Dual-Task Self-supervision
for Cross-modality Domain Adaptation

Yingying Xue, Shixiang Feng, Ya Zhang[(✉)], Xiaoyun Zhang,
and Yanfeng Wang

Cooperative Medianet Innovation Center,
Shanghai Jiao Tong University, Shanghai, China
{xueyingying,fengshixiang,ya_zhang,xiaoyun.zhang,wangyanfeng}@sjtu.edu.cn

Abstract. Data annotation is always an expensive and time-consuming
issue for deep learning based medical image analysis. To ease the need of
annotations, domain adaptation is recently introduced to generalize neu-
ral networks from a labeled source domain to unlabeled target domain
without much performance degradation. In this paper, we propose a novel
target domain self-supervision for domain adaptation by constructing an
edge generation auxiliary task to assist primary segmentation task so as
to extract better target representation and improve target segmentation
performance. Besides, in order to leverage detailed information contained
in low-level features, we propose a hierarchical low-level adversarial learn-
ing mechanism to encourage low-level features domain uninformative in a
hierarchical way, so that the segmentation performance can benefit from
low-level features without being affected by domain shift. Following these
two proposed approach, we develop a cross-modality domain adaptation
framework which employs the dual-task collaboration for target domain
self-supervision, and encourages low-level detailed features domain unin-
formative for better alignment. Our proposed framework achieves state-
of-the-art results on public cross-modality segmentation datasets.

Keywords: Domain adaptation · Dual-task collaboration ·
Hierarchical adversarial learning

1 Introduction

For medical image analysis based on deep learning, a great challenge remains
that deep learning models require high-quality and large quantities of annotated
images. This problem results in expensive data collection and repeatedly anno-
tation workload. Furthermore, annotating different image modalities such as CT
and MR of the same organ makes the issue more pronounced. Consequently, an
annotation-efficient deep learning method, namely unsupervised domain adap-
tation (UDA), is introduced to address cross-modality medical image analysis.

Unsupervised domain adaptation generalizes the learning model trained on
annotated source domain to another unlabeled target domain without any target
label supervision. For semantic segmentation task, many existing UDA methods
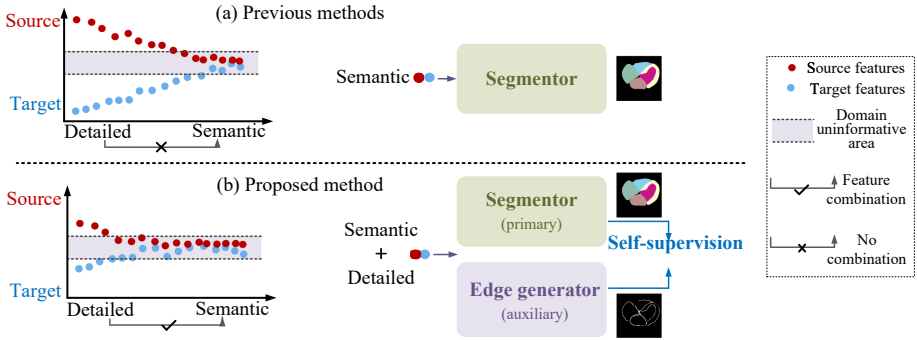
Fig. 1. Comparison between our proposed method and previous methods. (a) Previous methods where the detailed low-level domain informative features are not utilized for single way segmentation. (b) Proposed method encourages low-level detailed features to be domain uninformative for following segmentation, and employs self-supervision on target domain by constructing an edge generation task as auxiliary task.

[1–3,8] borrow the idea of image-to-image translation from CycleGAN [15] and multi-modal image translation network [6], so that the aligned images of two domains can be learned together under source domain supervision. Another main stream of UDA strategies employs adversarial learning to align source and target domain, where a common way is to follow [11] to utilize adversarial learning at the segmentation output [2,13], or at the segmentation entropy map [12,14], or in the VAE-based latent space [9].

