Improved unsupervised domain adaptation network based on category attention

Longhao Fan¹, Shouwei Gao¹, Fan Zhu², Zhenzhong Zhu³, Chaozheng Zhou⁴ and Yaxin Peng*²

¹ School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, 200444, China
² College of Science, Shanghai University, Shanghai, 200444, China
³ Department of Orthopedic Surgery, Shanghai Jiao Tong University Affiliated Shanghai Sixth People’s Hospital, Shanghai, 200233, China
⁴ Shanghai Electric Central Research Institute, Shanghai, 200336, China
*Corresponding author’s e-mail: yaxin.peng@shu.edu.cn

Abstract. Domain adaptation method can significantly reduce the distribution difference between images in variant domains, which plays an important role in unsupervised medical image segmentation. In this paper, an improved unsupervised domain adaptation framework is proposed based on category attention mechanism. The framework considers both image-level and feature-level alignment, and realizes semantic segmentation in different domains through a shared encoder-decoder. A category attention based classifier is proposed to compute the category attention feature and refine the semantic segmentation prediction. Sufficient experiments on MMWHS-2017 dataset indicate that the proposed method achieves the best segmentation performance among all comparison algorithms.

1. Introduction
Medical image segmentation is one of the basic tasks of intelligent medicine. It aims to extract regions of interest such as organs and tissues in medical images to facilitate subsequent processing. In recent years, due to the popularity of deep learning, many scholars have applied deep convolutional networks in medical image segmentation and achieved good results. However, medical images of different modalities are quite disparate. It is difficult to use a unified deep learning network for images of multiple modalities. Furthermore, medical image segmentation is different from segmentation in road scenes. The structure of medical image is complex, people can easily distinguish vehicles and pedestrians in the street but most people are not able to distinguish substructures in MRI images. Labelling of medical image often requires professional medical practitioners to complete. This undoubtedly increases the cost of medical image labelling and also increases the difficulty of implementing deep learning networks. Therefore, it is of great significance to study unsupervised medical image segmentation methods.

Domain adaptation is commonly used to solve the problem of unsupervised image segmentation. The main idea is to reduce the domain shift by aligning the feature distribution of the image from different domains. [1-5] utilize image transformation model to align the appearance of images in
different domains. [6-9] align the high-dimensional features of images, and classify them through a shared classifier. [10-13] consider both image alignment and feature alignment.

How to efficiently exploit context is essential for pixel-level segmentation. Chen et al. [14] proposed the Atrous Spatial Pyramid Pooling (ASPP) to aggregate spatial regularly sampled pixels at different dilated rates around a pixel as its context. PSPNet [15] obtained context of different scales through the pyramid pooling module, and merged them into a global context. These of methods [15-18] focus on exploiting different spatial strategies to capture richer contextual information. However, exploiting the class-level context is also critical for semantic segmentation task. ACFNet [19] exploited context from a categorical perspective, and proposed a class center which uses all pixels of the same category to calculate a class-level representation.

Based on the above research, we propose an unsupervised domain adaptation network. On the basis of SIFA [12] framework, a two-stage classifier with category attention module is adopted to advance the segmentation performance.

2. Related Works

2.1. ACFNet

Different from ordinary semantic segmentation methods, which use context from a spatial perspective, the Attentional Class Feature Network (ACFNet) [19] introduces the concept of category center, which extracts global context from a classification perspective, and combines the attention mechanism to propose the Attentional Class Feature (ACF) module. The overview of ACFNet is shown in figure 1. The network adopts a coarse-refine structure. First, the image feature is extracted through the base network and the coarse segmentation result of the image is obtained. Then the category attention feature is extracted through the ACF module. The original image feature map is combined with the category attention to refine the segmentation result.

![ACFNet Architecture](image)

The ACF module is composed of a class center block (CCB) and a class attention block (CAB). For a certain category, according to the probability that each pixel in the coarse segmentation map belongs to the category, the class center block weights the pixel values on the extracted high-dimensional feature maps to obtain the class center. Through the class attention block, the class center and the original feature map are point-wise multiplied to get the category attention. Finally, the category attention feature is concatenated with the original feature, and a fine classifier is used to refine the classification prediction. ACFNet utilizes ResNet-101 as the feature extraction network, and achieves the SOTA segmentation accuracy on the Cityscapes dataset.

