COMBINING MULTIPLE STYLE TRANSFER NETWORKS AND TRANSFER LEARNING FOR LGE-CMR SEGMENTATION

Bo Fang[1], *Junxin Chen*[1] , *Wei Wang*[2], *Yicong Zhou*[3]

[1]College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, China.
[2]School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen 518107, China.
[3]Department of Computer and Information Science, University of Macau, Macau, China.

ABSTRACT

This paper presents an algorithm for segmenting late gadolinium enhancement cardiac magnetic resonance (LGE-CMR) in the absence of labeled training data. The proposed method includes a data augmentation part and a segmentation network. Multiple style transfer networks are employed for data augmentation to increase the data diversity, and then the synthetic images are used for training an improved U-Net. Finally, the trained model is fine-tuned with a few LGE images and labels. Experiment results demonstrate the effectiveness and advantages of the proposed method.

***Index Terms*—** Cardiac segmentation, Style transfer networks, Transfer learning, Multiscale dilation fusion

1. INTRODUCTION

Cardiovascular diseases (CVDs) refer to diseases of the heart and blood vessels, and are a series of diseases involving the circulatory system. In clinical practice, physicians diagnose heart diseases by physiological parameters and the features of medical images [1]. Magnetic resonance imaging (MRI) is a type of tomographic imaging that uses the principles of nuclear magnetic resonance to map images of the structure inside an object. Cardiac magnetic resonance (CMR) is especially helpful to diagnose CVDs, since the information from multiple sequences (e.g., T1, T2, etc.) of MRI images is often different and complementary. Compared to other MRI sequences, the distinctive brightness contrast between diseased and healthy tissues in late gadolinium enhancement (LGE) CMR is a unique advantage [2]. Segmenting LGE-CMR can yield contours of the ventricles and myocardium, which provides a basis for diagnosing CVDs. However, LGE images with labels are very rare, it is valuable to develop novel segmentation methods in the absence of LGE labels.

In recent years, deep learning-based methods [3, 4] have made great progress for image segmentation. However, due to the large amount of time required for manual labeling, cross-modal segmentation becomes a difficult but practically valuable topic. Liu *et al.* [5] used the balanced steady state free precession (bSSFP) images as a priori knowledge for the cross-modal segmentation of LGE images. Meanwhile, with the development of generative adversarial networks (GAN), more researchers have used style transfer networks to train cross-modal segmentation. Chen *et al.* [6] proposed an algorithm combining a style transfer network and a cascaded U-Net. Qiu *et al.* [7] used GAN to expand the data, which improves the generalization ability and robustness of the model. However, these methods have complex structures, more parameters, and low visualization of training.

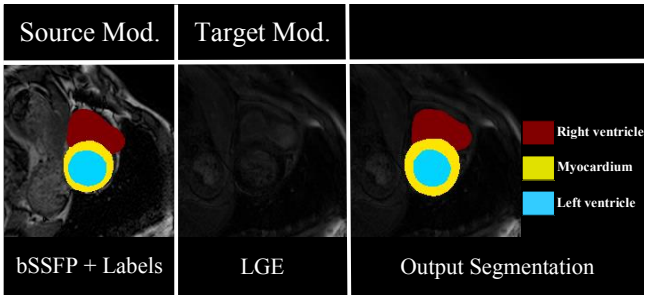

Fig. 1. Source domain image, target domain image, and segmentation result of the proposed method.

In this paper, we propose an algorithm for LGE-CMR segmentation. It can train the segmentation network from unpaired LGE-CMR and bSSFP-CMR samples without using LGE-CMR labels (as demonstrated in Fig. 1), and then segment the real LGE-CMR with satisfactory accuracy. The proposed method consists of two parts, the data augmentation procedure using multiple style transfer networks and a transfer learning based segmentation network. We use three different style transfer networks in the data augmentation module to generate synthetic LGE images from bSSFP-CMR images. The LGE images synthesized by different style transfer networks are better to reflect the contours of different tissues. Then, the synthetic LGE images together with corresponding bSSFP-CMR labels are used for training the segmentation network. In addition, we develop an improved U-Net

Corresponding Author: Junxin Chen (chenjx@bmie.neu.edu.cn, junxinchen@ieee.org). This work is funded by the National Natural Science Foundation of China (No. 62171114).

