ventricle segmentation in late gadolinium-enhanced MRI for objective myocardial scar assessment. Journal of Magnetic Resonance Imaging, 42(2), 390-399.

5. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).
6. Wang, T. C., Liu, M. Y., Zhu, J. Y., Tao, A., Kautz, J., Catanzaro, B. (2018). High-resolution image synthesis and semantic manipulation with conditional gans. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8798-8807).
7. Park, T., Liu, M. Y., Wang, T. C., Zhu, J. Y. (2019). Semantic image synthesis with spatially-adaptive normalization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2337-2346).
8. Campello, V. M., Martín-Isla, C., Izquierdo, C., Petersen, S. E., Ballester, M. A. G., Lekadir, K. (2019, October). Combining Multi-Sequence and Synthetic Images for Improved Segmentation of Late Gadolinium Enhancement Cardiac MRI. In International Workshop on Statistical Atlases and Computational Models of the Heart (pp. 290-299). Springer, Cham.
9. Zhuang, X. (2018). Multivariate mixture model for myocardial segmentation combining multi-source images. IEEE transactions on pattern analysis and machine intelligence, 41(12), 2933-2946.
10. Zhuang, X. (2016, October). Multivariate mixture model for cardiac segmentation from multi-sequence MRI. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 581-588). Springer, Cham.
11. Azad, R., Asadi-Aghbolaghi, M., Fathy, M., Escalera, S. (2019). Bi-directional ConvLSTM U-net with Densley connected convolutions. In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. pp. 406-415).
12. Ronneberger, O., Fischer, P., Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
13. Song, H., Wang, W., Zhao, S., Shen, J., Lam, K. M. (2018). Pyramid dilated deeper convlstm for video salient object detection. In Proceedings of the European conference on computer vision (ECCV) (pp. 715-731).
14. Baumgartner, C. F., Koch, L. M., Pollefeys, M., Konukoglu, E. (2017, September). An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation. In International Workshop on Statistical Atlases and Computational Models of the Heart (pp. 111-119). Springer, Cham.