2.2. SIFA

Synergistic Image and Feature Alignment (SIFA) [12] is an unsupervised domain adaptation framework for image segmentation which proposes to perform collaborative alignment in both image space and feature space. The model extracts image features through an encoder shared by the source and target domains. The feature map is reconstructed into image in a specific domain through the
decoder. Multiple pairs of generators and discriminators are used to establish a cyclic consistency loss which aims to reduce the appearance difference of two domains in generated image space. Besides, a point-wise classifier is employed to obtain the semantic segmentation prediction of two domains. An adversarial loss is applied to align the images of different domains in the semantic prediction space.

3. Methods

In this paper, the category attention mechanism is adopted to improve the unsupervised domain adaptation framework SIFA. The overview of our model is shown in figure 2. A new two-stage classifier $C_{ca}$ is proposed based on the category attention mechanism. $C_{ca}$ is composed of a category attention module and two classifiers. The category attention features are computed through the segmentation results of the coarse classifier, and the image is re-classified with an attention-fused feature map through the fine classifier, which improves the image segmentation performance of our model. The introduction of the proposed method is divided into two sections, i.e., image space alignment and semantic prediction space alignment.

3.1. Image space alignment

There is huge distribution gap between cross-modality medical images. In this section, we use generative adversarial network to eliminate the appearance difference between images of two domains.

A generative adversarial network $\{G^t, D^t\}$ is used to transform the image from source domain to target domain. The objective function maintains the structure of the generated image, which is:

\[
L^t_{adv}(G^t, D^t) = \mathbb{E}_{x^t \sim x^t}[\log D^t(x^t)] + \mathbb{E}_{x^s \sim x^s}[\log(1 - D^t(x^{s \rightarrow t}))]
\]  

(1)

where $x^t$ is the target domain image, $x^{s \rightarrow t}$ is the target domain image generated from source domain.

A reverse generative adversarial network $\{E, U, D^s\}$ is introduced to convert the target domain image back to the source domain. $E$ is an encoder, $U$ is a decoder. $E \circ U$ forms a generator, $D^s$ is the corresponding discriminator. $\{E, U, D^s\}$ keeps the distribution of the generated image $x^{t \rightarrow s}$ consistent with that of $x^s$. The objective function of the reverse generative adversarial network is:

\[
L^s_{adv}(E, U, D^s) = \mathbb{E}_{x^s \sim x^s}[\log D^s(x^s)] + \mathbb{E}_{x^t \sim x^t}[\log(1 - D^s(x^{t \rightarrow s}))]
\]  

(2)

Inspired by CycleGAN [20], a cycle-consistency constraint is established for the images generated by the two sets of generators, so that the generated images can maintain their original structure. The cycle-consistency constraint is as follows:

\[
L^s_{cyc}(G^t, E, U) = \mathbb{E}_{x^t \sim x^t}[\|x^{s \rightarrow t \rightarrow s} - x^s\|_1] + \mathbb{E}_{x^s \sim x^s}[\|x^{t \rightarrow s \rightarrow t} - x^t\|_1]
\]  

(3)

where $x^{t \rightarrow s \rightarrow t}$ means the image generated from $x^{t \rightarrow s}$ through $G^t$. The cycle-consistency constraint ensures that the output of the generator $G^t$ and $E \circ U$ have the same structure as the input image.
The feature alignment in generated image space is shown in the blue dashed box in figure 2. For the images generated through \( E \circ U \), the discriminator \( D^s \) is adopted to determine whether the input image is generated from \( x^{s-t} \) or \( x^t \). If the discriminator can classify the input image correctly, it means that the output space still contains the information from the origin domain. The objective function of feature alignment in generated image space is:

\[
L_{adv}^s(E, U, D^s) = \mathbb{E}_{x^{s-t}, x^s \sim \mathcal{X}} [\log D^s(x^{s-t} \sim x^s)] + \mathbb{E}_{x^t \sim \mathcal{X}} [\log (1 - D^s(x^t))] \quad (4)
\]

3.2. Semantic prediction space alignment

3.2.1 Category attention based segmentation network. Based on ACFNet, we introduce a two-stage classifier \( C_{ca} \) after the encoder \( E \) to construct the image segmentation framework \( E \circ C_{ca} \). The two-stage classifier \( C_{ca} \) is composed of a coarse classifier \( C \), a category attention module and a fine classifier \( C_{fine} \), as shown in figure 3.

