Adversarial Convolutional Networks with Weak Domain-Transfer for Multi-sequence Cardiac MR Images Segmentation

Jingkun Chen[1,4], Hongwei Li[2], Jianguo Zhang[1,3,4(✉)], and Bjoern Menze[2]

[1] Southern University of Science and Technology, Shenzhen, China
chenjk@mail.sustech.edu.cn, zhangjg@sustech.edu.cn
[2] Technical University of Munich, Munich, Germany
hongwei.li@tum.de
[3] Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen, China
[4] University of Dundee, Dundee, UK

Abstract. Analysis and modeling of the ventricles and myocardium are important in the diagnostic and treatment of heart diseases. Manual delineation of those tissues in cardiac MR (CMR) scans is laborious and time-consuming. The ambiguity of the boundaries makes the segmentation task rather challenging. Furthermore, the annotations on some modalities such as Late Gadolinium Enhancement (LGE) MRI, are often not available. We propose an end-to-end segmentation framework based on convolutional neural network (CNN) and adversarial learning. A dilated residual U-shape network is used as a segmentor to generate the prediction mask; meanwhile, a CNN is utilized as a discriminator model to judge the segmentation quality. To leverage the available annotations across modalities per patient, a new loss function named *weak domain-transfer loss* is introduced to the pipeline. The proposed model is evaluated on the public dataset released by the challenge organizer in MICCAI 2019, which consists of 45 sets of multi-sequence CMR images. We demonstrate that the proposed adversarial pipeline outperforms baseline deep-learning methods.

Keywords: Adversarial convolutional network · Multi-sequence cardiac segmentation

1 Introduction

Automatic segmentation of the tissues in cardiac magnetic resonance (CMR) images can provide the initial geometric information for surgical guidance [5]. However, manual delineation of heart structures in CMR scans is laborious and time-consuming. Late Gadolinium Enhancement (LGE) MR imaging is one of the most effective imaging modalities that can predict heart failure and sudden

M. Pop et al. (Eds.): STACOM 2019, LNCS 12009, pp. 317–325, 2020.
https://doi.org/10.1007/978-3-030-39074-7_34

death [16]. It enables doctors to visually exam the changes in the myocardium (myo) and confirm the existence of 'cardiomyopathy' and the degree of fibrosis.

There are three main challenges in CMR image segmentation: (1) the large anatomy variations between individuals, and the big diversity of imaging quality in the LGE. For example, due to microvascular occlusion, the contrast agent cannot reach certain areas of the heart, resulting in different enhancements; (2) the ambiguities of boundaries between different cardiac tissues, i.e., the intensity range of the myocardium in LGE CMR overlaps with the surrounding muscle tissue [4]; (3) Despite its clinical importance, LGE slice is much more difficult to annotate than both T2-weight and bSSFP, thus the annotations of LGE CMR are often not accurate or not available. In contrast, the annotations of T2-weight and bSSFP are easier and often available. To tackle these challenges, various methods have been proposed for whole-heart segmentation [8], ventricles segmentation [9,10], etc.

In recent years, deep convolutional neural networks (CNNs) [11] have achieved remarkable success in various computer vision tasks [12,13] as well as medical image segmentation [1]. Generative adversarial networks [2] as a recent machine learning technique, offers a promising avenue in image synthesis [6] as well as image segmentation [7].

We propose a framework to segment ventricles and myocardium from LGE CMR images based on CNNs and adversarial learning, when the annotations of LGE images are rather limited for training. Our contributions in this work are three folds: (1) we proposed an adversarial segmentation network containing two tailored modules: a segmentation model and a discriminator model, trained and optimized in an end-to-end fashion. The segmentation network generates the predicted masks, and the discriminator network aims to identify the segmentation mask and the ground-truth mask. The segmentation quality is improved in the min-max game. (2) since different modalities share structure information, we introduced a loss function named *weak domain-transfer loss* to leverage information from available modalities with rich annotations; (3) results show that the proposed method outperforms traditional CNN-based method.

2 Method

Our adversarial segmentation framework consists of a segmentation network and discrimination network. A dilated residual U-shape networks [14] is used as a segmentor (i.e. mask generator) G and a CNN classifier as a discriminator D. D is used to ensure that a generated mask being close to its ground truth mask conditioned on the same raw image; the segmentor and the discriminator are updated to improve the performance in an adversarial manner. We also leverage information from other common modalities using a *weak domain-transfer loss*. Figure 1 shows the framework of the proposed method.