These UDA methods have two drawbacks. Firstly, simple adversarial learning between source and target domain is not enough to completely align two domains, especially when unpaired source and target modalities vary much in medical images. Under unsupervised conditions, the edge region of target domain segmentation mask may be very inaccurate and has a high probability to over segmented or under segmented. Therefore, we propose a novel self-supervision on target domain to directly improve target domain performance. Specifically, we propose an auxiliary task that generates edges to assist primary segmentation task to improve prediction accuracy around contour. These two tasks are collaborated through a designed edge consistency function and their partially shared parameters, where two tasks share a common feature extractor and partial layers in decoder.

Secondly, existing methods employ segmentation task on aligned semantic features, without considering rich detailed information in low-level features, because domain information contained in low-level features can harm the adaptation performance. But detailed features can benefit medical image segmentation which are proved by the great success of U-net [10], and should also be considered. Therefore, to leverage detailed information in low-level features, while simultaneously reduce the adaptation degradation results from the skip-connection in U-net, we propose a hierarchical low-level adversarial learning mechanism to encourage low-level detailed features domain uninformative in a hierarchical way according to the content of domain.
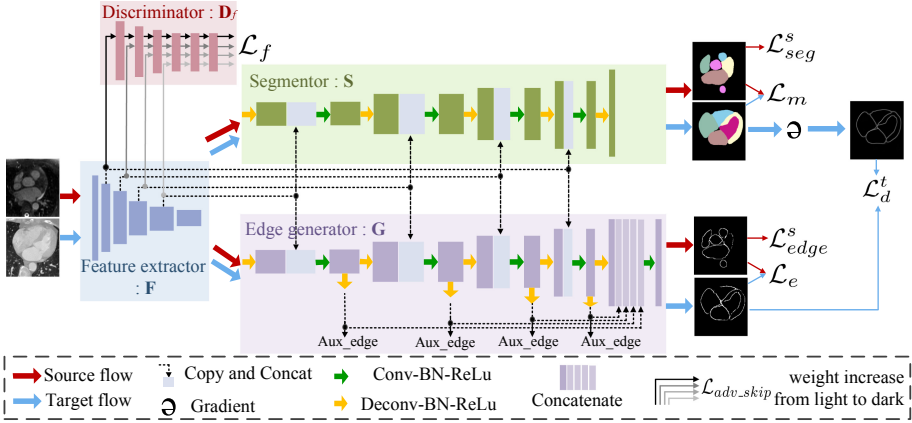
Fig. 2. Our proposed framework. The feature extractor F generates domain invariant features, and the hierarchical discriminator D_f differentiates input features accordingly. Segmentor S and edge generator G take features from corresponding layers of F to generate segmentation masks and edges. Two discriminators D_m and D_e (omitted in this figure) are employed at the output of S and G for adversarial learning.

In general, the comparison between our Dual-task and Hierarchical learning Network (DualHierNet) and previous UDA methods is shown in Fig. 1.

2   Methodology

Given N_s pixel-level labeled source domain data $\{X^s, Y^s\} = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ and N_t unlabeled target domain data $X^t = \{x_i^t\}_{i=1}^{N_t}$, unsupervised domain adaptation aims to use these data to learn a source to target adaptation network to correctly segment target images without any target domain supervision.

The architecture of proposed DualHierNet is shown in Fig. 2. Target domain self-supervision is achieved through the edge consistency between partially shared primary segmentation task S and auxiliary edge generation task G. Also, the low-level features extracted by feature extractor F are encouraged to be domain invariant through the adversarial learning with discriminator D_f in a hierarchical way according to the domain content. Lastly, two discriminators D_m and D_e are employed on output semantic space to align generated segmentation masks and edges respectively.

2.1   Dual-Task Collaboration for Target Domain Self-supervision

Under unsupervised conditions, the edge region of target domain segmentation mask may be inaccurate and has a high probability to over or under segmented. We therefore propose a novel target self-supervision by constructing an auxiliary task to generate edges, and making it collaborate with primary segmentation task to obtain a more accurate target segmentation mask at the edge region.