![Figure 3. Image segmentation network with ACF based classifier.](image)

The category attention module first extracts the center of each category, i.e., the average feature vector of each category. The equation for the center of i-th category is:

\[
F_{i}^{cate} = \frac{\sum_{j=1}^{H\times W} \mathbb{1}[y_j^s = i] \cdot F_{j}^{s-t}}{\sum_{j=1}^{H\times W} \mathbb{1}[y_j^s = i]} \quad (5)
\]

where \( y_j^s \) represents the ground-truth label of pixel j. \( \mathbb{1}[y_j^s = i] \) is a binary indicator function, indicating whether pixel j belongs to the i-th category. If pixel j belongs to the i-th category, \( \mathbb{1}[y_j^s = i] \) is equal to 1, otherwise 0. \( F_{j}^{s-t} \) represents the feature vector of pixel j.

Since the test data have no label, we use coarse segmentation probability map \( \hat{y}_{s-t}^s \) predicted by classifier \( C \) and normalized by softmax function to replace the ground-truth label \( y^s \). Let the softmax function be \( s(\cdot) \), the equation for the i-th category center can be written as:

\[
F_{i}^{cate} = \frac{\sum_{j=1}^{H\times W} s(\hat{y}_{s-t}^s)_{i,j} \cdot F_{j}^{s-t}}{\sum_{j=1}^{H\times W} s(\hat{y}_{s-t}^s)_{i,j}} \quad (6)
\]

Having the center of each category, for pixel j, we use its coarse segmentation probability map as the attention map, and combine the various category centers to obtain the category attention. The category attention feature \( F_{j}^{a} \) for pixel j is:

\[
F_{j}^{a} = \sum_{i=1}^{K} s(\hat{y}_{s-t}^s)_{i,j} \cdot F_{i}^{cate} \quad (7)
\]

The structure of category attention module is shown in figure 4. In order to reduce the network calculation cost, we use \( 1 \times 1 \) convolutional layer decreasing the channel number of the feature map \( F_{j}^{s-t} \) to \( C' \), and get a new feature map \( F^{s-t} \). Then, multiply the coarse segmentation probability map
s(\hat{y}^{s\rightarrow t}) with the new feature map to obtain the category center \( F^{cate} \in \mathbb{R}^{K \times C'} \). Finally, multiply the category center \( F^{cate} \) by the coarse segmentation probability map \( s(\hat{y}^{s\rightarrow t}) \) and acquire the category attention feature \( F^a \) through a 1 \times 1 convolutional layer.

\[
\begin{align*}
\text{Figure 4. The architecture of category attention module.}
\end{align*}
\]

Input the source domain image of the target domain modality \( x^{s\rightarrow t} \) and the target domain image \( x^t \) into the segmentation network to obtain the corresponding semantic segmentation predictions \( \hat{y}^{s\rightarrow t} \) and \( \hat{y}^t \) through classifier \( C \). Then the category attention of the source domain prediction is extracted by category attention module. The attention feature is concatenated with the original high-dimensional features and feed to the fine classifier \( C_{fine} \) to get the refined segmentation prediction \( \hat{y}_{fine} \). The objective function of the segmentation framework \( E \circ C \) is:

\[
L_{seg}(E, C_{ca}) = H(y^s, \hat{y}_{fine}^{s\rightarrow t}) + \alpha_1 \cdot \text{Dice}(y^s, \hat{y}_{fine}^{s\rightarrow t}) \tag{8}
\]

where \( H \) represent the cross-entropy function, Dice means dice similarity coefficient; \( \alpha_1 \) is a hyperparameter to balance cross-entropy and dice coefficient. The feature alignment in semantic prediction space is shown in the orange dashed box in figure 2. Similar to the generated image space, the model introduces a discriminator \( D^p \) in the semantic prediction space, and inputs the semantic prediction results of the two domains \( \hat{y}_{fine}^{s\rightarrow t} \) and \( \hat{y}^t \) into discriminator \( D^p \) to judge the source of them. The objective function of feature alignment in semantic prediction space is:

\[
L_{adv}^p(E, C_{ca}, D^p) = \mathbb{E}_{x^{s\rightarrow t}, x^{s\rightarrow t}}[\log D^p(\hat{y}_{fine}^{s\rightarrow t})] + \mathbb{E}_{x^t, x^t}[^{1}(1 - D^p(\hat{y}^t))] \tag{9}
\]