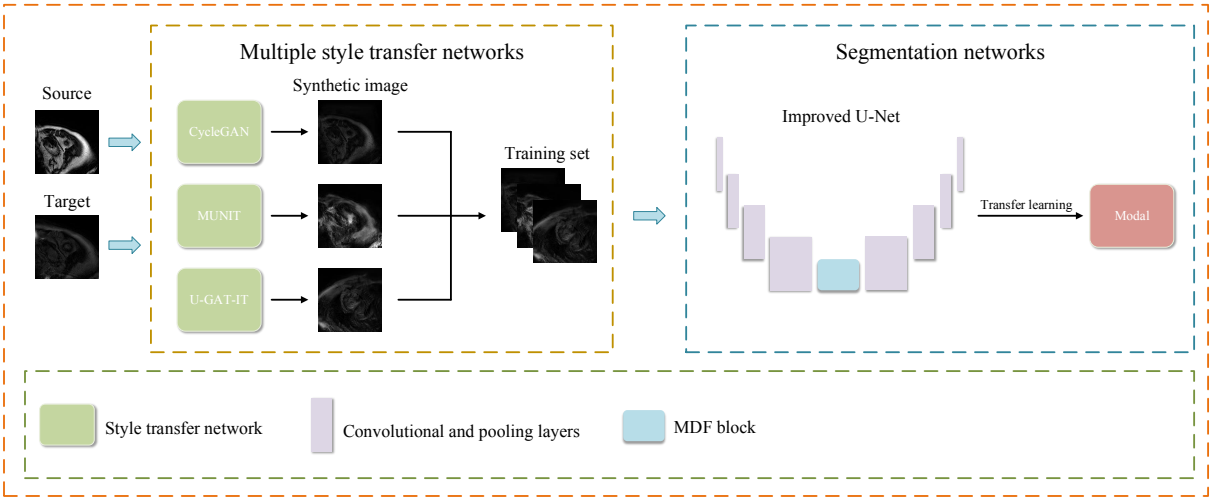

Fig. 2. Overview of the proposed algorithm.

as the segmentation network, which adds a multiscale dilation fusion (MDF) module to improve the network's ability to recognize multiscale features. Finally, inspired by transfer learning, we fine-tune the trained model using a small portion of the LGE images and labels.

Our main contributions can be summarized in three folds. 1) We propose to use multiple style transfer networks for data augmentation, so as to train the segmentation network without real LGE labels. 2) We fine-tune the trained model by transfer learning for further promoting the segmentation accuracy. 3) We train the data augmentation module and the segmentation network separately to increase the robustness and flexibility of the algorithm.

2. METHOD

2.1. Multiple Style Transfer Networks for Data Augmentation

Fig. 2 gives an overview of the proposed algorithm. In the data augmentation module, we choose three different style transfer networks to synthesize LGE images. The segmentation network is trained with synthetic images and corresponding bSSFP labels. We will analyze the effectiveness of using multiple style transfer networks in Section 4.

2.1.1. CycleGAN

The CycleGAN is proposed to obtain high-quality synthetic images under the training of unpaired images [8]. It consists of two generators and two discriminators. The discriminator determines whether the images synthesized by the generator is a synthetic image or real image. By continuous learning, the synthetic image is styled like the target domain image while preserving the structure of the source domain image.

2.1.2. MUNIT

Common style transfer network can only generate identical counts of synthetic images as the source domain images on a one-to-one basis. Therefore, the images synthesized by these networks lack diversity. To solve this problem, Huang *et al.* [9] proposed the multimodal unsupervised image-to-image translation (MUNIT) network.