Specifically, feature extractor F generates domain invariant features from input images of source and target domain, where this part will be illustrated in later subsection. Domain invariant features are input to both segmentor S and edge generator G. The edge generator G has a similar network structure to segmentor S, with low-level features of F copied and concatenated to corresponding high-level features. Besides, edge generator employs deep-supervision for the purpose of providing auxiliary supervision to improve edge generation quality by outputting upsampled features in G as auxiliary edges shown in Fig. 2. These auxiliary edges are fused together to obtain the final generated edge p_e.

For source domain supervision, we use a combination of weighted cross-entropy loss and Dice loss: $\mathcal{L}^s(p^s, y^s) = \mathcal{L}^s_{wCE} + \mathcal{L}^s_{Dice}$, where p^s and y^s are segmentation mask and ground truth. We employ multi-class and two-class segmentation for segmentor S and edge generator G:

$$
\begin{aligned}
\mathcal{L}^s_{seg} &= \mathcal{L}^s(p^s_m, y^s_m), \\
\mathcal{L}^s_{edge} &= \mathcal{L}^s(p^s_e, y^s_e) + \sum_{Ae} \mathcal{L}^s(p^s_{Ae}, y^s_e),
\end{aligned}
\tag{1}
$$

where $\mathcal{L}^s_{seg}$ and $\mathcal{L}^s_{edge}$ are objective functions of S and G respectively. p^s_m and y^s_m is the segmentation mask and ground truth of source domain, and p^s_e and y^s_e is the generated edge and ground truth edge. Noted that y^s_e is obtained by calculating the first derivative of y^s_m. p^s_{Ae} are auxiliary edges shown in Fig. 2.

For target domain self-supervision, we encourage the segmentation mask p^t_m and generated edge p^t_e to keep consistency at the edges and we propose a dual-task consistency loss $\mathcal{L}^t_d$ on target domain. An operation $\boldsymbol{\partial}$ calculates the first derivative of soft segmentation mask p^t_m on two spatial axes i, j to obtain a soft edge, which should possess structural consistency with generated edge p^t_e. The consistency loss and the soft edge calculation formula are:

$$
\mathcal{L}^t_d = \left\| p^t_e - \boldsymbol{\partial}(p^t_m) \right\|^2_2,
$$
$$
\boldsymbol{\partial}(p^t_m) = \frac{1}{2}\left(\left|\frac{\partial p^t_m}{\partial i}\right| + \left|\frac{\partial p^t_m}{\partial j}\right|\right) \approx \frac{1}{2}\left(\sum_c \left|p^t_{m,i+1} - p^t_{m,i}\right| + \sum_c \left|p^t_{m,j+1} - p^t_{m,j}\right|\right),
\tag{2}
$$

where the summation symbol is applied to channel dimension c. The soft edge $\boldsymbol{\partial}(p^t_m)$ has a probability between [0,1].

2.2 Hierarchical Adversarial Learning for Better Alignment

Hierarchical Adversarial Learning. We follow the success of U-net in medical image segmentation [10] to combine low-level detailed features with high-level semantic features. However, low-level features are domain informative, and severe domain gap in detailed features can harm adaptation performance when combined with domain uninformative semantic features. We thus develop a hierarchical adversarial skip connection mechanism to make low-level detailed features domain invariant when concatenating them to semantic features simultaneously.

Specifically, feature extractor F maps input images to feature space, and we propose a hierarchical discriminator D_f to differentiate input domains accordingly. Features of each layer in F: l_1, l_2, l_3, l_4 and l_5, are gradually decreasing in domain information and increasing in semantic information. l_5 is directly input to following segmentor S and edge generator G, while l_1, l_2, l_3, l_4 are input to different layers of discriminator D_f in a hierarchical way according to their distinct resolutions for domain alignment. The objective function of layer $l_{k,k=1,2,...,K}$ is formulated as follows where F and D_f play a min-max game:

$$\mathcal{L}_f = \sum_{k=1}^{K} \gamma_k \mathcal{L}_{f,k},$$

$$\mathcal{L}_{f,k} = \mathbb{E}_{l_k^s \in F(X^s)}[\log(D_f(l_k^s))] + \mathbb{E}_{l_k^t \in F(X^t)}[\log(1 - D_f(l_k^t))],$$

(3)

where γ_k increases as k decreases, indicating that lower layer features contained more domain information is assigned with larger weights for attention.