3.2.2 Auxiliary segmentation. Considering the distance between the output space and the shallow features, the model introduces an auxiliary classifier \( C_a \) in the penultimate layer of the encoder. A discriminator \( D^{p_a} \) is introduced as well, so that the objective function of the adversarial learning can effectively align the shallow features during back propagation. The configuration of this part is the same as the segmentation network in section 3.2.1. The corresponding auxiliary segmentation objective function and the auxiliary semantic prediction space alignment objective function are:

\[
L_{seg}(E, C_a) = H(y^s, \hat{y}_a^{s\rightarrow t}) + \alpha_2 \cdot \text{Dice}(y^s, \hat{y}_a^{s\rightarrow t}) \tag{10}
\]

\[
L_{adv}^{p_a}(E, C_a, D^{p_a}) = \mathbb{E}_{x^{s\rightarrow t}, x^{s\rightarrow t}}[\log D^{p_a}(\hat{y}_a^{s\rightarrow t})] + \mathbb{E}_{x^t, x^t}[\log(1 - D^{p_a}(\hat{y}_a^t))] \tag{11}
\]

where \( \hat{y}_a^{s\rightarrow t} \) and \( \hat{y}_a^t \) are the auxiliary semantic prediction results of \( x^{s\rightarrow t} \) and \( x^t \). \( E(\cdot) \) in equation (10) and (11) represent the output of the penultimate layer of the encoder. \( \alpha_2 \) is the hyperparameter.

In conclusion, the overall objective function of the whole framework is:

\[
L(G^t, D^t, E, U, D^s, C_{ca}, C_a, D^{p_a}, D^p) = \lambda_{adv}^l L_{adv}^l(G^t, D^t) + \lambda_{adv}^s L_{adv}^s(E, U, D^s) + \lambda_{cyc} L_{cyc}(G^t, E, U) + \lambda_{seg} L_{seg}(E, C_{ca}) + \lambda_{seg}^p L_{seg}^p(E, C_a, D^{p_a}) + \lambda_{adv}^p L_{adv}^p(E, C_a, D^{p_a}) \tag{12}
\]
where function \( L_{adv}^I, L_{adv}^s \) and \( L_{cyc} \) correspond to the image level alignment; Function \( L_{adv}^P, L_{adv}^{P_a} \) and \( L_{adv}^{P_a} \) correspond to the feature level alignment. \( \lambda_{adv}^I, \lambda_{adv}^s, \lambda_{cyc}, \lambda_{seg}^1, \lambda_{seg}^2, \lambda_{adv}^P, \lambda_{adv}^{P_a} \) and \( \lambda_{adv}^{P_a} \) are hyperparameters used to balance various objective functions.

3.3. Implementation detail
As shown in figure 2, the modules are divided into five types: encoder, decoder, classifier, generator and discriminator. The configurations of each type of modules are the same as SIFA\cite{12}.

The whole framework is implemented using TensorFlow 1.12.0 software package. All models are iteratively trained 20k times on Tesla P400 GPU. The batch size is set to 8. Adam optimizer is adopted to optimize the network with a learning rate of 2e-4.

4. Experiments

4.1. Dataset
We evaluate the proposed algorithm on the dataset of the 2017 Multi-Modality Whole Heart Segmentation Challenge (MMWHS-2017) \cite{21}. The MMWHS-2017 dataset contains 20 sets of cardiac MRI data and 20 sets of CT data from different patients. In this section, four important cardiac substructures, i.e., Ascending Aorta (AA), Left Atrium blood Cavity (LAC), Left Ventricle blood Cavity (LVC) and MYOcardium of the left ventricle (MYO), are selected for the domain adaptation from MRI to CT. MRI data is regarded as source domain data and CT data is regarded as target domain data. The training of the model is carried out on two-dimensional coronal slices. During training, the source domain data is labelled, and the target domain data is unlabelled.

For MRI data, 16 out of 20 sets are used as the training set and 4 sets are used as validation set. For CT data, 14 sets are selected for training, 2 sets for validation and 4 sets for testing. In training, the pre-process provided by SIFA is adopted, including clip, resampling, normalization, augmentation, etc.

4.2. Evaluation Metrics
Two widely used evaluation metrics in segmentation, Dice similarity coefficient (Dice) and average symmetric surface distance (ASSD), are introduced to evaluate the segmentation performance of models. The higher Dice and lower ASSD demonstrates the better segmentation performance. The expressions of Dice and ASSD are as follows:

\[
\text{Dice} = \frac{2(A \cap B)}{A + B}
\]  \hspace{1cm} (13)

\[
\text{ASSD} = \frac{\sum_{a \in S(A)} \min_{b \in S(B)} \|a - b\| + \sum_{b \in S(B)} \min_{a \in S(A)} \|b - a\|}{S(A) + S(B)}
\]  \hspace{1cm} (14)

where A and B represent the 3-dimensional prediction result and the ground-truth respectively. \( S(\cdot) \) means the set of voxels in 3-dimensional surface.