Data and Preprocessing. The dataset is provided by the challenge organizers [3] and [4]. It consists of 45 patients, each with three MRI modalities

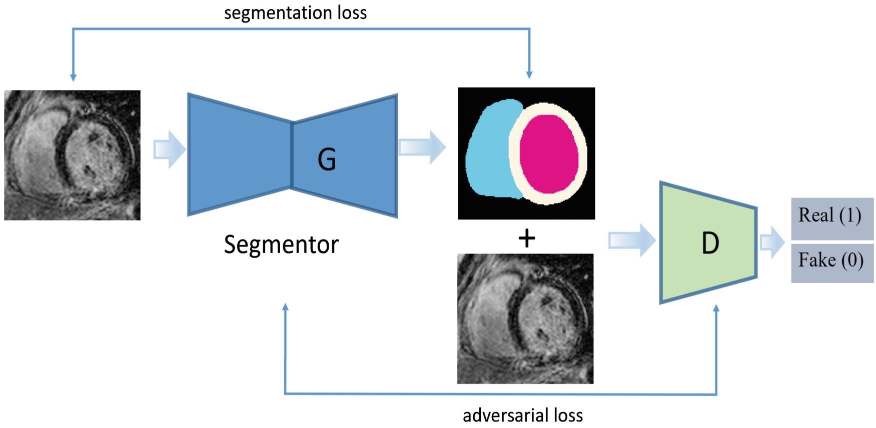

Fig. 1. Adversarial segmentation network architecture. It consists of a generator based on a dilated residual U-shape network and a CNN discriminator. The two networks are simultaneously optimized during the process of supervised learning and adversarial learning. Segmentation loss is a combination of individual-domain and domain-transfer loss, while the adversarial loss is a combination of the segmentation loss and the discriminator loss.

(LGE, T2-weight and bSSFP). It is noted that not all of the modalities come with the annotations of three heart regions (i.e., left ventricles, myocardium, and right ventricles). Annotations of all the three modalities are provided for patients 1–5; while patients 6–35 have manual annotations of T2-weight and bSSFP. Patients 36–45 have the raw MR scans of three modalities but without any annotations. When constructing the training set, only those MR scans with manual annotations are included. The test data contains the MR scans of LGE from patients 6 to 45, tasked to predict the masks of the three heart regions. Data augmentation is used for robust training. Three geometrical transformations (rotation, shear, zooming) are applied to all of the images and their corresponding masks. For each slice, we also crop a region with a fixed bounding box ($224 * 224$), enclosing all the annotated regions but at different locations to capture the shift invariance, resulting in 5 groups of cropped regions of interests. Before training the networks, the intensities of each 2D slice from three modalities are normalized using *z-scores* normalization to calibrate the range of intensities.

Weak Domain Transfer. Figure 2 shows some sample images with annotation masks of different modalities from the same patient. In Fig. 2, we can further observe from the annotations that the *bSSFP*, *T2* and *LGE* share some anatomical and structure information; For example, the right ventricle is always surrounded by myocardium, left ventricle is next to myocardium. The annotation masks of the corresponding slices from the three modalities have a certain level of overlap. Based on those observations, we hypothesize that the information from bSSFP and T2 can facilitate the segmentation of LGE. Hence we propose to use the annotation masks on bSSFP and T2 modalities as the *pseudo* masks for the unlabelled LGE modalities.

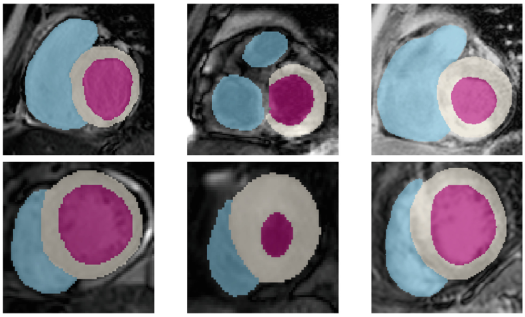

Fig. 2. From left to right are the images of the bSSFP, T2, LGE modalities from the same patient, with ground truth masks imposed (best viewed in color). (Color figure online)

The masks of bSSFP and T2 scans are transferred to LGE by using a normalized index which identifies the correspondence between axial slices from different modalities. These masks from bSSFP or T2 are directly used as the *pseudo* masks for the corresponding LGE. Specifically, for an axial slice i in bSSFP (or T2) with annotations, its corresponding slice index j in LGE is computed as below:

$$j = \lfloor i * \frac{n}{m} \rfloor \tag{1}$$

where $\lfloor \cdot \rfloor$ is the floor function. n denotes the number of axial slices of LGE; while m is the number of axial slice in bSSFP (or T2) respectively. Therefore the mask of slice i in bSSFP (or T2) is treated as the pseudo mask of the slice j in LGE.