Output Alignment. Finally, two discriminators D_m and D_e are employed in output space to align segmentation mask p_m and generate edge p_e with adversarial learning. $\mathcal{L}_m$ and $\mathcal{L}_e$ are the adversarial objective as follows respectively:

$$\mathcal{L}_m = \mathbb{E}_{x^s \sim X^s}[\log(D_m(p_m^s))] + \mathbb{E}_{x^t \sim X^t}[\log(1 - D_m(p_m^t))],$$

$$\mathcal{L}_e = \mathbb{E}_{x^s \sim X^s}[\log(D_e(p_e^s))] + \mathbb{E}_{x^t \sim X^t}[\log(1 - D_e(p_e^t))].$$

(4)

Therefore, with trade-off parameters $\lambda_0, \lambda_1, \lambda_2, \lambda_3$, the total objective function of the model is formulated as:

$$\min_{F,S,G} \max_{D_f,D_m,D_e} \mathcal{L}_{seg}^s + \mathcal{L}_{edge}^s + \lambda_0 \mathcal{L}_d^t$$

$$+ \lambda_1 \mathcal{L}_f + \lambda_2 \mathcal{L}_m + \lambda_3 \mathcal{L}_e.$$

(5)

3   Experiments and Results

Dataset and Implementation Details. The proposed framework is evaluated on the *Multimodality Whole Heart Segmentation Challenge* MMWHS2017 dataset [17] which consists of unpaired 20 CT and 20 MR volumes with pixel-level annotation of seven heart structures: left ventricle blood cavity (LV), right ventricle blood cavity (RV), left atrium blood cavity (LA), right atrium blood cavity (RA), myocardium of the left ventricle (Myo), ascending aorta (AA) and pulmonary artery (PA). We follow Pnp-AdaNet and SIFA [2,4] to use randomly selected sixteen MR volumes as source and sixteen CT as target for training. The remaining four CT volumes are for testing. Each volume is split into transverse view slices as inputs since doctors observe transverse view to diagnose cardiac diseases, and is augmented with flipping, rotation and scaling, and normalized to zero mean and unit variance and resized to 256×256. The volume metrics Dice score and Average Surface Distance (ASD) are employed for evaluation.

For fair comparison, 5-fold cross validation is employed. All annotations of CT are only used for evaluation without being presented during training.

We also validate our proposed method on another multi-modality cardiac dataset: MS-CMRSeg 2019 [16] which consists of 45 patients and each patient has cardiac images of three MR modalities: bSSFP, T2 and LGE. For fair comparison, we re-implement methods [1,13] under the same experiment setup with us, and follow [13] to combine labeled bSSFP and T2 as source and unlabeled LGE as target, where the target LGE is divided by competition, and use transverse view slices with the same preprocessing and augmentation as above.

The detailed dual-task architecture is shown in Fig. 2. Discriminators follow [7] to have 6 convolutional layers where the first 3 use instance normalization. Adam optimizers are utilized with a learning rate of 1.0×10^{-3} for segmentation and edge generation, but with a decay rate of 0.9 every 2 epochs for segmentation and no decay for edge generation, since we empirically found edge generation task converges slower than segmentation. The model is trained for 100 epochs with a batch size equals 4. Hyper-parameters λ_0 is 10, λ_1 is 1.0, while λ_2, λ_3 grow linearly from 0.0 to 1.0 as epoch increases to 40, and remain 1.0 subsequently.

Quantitative and Qualitative Analysis. For MMWHS2017 dataset, we validate our methods on seven structures and show the results in Table 1, and also follow [2,4] to validate on four left-side structures in Table 2. We compare with several state-of-the-art UDA methods including CyCADA [5], Pnp-AdaNet [4], BEAL [14], Cascaded U-net [1] and SIFA [2]. We re-implemented all above methods under the same experiment setup with five-fold cross validation shown in the mean $\pm$ std manner, and no post-processing is employed.