4.3. Main results
In order to comprehensively evaluate the impact of domain shift on the segmentation performance of our model, we adopt the basic segmentation network \((E \circ C\) in figure 3) training in source domain to test target domain images as the without adaptation control group. Meanwhile, the basic network also carries on a fully-supervised training for comparison. The proposed method is compared with the above model and other advanced algorithms to verify the effectiveness of our method in unsupervised segmentation tasks.

Table 1 compares the image segmentation performance of the proposed model and other comparison models on the CT test dataset. Without domain adaptation, the dice coefficient of the basic segmentation network is only 26.7%, and the ASSD reaches up to 24.5. After adopting the SIFA
framework, the dice coefficient is increased to 80.0%, the ASSD is reduced to 6.0. This result confirms the effectiveness of SIFA framework. In this paper, we employ the category attention mechanism to improve the classifier in SIFA. The improved model achieves an average dice coefficient of 82.9%, an increase of 2.9% on the basis of SIFA; The average ASSD fell to 3.2, a decrease of 2.8. As shown in table 1, our model has better results in all categories, reflecting the effectiveness of category attention mechanism. The result indicates that the proposed unsupervised domain adaptation model achieves a comparable performance to that of the fully-supervised training model, and gets the best segmentation results in comparison with other algorithms.

| Methods        | Dice | ASSD |
|----------------|------|------|
|                | AA   | LAC  | LVC  | MYO  | Average | AA   | LAC  | LVC  | MYO  | Average |
| W/o adaptation| 29.4 | 53.5 | 4.7  | 19.0 | 26.7     | 36.2 | 11.8 | 29.7 | 20.5 | 24.5    |
| Supervised training | 81.7 | 90.1 | 92.2 | 87.0 | 87.7     | 2.7  | 2.8  | 1.6  | 1.9  | 2.2     |
| PnP-AdaNet[11] | 74.0 | 68.9 | 61.9 | 50.8 | 63.9     | 12.8 | 6.3  | 17.4 | 14.7 | 12.8    |
| SynSeg-Net[22] | 71.6 | 69.0 | 51.6 | 40.8 | 58.2     | 11.7 | 7.8  | 7.0  | 9.2  | 8.9     |
| AdaOutput[23]  | 65.2 | 76.6 | 54.4 | 43.6 | 59.9     | 17.9 | 5.5  | 19.7 | 17.9 | 11.5    |
| CycleGAN[20]   | 73.8 | 75.7 | 52.3 | 28.7 | 57.6     | 11.5 | 13.6 | 9.2  | 8.8  | 10.8    |
| CyCADA[10]     | 72.9 | 77.0 | 62.4 | 45.3 | 64.4     | 9.6  | 8.0  | 9.6  | 10.5 | 9.4     |
| SIFA[12]       | 87.6 | 86.7 | 80.0 | 65.8 | 80.0     | 5.6  | 4.1  | 7.0  | 7.3  | 6.0     |
| ours           | 89.3 | 87.7 | 83.5 | 71.0 | 82.9     | 3.0  | 3.2  | 3.4  | 3.2  | 3.2     |

Figure 5. Part of visualization results.

Figure 5 presents part of the visualization results of our method. The segmentation results of the model without adaptation are messy and disorganized. With adopting the unsupervised domain adaptation framework SIFA, the segmentation results of each organization become clear, but still have some mistakes. However, the results of our model have almost no difference from the fully-supervised
results as well as the ground-truth. This also proves that our method can significantly reduce the performance gap between unsupervised domain adaptation networks and fully-supervised networks. The proposed domain adaptation method achieves similar results with supervised learning on unlabelled datasets.

4.4. Ablation studies
For evaluating the impact of the important components of the unsupervised domain adaptation segmentation model proposed in this paper, we also conduct a series of ablation experiments on MMWHS dataset. These experiments mainly analyse the influence of CycleGAN and category attention module on segmentation performance of our model. The experimental results are shown in table 2, where CG stands for CycleGAN, CA represents category attention module.