Notably, those masks are *pseudo*, therefore, the domain-transfer loss should be set as a *weaker* one when combined with loss defined on ground truth annotations from expert. We will discuss this further in next section.

It is worth noting that our **transfer** is different from the *conventional transfer*, which often used a pre-trained model (e.g. on ImageNet), or a knowledge distillation framework of teacher-student learning [15]. Instead, our *transfer* is built as part of the whole model, specifically tailored for the cross-domain annotation-transfer problem.

Generator. Figure 3 shows the overview of the generator model, where a dilated residual U-shape network is tailored and used for the segmentation network. Residual blocks in downsampling and upsampling parts are connected through skip connections. In total the entire network consists of only 0.16 million trainable parameters.

In training a segmentation model, it is aware that cross-entropy loss focuses on individual pixels while Dice loss focuses on the overlap of regions. Thus, a combination of cross-entropy loss and Dice loss is chosen to optimize the network. Images and ground truth masks from the three sequences as well as the transferred masks mentioned above are used. Therefore, the training loss includes two parts: individual-domain loss and domain-transfer loss. Individual-domain loss, denoted as $\mathcal{L}_{ID}$, is the difference between the ground truth mask and prediction while *domain-transfer loss* denoted as $\mathcal{L}_{DT}$, is the difference between transferred masks (pseudo masks) and predicted ones.

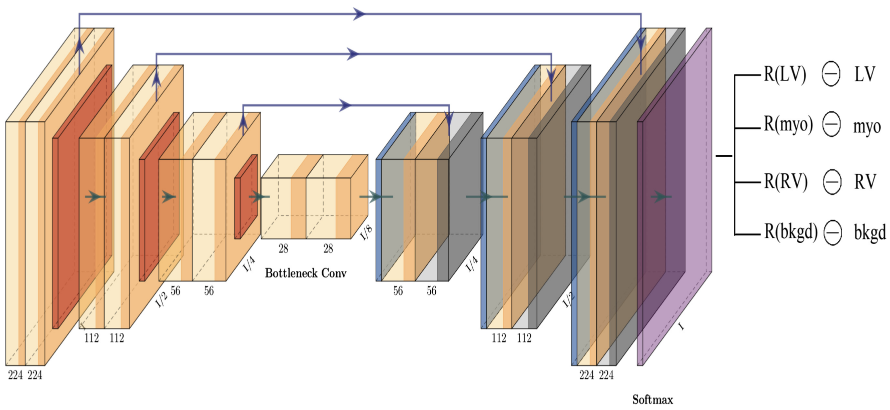

Fig. 3. Generator network architecture, composed of a downsampling tower and an upsampling tower.

Both of $\mathcal{L}_{ID}$ and $\mathcal{L}_{DT}$ consist of a linear combination of the multi-class cross-entropy loss $\mathcal{L}_{ce}$ and the Dice loss $\mathcal{L}_{Dice}$, formulated as:

$$\mathcal{L}_{ID} = \beta_1 \cdot \mathcal{L}_{ce} + (1 - \beta_1) \cdot \mathcal{L}_{Dice} \tag{2}$$

$$\mathcal{L}_{DT} = \beta_2 \cdot \mathcal{L}_{ce} + (1 - \beta_2) \cdot \mathcal{L}_{Dice} \tag{3}$$

The total loss function $\mathcal{L}_G$ is formulated as:

$$\mathcal{L}_G = \lambda \cdot \mathcal{L}_{ID} + (1 - \lambda) \cdot \mathcal{L}_{DT} \tag{4}$$

Notably, the domain-transfer loss leverages the information from bSSFP and T2 modalities. It is worth noting that λ in Eq. (4) is used to control the balance of the transfer; and it is set to 0.9, thus giving a much lower weight of the transfer loss 0.1 which is weak. In our experiments, β_1, β_2 are set to 0.9 after observing the segmentation performance on a validation set.

Mask Discriminator. We use a CNN as a discriminator to drive the generator to generate good-quality masks similar to the ground truth ones. The architecture contains several residual blocks with max-pooling layers. The raw images and the masks are spatially concatenated as a multi-channel input to the CNN. A (negative) binary cross-entropy loss $\mathcal{L}_D$ is used to train the model, defined as:

$$\mathcal{L}_{\mathcal{D}}(\mathcal{S}, \mathcal{T}, D, G) = \mathbb{E}_{(x,y)\sim\mathcal{S}}[\log D(y|x)] + \mathbb{E}_{(x',y')\sim\mathcal{T}}[\log D(y'|x')] + \\ \mathbb{E}_{(x,y)\sim\mathcal{S}}[\log(1 - D(G(x,y)|x))] + \\ \mathbb{E}_{(x',y')\sim\mathcal{T}}[\log(1 - D(G(x',y')|x'))] \quad (5)$$

where $\mathcal{S}$ is the set of training data x with ground truth masks y, and $\mathcal{T}$ is the set of LGE data x' without masks, but with pseudo masks y'.