In Table 1, we first obtain the unadapted results by directly testing a source domain trained U-net on target domain, and a Dice score of 30.43% reflects severe domain shift between different modalities. A supervised target domain upper bound of 84.95% is also obtained through a supervised U-net. Our proposed method outperforms several UDA methods by a great margin and achieves superior performance of 73.68% in average Dice and 7.3 in average ASD. Note that our approach significantly improves the accuracy of LA by achieving a performance gain up to 9.4% in Dice, and even the most difficult structure to segment: Myo, is also improved to 64.03%. For four class segmentation of LV, LA, Myo and AA shown in Table 2, we achieve an average Dice of 76.98% and average ASD of 4.6, with great margin compared with other methods. Results on MS-CMRSeg shown in Table 3 prove the generalization ability of our method on cross-MR modalities by achieving an average Dice of 84.85%.

Visual results are shown in Fig. 3. Our DualHierNet has a smoother 3D heart with clearer contours, and better segmentation masks inside cardiac structures. For generated edges in lower part of Fig. 3, figures inside the red box are good examples that generated edge p_E^t and $\partial(p_M^t)$ are well constrained to be similar. Figures inside the blue box are poor examples, where the blue arrows point to boundary area that are distinct in p_E^t and $\partial(p_M^t)$. This usually results from incoherent annotation between two adjacent slices.

Table 1. Comparison results on *MMWHS2017* for 7 cardiac structures.

Methods	↑ Dice [%]								↓ ASD [voxel]							
	LV	RV	LA	RA	Myo	AA	PA	Average	LV	RV	LA	RA	Myo	AA	PA	Average
Upper Bound	88.02	86.01	85.42	87.84	81.27	82.22	83.91	84.95	5.4	4.7	4.2	4.6	5.4	2.3	4.8	4.5
	±4.34	±3.91	±5.28	±6.01	±4.24	±6.12	±7.71	±5.72	±3.2	±2.9	±2.4	±3.6	±3.7	±1.8	±2.6	±3.7
W/o Adaptation	15.87	15.21	52.14	45.65	6.54	43.82	33.81	30.43	52.4	56.8	31.9	24.4	36.0	42.1	31.2	39.26
	±3.72	±4.54	±13.74	±6.20	±3.98	±13.24	±5.77	±9.26	±16.4	±17.1	±15.0	±14.3	±29.6	±20.7	±22.8	±21.1
Pnp-AdaNet [4]	50.49	51.43	70.18	72.24	31.18	64.92	57.49	56.85	13.6	8.5	15.4	12.8	9.4	9.7	8.9	11.2
	±10.46	±6.72	±5.41	±4.18	±10.09	±8.44	±9.23	±8.20	±2.5	±3.7	±6.8	±5.1	±2.8	±4.9	±5.1	±5.2
CyCADA [5]	55.49	54.16	66.90	68.70	50.03	68.42	65.77	61.35	10.6	8.1	9.3	9.4	8.2	8.6	7.4	8.8
	±9.75	±13.47	±10.05	±5.42	±8.80	±9.67	±6.96	±9.58	±3.4	±4.2	±2.3	±5.5	±4.7	±4.6	±3.8	±4.2
BEAL [14]	65.23	58.01	64.00	62.04	56.10	72.05	61.95	62.77	10.2	7.9	9.2	10.1	8.4	8.2	7.0	8.7
	±7.85	±9.22	±8.44	±9.91	±6.80	±9.80	±9.67	±8.21	±3.8	±5.1	±2.1	±4.8	±3.6	±4.3	±3.9	±4.3
Cascaded U-net [1]	72.35	60.14	67.75	69.88	60.74	73.25	62.45	66.65	9.6	7.7	8.8	9.7	8.2	8.2	8.4	
	±8.33	±9.24	±9.18	±8.74	±5.65	±7.42	±8.77	±9.33	±4.1	±5.6	±3.0	±3.6	±3.0	±3.1	±4.2	±4.0
SIFA [2]	79.56	**70.79**	69.77	72.02	61.48	75.59	63.82	70.43	8.4	7.2	8.1	8.6	8.0	6.6	6.8	7.7
	±8.17	±11.24	±9.55	±7.00	±6.57	±9.45	±8.45	±8.73	±4.1	±3.2	±4.6	±4.9	±2.8	±3.1	±3.2	±3.9
DualHierNet (Ours)	**81.58**	70.23	**79.17**	**73.30**	**64.03**	**80.25**	**67.22**	**73.68**	**8.2**	**7.2**	**7.0**	**8.4**	**7.7**	**6.4**	**6.3**	**7.3**
	±9.70	±8.45	±9.19	±8.29	±7.97	±7.24	±7.16	±8.93	±3.6	±3.0	±2.8	±2.9	±2.6	±3.2	±2.7	±3.4