The MUNIT network considers that an image can be decomposed into content code and style code, where the style code captures the specific properties of the domain. When a source image is input to the trained style transfer network, the network only needs to recombine the content code of the source domain and the style code of the target domain to generate multiple synthetic images. The decoder of the network uses adaptive instance normalization to reconstruct the image based on the input content code and style code.

2.1.3. U-GAT-IT

Kim *et al.* [10] proposed a new style transfer network named the unsupervised generate attentional networks with adaptive layerinstance normalization for image-to-image translation (U-GAT-IT). The U-GAT-IT combines new attention modules and learnable normalization features in an end-to-end manner.

The attention mechanism in U-GAT-IT uses a class activation map (CAM) under global and average pooling. The CAM is embedded in the generator and discriminator to guide the transformation to focus on the crucial areas and ignore the minor areas. Since the selection of the normalization function has a great impact on the transformation results for datasets with different textures and amounts of texture variation, U-GAT-IT proposed to use adaptive layer-instance normalization.

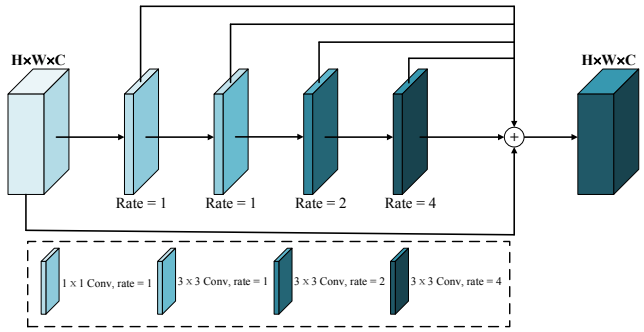

Fig. 3. The details of MDF block.

2.2. Segmentation Network

2.2.1. Improved network

In recent years, U-Net [11] has been widely used in various segmentation algorithms. Here, we improve the U-Net and use it for segmentation. Specifically, we add a MDF block between the encoder and decoder of U-Net, so as to improve the U-Net's recognition capability for multi-scale features. As shown in Fig. 3, the MDF block improves the ability to learn features by connecting multiple sub-branches of the feature map. The convolutional layers of the expansion are gradually increased at the rate of 1, 2, and 4. Finally, the model's robustness is improved by fusing multi-scale features.

On the other hand, to solve the overfitting problem, a spatial mean-variance normalization layer is added after each convolutional layer.

2.2.2. Overall loss function

To better improve the segmentation accuracy, we use weighted cross entropy and weighted Dice coefficients as the loss function of the segmentation network. The final loss function is the weighted sum of the two loss functions, where the coefficients are both 0.5.

3. EXPERIMENTAL RESULT

3.1. Dataset

The proposed method is validated on the Multi-sequence Cardiac MR Segmentation Challenge (MS-CMRSeg) 2019 dataset [2] which contains the CMR images of 45 patients. The dataset contains three sequences (LGE, bSSFP, T2) of CMR images, with approximately 11 bSSFP images and 15 LGE images for each patient. We use unpaired LGE and bSSFP images to train three style transfer networks, and the LGE images of patients #6 to #45 are used as the test set. Then, the LGE images' style is transferred to the bSSFP images to synthesize more LGE images. The synthetic images and their corresponding bSSFP-CMR labels are used as the training data of the segmentation network. Finally, the model trained by the synthetic images is fine-tuned with labeled LGE images of five patients.