Adversarial Training of Generator and Discriminator. The objective of the proposed system is to produce appropriate segmentation masks on the target class during the min-max game of the two networks. Firstly we perform a supervised training on G using the MR scans with ground truth masks, the objective of G is to generate a good mask to deceive the discriminator network D. The goal of D is to identify the generated masks from the real masks. We aim to improve the segmentation quality by merging the generated masks with the original images as condition labels and putting them into the discriminator for adversarial learning training. The adversarial model is designed to minimize the adversarial loss which will reverse optimize the generator loss.

Equation 6 represents the total loss in the adversarial model. G and D are simultaneously optimized.

$$\min_G \max_D \mathcal{L}_{adv} = \mathcal{L}_D + \mathcal{L}_G \quad (6)$$

Algorithm 1. Training procedure of the adversarial model

Input: training images X, training masks Y, iteration j and k, batch size n
Output: Models: Segmentation model G, Discriminator D

i = 0
while i ¡ j **do**
  update G by $\mathcal{L}_{IN}$
  i = i+1
end
while i ¡ k **do**
  update D by maximizing $\mathcal{L}_{adv}$ using a mini-batch while keep G fixed
  update G by minimizing $\mathcal{L}_{adv}$ using a mini-batch while keep D fixed.
  i = i+1
end
return G

3 Experiment

Implementation. The proposed method is implemented using *Keras* library. The codes are available at https://github.com/jingkunchen/MS-CMR_miccai_2019. α is set as 0.9 thus, giving the weight of 0.9 for the categorical cross-entropy loss and 0.1 for Dice loss. Learning rate is set to 2×10^{-4}, and the learning decay is 1×10^{-8}. We use a batch size of 16. For the transfer loss L_{DT}, we use the ground truth (whenever available) masks of T2-weight and bSSFP, as the *pseudo* ground truth masks for the corresponding LGE slices. The correspondence between the LGE slices and the T2-weight (or the bSSFP) slices are established based on the simple index normalization along the z-axis of the 3D MRI scans[1]. We use Adam optimizer.

3.1 Results

It is noted that only 5 patients have LGE annotations available, thus we performe a very preliminary experiment to test the proposed method. We held out patients 4 and 5 for testing and the rest for training. Results are reported in Table 1 in terms of Dice score and Hausdorff distance (LV, myo, RV). We further compare three methods: dilated residual U-shape networks with Dice loss (U+D), adversarial model with Dice coefficient loss (U+A+D), adversarial model with Dice coefficient loss and transfer loss (U+A+D+T). The U-shape networks are specifically designed to segment biomedical images and perform well in myocardial segmentation of bSSFP CMR images [3]. Here we use dilated residual U-shape networks with Dice loss (U+D) as our baseline for a comparison. It could be observed that adding adversarial training improves the segmentation performance on both the myocardium and right ventricles, but performs worse on left ventricles. The proposed method with transfer loss outperforms both of them with only one exception of the lower Dice score the right ventricle.

Table 1. Average Dice and Hausdorff distance on patients 4 and 5

Method	Dice	Hausdorff Dist.
	LV, myo, RV	LV, myo, RV
U-shape network (U+D)	70.5%, 50.0%, 70.0%	13.2, 12.0, 24.6
Adversarial model (U+A+D)	65.1%, 53.9%, **74.7**%	38.0, 16.1, 19.4
Adversarial transfer (U+A+D+T)	**76.0**%, **59.6**%, 71.7%	**10.2**, **12.1**, **12.9**

Results on Challenge Test Set. We submitted the results of the methods of (U+A+D) and (U+A+D+T) on the testing set containing patients 6 to 45 LGE. Tables 2 and 3 summarize the average and median values of the results returned by the organizers. It could be seen that overall the approach of (U+A+D+T) outperforms (U+A+D), which confirms that promise of the proposed method.

[1] In practice, we find this works well. Ideally, registration could be performed to find the correspondence, which will be investigated further.