Table 2. Comparison results on *MMWHS2017* for 4 cardiac structures.

Methods	↑ Dice [%]					↓ ASD [voxel]				
	LV	LA	Myo	AA	Average	LV	LA	Myo	AA	Average
Upper Bound	89.54	87.72	86.06	85.21	87.13	3.3	4.3	4.9	1.6	3.5
	±5.27	±1.58	±1.97	±6.67	±5.24	±3.0	±2.6	±1.2	±1.1	±3.3
W/o Adaptation	13.42	46.05	12.25	20.39	23.03	47.1	22.8	24.5	42.7	34.3
	±2.31	±20.32	±10.92	±10.96	±18.47	±17.9	±19.9	±12.1	±12.6	±19.8
Pnp-AdaNet [4]	52.32	75.75	28.73	73.86	57.67	9.2	13.6	8.8	11.5	10.8
	±21.00	±4.35	±13.31	±7.48	±13.85	±3.9	±3.6	±4.3	±2.9	±4.8
CycADA [5]	61.90	68.95	50.83	74.08	63.94	11.4	6.3	14.7	10.8	10.8
	±10.78	±5.26	±7.06	±7.30	±8.59	±7.0	±2.3	±3.8	±3.2	±5.1
BEAL [14]	68.49	62.77	57.93	75.47	66.17	9.8	7.0	8.9	7.7	8.4
	±11.23	±9.47	±7.59	±9.67	±10.25	±5.0	±3.4	±4.9	±5.2	±4.9
Cascaded U-net [1]	70.05	65.28	60.66	77.36	68.34	9.2	6.8	8.2	7.6	8.0
	±9.60	±10.16	±9.74	±6.28	±9.31	±4.2	±4.9	±4.9	±3.4	±4.8
SIFA [2]	77.39	73.35	60.68	84.55	73.99	4.2	5.5	5.9	**4.4**	5.0
	±11.04	±12.20	±5.02	±4.05	±8.24	±2.5	±3.8	±3.7	±2.1	±3.4
DualHierNet (Ours)	**83.42**	**74.61**	**65.19**	**84.70**	**76.98**	**3.6**	**5.3**	**4.8**	4.5	**4.6**
	±7.46	±10.07	±6.33	±6.41	±7.84	±1.7	±2.0	±2.2	±2.8	±2.3

Table 3. Comparison results on *MS-CMRSeg*.

Methods	↑ Dice [%]				↓ ASD [voxel]			
	LV	RV	Myo	Average	LV	RV	Myo	Average
Cascaded U-net [1]	82.05	80.90	81.47	81.47	1.8	1.9	1.8	1.8
U-net+GFRM [13]	85.42	83.64	79.62	82.89	1.6	1.7	1.8	1.7
DualHierNet (Ours)	**88.91**	**84.01**	**81.62**	**84.85**	**1.5**	**1.7**	**1.8**	**1.7**