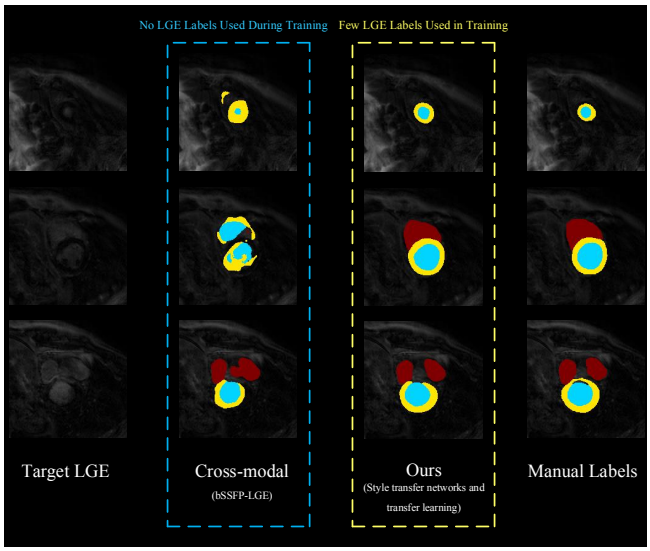

Fig. 4. Sample segmentation results by different strategies

3.2. Algorithm Implementation

For fair comparison, all experiments are trained in the same setting. The experiments are conducted on Tensorflow-1.13.1[1] platform, using the Microsoft Windows 10 operating system. The program is deployed on the workstation equipped with the GPU of Nvidia GeForce GTX 2080Ti with 11G memory. We use the Dice similarity coefficient as the metric to evaluate the segmentation performance. The Dice coefficient is an ensemble similarity index that measures the degree of overlap between the predicted profile P and the ground truth G, as defined in Eq. (1).

$$D(P,G) = 2\frac{P \bigcap G}{P+G}. \quad (1)$$

3.3. Result and Comparison

As listed in Table 1, the Dice values of our model for segmenting the left ventricle (LV), myocardium (MYO), and right ventricle (RV) are 0.91, 0.82 and 0.88, respectively. In addition, the segmentation results are visually plotted in Fig. 4, we can observe that method's segmentation results are very close to the manual labels. By visualizing the prediction results, the model predicts the contour of each part with little deviation from the manually labeled ground truth.

To better evaluate the performance of the proposed method, we choose the algorithms that have good results

[1]https://www.tensorflow.org/.

Table 1. Comparison between our method and some counterparts.

Dice score	LV	MYO	RV	AVG
Zhuang *et al.* [2](GMM+bSSFP)	0.836±0.071	0.635±0.120	-	-
Zhuang *et al.* [2](MvMM)	0.866±0.063	0.717±0.076	-	-
Tao *et al.* [12]	0.847±0.054	0.686±0.078	0.776±0.048	0.770±0.060
Liu *et al.* [5]	0.807±0.074	0.617±0.084	0.680±0.117	0.701±0.092
Unsupervised (bSSFP-LGE)	0.766±0.231	0.573±0.221	0.604±0.391	0.648±0.228
Ours	**0.912±0.127**	**0.820±0.166**	**0.877±0.216**	**0.870±0.119**

Table 2. The Dice values of models trained by different style transfer networks.

Dice	LV	MYO	RV	AVG
CycleGAN	0.853±0.199	0.688±0.197	0.798±0.291	0.780±0.167
MUNIT	0.866±0.145	0.729±0.163	0.810±0.248	0.802±0.125
U-GAT-IT	0.829±0.145	0.622±0.186	0.791±0.253	0.747±0.132
All (without transfer learning)	0.894±0.151	0.762±0.169	0.857±0.230	0.838±0.128
All (with transfer learning)	**0.912±0.127**	**0.820±0.166**	**0.877±0.216**	**0.870±0.119**

in the same test set for further comparison. Zhuang *et al.* proposed two machine learning methods in [2], and Liu *et al.* [5] used bSSFP-CMR images as a priori knowledge for LV localization and then segmented LGE images. Furthermore, Tao *et al.* [12] improved the segmentation accuracy by proposing a novel shape transfer GAN network to preserve the anatomical structure of bSSFP-CMR. Table 1 summarizes the performance records. As indicated, the proposed method has advantages in terms of accuracy, complexity, and has great potential for clinical application.

In addition, a unsupervised segmentation model which directly adopts the bSSFP images for training the segmentation network and then uses the trained model for segmenting LGE-CMR images is also evaluated. The segmentation performance of this unsupervised model is listed in Table 1. As indicated, the performance records obtained by our method are dramatically higher than the unsupervised model, though the prior knowledge (the bSSFP images) of this model is identical to that of mine. The innovations of our method in terms of using style transfer network for data augmentation and using transfer learning for performance promotion can be well demonstrated.