Table 2. Average Dice, Jaccard, Surface Distance and Hausdorff distance on patients 6 to 45

Method	Dice	Jaccard	Surface Dist	Hausdorff Dist.
	LV, myo, RV	LV, myo, RV	LV, myo, RV	LV, myo, RV
U+A+D	76.6%, 42.0%, 69.5%	0.62, 0.27, 0.54	5.5, 4.7, 5.5	22.1, 42.0, 32.7
U+A+D+T	**82.4%**, **61.0%**, **71.0%**	**0.71**, **0.45**, **0.57**	**3.9**, **4.0**, **5.0**	23.7, **24.6**, **23.5**

Table 3. Median of Dice, Jaccard, Surface Distance and Hausdorff distance on patients 6 to 45

Method	Dice	Jaccard	Surface Dist	Hausdorff Dist.
	LV,myo,RV	LV,myo,RV	LV,myo,RV	LV,myo,RV
U+A+D	77.8%, 42.7%, 71.1%	0.63, 0.27, 0.55	5.3, 4.3, 5.0	18.5, 41.2, 28.5
U+A+D+T	**82.1%**, **60.8%**, **72.8%**	**0.70**, **0.44**, **0.57**	**3.8**, **3.9**, **4.6**	**15.4**, **19.6**, **22.8**

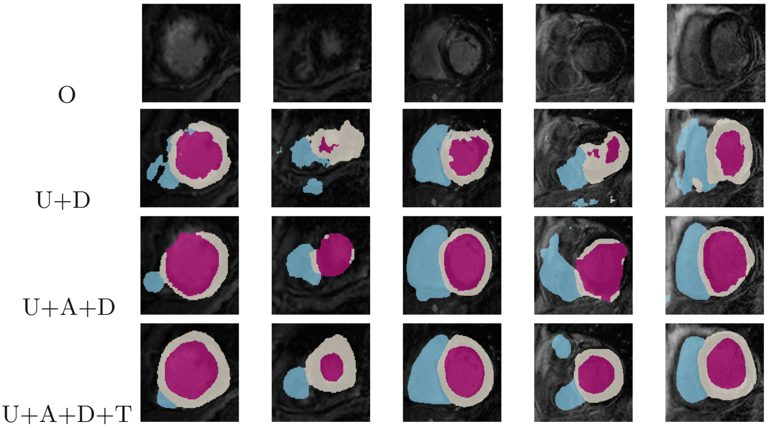

Fig. 4. The results of the segmentation. Rows from top to bottom: original images (O), dilated residual networks (U+D), adversarial model (U+A+D), adversarial model with Dice coefficient loss and transfer loss (U+A+D+T) (best viewed in color). (Color figure online)

Visualisation. Figure 4 shows some predicted masks of the LGE slices of four patients. It could be seen that adversarial learning improves the results of only using the dilated residual network, and the cross-modality transfer further refine the segmentation masks, especially for the left ventricles. Those observations are consistent with the results shown in Tables 1, 2 and 3.

4 Conclusions

We propose an automated method for heart segmentation based on multi-modality MRI images, which is trained in an adversarial manner. Specifically, our architecture consists of two modules, a multi-channel mask generator and a discriminator. In particular, we further introduce a domain-transfer loss function to leverage the information across different modalities for the same patients. Results show that such an idea is effective, and the overall framework performs better than the baseline methods.

References

1. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds.) Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015. Lecture Notes in Computer Science, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
2. Goodfellow, I., et al.: Generative adversarial nets. In: NIPS, pp. 2672–2680 (2014)
3. Zhuang, X.: Multivariate mixture model for myocardial segmentation combining multi-source images. In: TPAMI (2018)
4. Zhuang, X.: Multivariate mixture model for cardiac segmentation from multi-sequence MRI. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 581–588. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_67
5. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Med. Image Anal. **31**, 77–87 (2016)
6. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: CVPR, pp. 1125–1134 (2017)
7. Luc, P., Couprie, C., Chintala, S., Verbeek, J.: Semantic segmentation using adversarial networks. arXiv preprint arXiv:1611.08408 (2016)
8. Zhuang, X., et al.: Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge. arXiv preprint arXiv:1902.07880 (2019)
9. Petitjean, C., et al.: Right ventricle segmentation from cardiac MRI: a collation study. Med. Image Anal. **19**(1), 187–202 (2015)
10. Avendi, M.R., Kheradvar, A., Jafarkhani, H.: A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. Med. Image Anal. **30**, 108–119 (2016)
11. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436 (2015)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)
13. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: CVPR, pp. 3431–3440 (2015)
14. Li, H., Zhygallo, A., Menze, B.: Automatic brain structures segmentation using deep residual dilated U-Net. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) BrainLes 2018. LNCS, vol. 11383, pp. 385–393. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11723-8_39
15. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015)
16. Moon, J.C.: What is late gadolinium enhancement in hypertrophic cardiomyopathy? In: Revista Española de Cardiología (2007)