As shown in table 2, without both components, the framework still achieves a dice coefficient of 78.8% and a ASSD of 5.9. After introducing CycleGAN to the basic framework, the dice coefficient increased by 1.5%, but the ASSD grows slightly. With category attention module, the base network gets 77.0% Dice and 4.0 ASSD. With the employment of both CycleGAN and category attention based classifier, the dice coefficient rise to 82.9%, and the ASSD is 3.2. The result of ablation experiments indicates that the proposed method can significantly improve the segmentation performance of the model.

Table 2. The impact of each component on segmentation result.

| CG | CA | Dice | ASSD |
|----|----|------|------|
| √  |  | 78.8 | 5.9  |
| √  | √ | 80.3 | 6.7  |
| √  | √ | 77.0 | 4.0  |
| √  | √ | 82.9 | 3.2  |

5. Conclusion
This paper presents an improved unsupervised domain adaptation method. Based on the SIFA framework, we introduce a category attention based fine classifier to refine the segmentation result. The proposed method and other advanced algorithms are compared on MMWHS-2017 dataset. Experimental results indicate that our method achieves the best results in unsupervised medical image segmentation.

Acknowledgments
This work was supported by Shanghai Science and Technology Innovation Action Plan 21S31901000. Thanks are due to Li SJ and Gao BQ for assistance with the translation.

References
[1] Zhang Y, Miao S, Mansi T, et al. (2018) Task driven generative modeling for unsupervised domain adaptation: Application to x-ray image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. 599-607.
[2] Zhao S, Li B, Yue X, et al. (2019) Multi-source domain adaptation for semantic segmentation. arXiv preprint arXiv:1910.12181.
[3] Yang Y, Soatto S. (2020) Fda: Fourier domain adaptation for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 4085-4095.
[4] Russo P, Carlucci F M, Tommasi T, et al. (2018) From source to target and back: symmetric bi-directional adaptive gan. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 8099-8108.
[5] Chen C, Dou Q, Chen H, et al. (2018) Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest X-ray segmentation. In: International Workshop on Machine Learning in Medical Imaging. 143-151.

[6] Ganin Y, Ustinova E, Ajakan H, et al. (2016) Domain-adversarial training of neural networks. The Journal of Machine Learning Research, 17: 2096-2030.

[7] Kamnitsas K, Baumgartner C, Ledig C, et al. (2017) Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: International Conference on Information Processing in Medical Imaging. 597-609.

[8] Dou Q, Ouyang C, Chen C, et al. (2018) Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss. arXiv preprint arXiv:1804.10916.

[9] Tzeng E, Hoffman J, Saenko K, et al. (2017) Adversarial discriminative domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 7167-7176.

[10] Hoffman J, Tzeng E, Park T, et al. (2018) Cycada: Cycle-consistent adversarial domain adaptation. In: International Conference on Machine Learning. 1989-1998.

[11] Dou Q, Ouyang C, Chen C, et al. (2019) PnP-AdaNet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation. IEEE Access, 7: 99065-99076.

[12] Chen C, Dou Q, Chen H, et al. (2020) Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. IEEE Transactions on Medical Imaging, 39: 2494-2505.

[13] Chen C, Dou Q, Chen H, et al. (2019) Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. 865-872.

[14] Chen L-C, Papandreou G, Schroff F, et al. (2017) Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.

[15] Zhao H, Shi J, Qi X, et al. (2017) Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2881-2890.

[16] Chen L-C, Zhu Y, Papandreou G, et al. (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV). 801-818.

[17] Yu C, Wang J, Peng C, et al. (2018) Bisenet: Bilateral segmentation network for real-time semantic segmentation. In: Proceedings of the European conference on computer vision (ECCV). 325-341.

[18] Yu C, Wang J, Peng C, et al. (2018) Learning a discriminative feature network for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 1857-1866.

[19] Zhang F, Chen Y, Li Z, et al. (2019) Acfnet: Attentional class feature network for semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 6798-6807.

[20] Zhu J-Y, Park T, Isola P, et al. (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. 2223-2232.

[21] Zhuang X, Shen J. (2016) Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Medical Image Analysis, 31: 77-87.

[22] Hua Y, Xu Z, Moon H, et al. (2018) Synseg-net: Synthetic segmentation without target modality ground truth. IEEE Transactions on Medical Imaging, 38: 1016-1025.

[23] Tsai Y-H, Hung W-C, Schulter S, et al. (2018) Learning to adapt structured output space for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 7472-7481.