4. ABLATION STUDIES

To demonstrate the effectiveness of using multiple style transfer networks for data augmentation, we have conducted three comparison experiments. As listed in Table 2, the results are different when using different style transfer networks separately for data augmentation. Meanwhile, our method achieves the best results at all parts, suggesting that combining the three networks is effective.

In addition, we chose peak signal to noise ratio (PSNR), learned perceptual image patch similarity (LPIPS), structural similarity (SSIM), visual information fidelity (VIF) to measure the quality of the synthesized images of different style transfer networks. As given in Table 3, each network has its advantages. Therefore, using multiple style transfer networks for data augmentation can improve the model's robustness. Meanwhile, the data diversity is improved, which makes the pixel distribution of the synthetic images and LGE-CMR images more similar.

Table 3. The quality of the synthetic images generated by different style transfer networks.

Metrics	PSNR	LPIPS	SSIM	VIF
CycleGAN	21.17	**0.21**	0.45	0.83
MUNIT	16.66	0.27	0.35	**0.90**
U-GAT-IT	**22.02**	0.23	**0.46**	0.79

5. CONCLUSION

In this paper, we propose a LGE segmentation method using multiple style transfer networks and transfer learning. For data augmentation, we improve the data diversity by combining synthetic images generated by multiple style transfer networks. In the segmentation network, we improve the model robustness and segmentation accuracy by adding the MDF module to U-Net and refining the loss function. Finally, a few LGE images and labels are used to fine-tune the model. The experimental results show that the proposed method achieves high segmentation accuracy. In addition, the advantages and practical value of the proposed method have been highlighted by comparing it with some state-of-the-art algorithms.

6. REFERENCES

[1] Junxin Chen, Shuang Sun, Li-bo Zhang, Benqiang Yang, and Wei Wang, "Compressed sensing framework for heart sound acquisition in internet of medical things," *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2021.

[2] Xiahai Zhuang, "Multivariate mixture model for myocardial segmentation combining multi-source images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 12, pp. 2933–2946, 2018.

[3] Wei Chen, Qiuli Wang, Sheng Huang, Xiaohong Zhang, Yucong Li, and Chen Liu, "Dfdm: A deep feature decoupling module for lung nodule segmentation," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 1120–1124.

[4] Changlu Guo, Márton Szemenyei, Yangtao Hu, Wenle Wang, Wei Zhou, and Yugen Yi, "Channel attention residual u-net for retinal vessel segmentation," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 1185–1189.

[5] Tao Liu, Yun Tian, Shifeng Zhao, XiaoYing Huang, Yang Xu, Gaoyuan Jiang, and Qingjun Wang, "Pseudo-3d network for multi-sequence cardiac MR segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2019, pp. 237–245.

[6] Chen Chen, Cheng Ouyang, Giacomo Tarroni, Jo Schlemper, Huaqi Qiu, Wenjia Bai, and Daniel Rueckert, "Unsupervised multi-modal style transfer for cardiac mr segmentation," 2019.

[7] Chen Chen, Chen Qin, Huaqi Qiu, Cheng Ouyang, Shuo Wang, Liang Chen, Giacomo Tarroni, Wenjia Bai, and Daniel Rueckert, "Realistic adversarial data augmentation for mr image segmentation," 2020.

[8] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2242–2251.

[9] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz, "Multimodal unsupervised image-to-image translation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 172–189.

[10] Junho Kim, Minjae Kim, Hyeonwoo Kang, and Kwanghee Lee, "U-GAT-IT: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation," *CoRR*, vol. abs/1907.10830, 2019.

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[12] Xumin Tao, Hongrong Wei, Wufeng Xue, and Dong Ni, "Segmentation of multimodal myocardial images using shape-transfer GAN," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2019, pp. 271–279.