Ablation Study. Firstly, we conduct an ablation experiment to evaluate the effectiveness of each component: (i) U-net with output adversarial learning (Base), (ii) Base equipped with dual-task collaboration (Base+Dual), (iii) Base with hierarchical adversarial learning (Base+Hier), and (iv) ours (Base+Dual+Hier). In Table 4, the performance is improved to 68.50% and
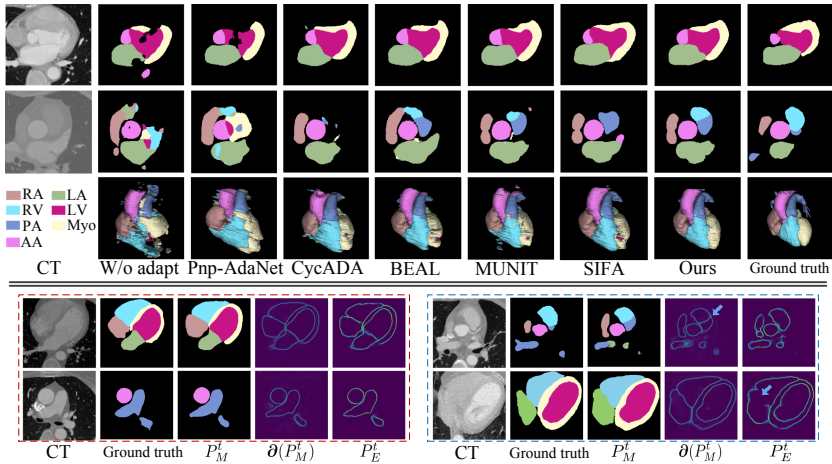
Fig. 3. Visual results of comparison and generated edges.

Table 4. Effects of each component.

Methods	LV	LA	Myo	AA	Avg Dice [%]
W/o Adaptation	13.42	46.05	12.25	20.39	23.03
Base	67.20	54.46	53.27	69.39	61.08
Base+Dual	76.82	60.62	57.48	79.09	68.50
Base+Hier	78.11	63.74	60.42	81.28	70.89
Base+Dual+Hier	**83.42**	**74.61**	**65.19**	**84.70**	**76.98**

Table 5. Effect of hierarchical weights.

$\gamma_1:\gamma_2:\gamma_3:\gamma_4$	LV	RV	Myo	AA	Avg Dice [%]
1 : 2 : 3 : 4	78.10	67.05	58.14	79.32	70.65
1 : 1 : 1 : 1	79.24	69.95	60.33	80.90	72.61
4 : 3 : 2 : 1	**83.42**	**74.61**	**65.19**	**84.70**	**76.98**

Table 6. Dual-task self-supervision extended on supervised setting.

Target domain	Model	LV	RV	Myo	AA	Avg Dice [%]
Supervised	Seg+Seg	90.02	87.62	86.98	89.40	88.51
	Seg+Edge	**90.17**	**89.06**	**88.11**	**91.24**	**89.65**
Adapted	Seg+Seg	79.64	68.11	60.28	79.69	71.93
	Seg+Edge	**83.42**	**74.61**	**65.19**	**84.70**	**76.98**

70.89% equipped with our proposed dual-task self-supervision and hierarchical strategy respectively. The further improvement to 76.98% in our DualHierNet confirms the effect of using dual-task as self-supervision and hierarchically aligning low-level features.

Secondly, we experiment on choice of hierarchical weights γ_k shown in Table 5. When we assign larger weights to higher layers, only an average Dice of 70.65% is achieved. A Dice of 72.61% is achieved if each layer shares a same weight. When we enlarge the weights of shallow layers which contain more domain information,

we can get a Dice of 76.98%. This further justify that low-level domain informative features should receive stronger adversarial learning attention.

Thirdly, we extend on target-only supervised segmentation to validate our proposed self-supervision. We replace *Seg+Edge* structure with two segmentors *Seg+Seg* so that they have nearly the same number of parameters. In supervised setting, *Seg+Edge* uses segmentation loss and dual_consist loss while *Seg+Seg* only uses segmentation loss to train the networks. Results shown in Table 6 reveals that auxiliary edge task assists segmentation even on supervised setting and achieve performance gain of 1.14%. While in adapted setting, a larger performance gain is obtained through our proposed dual-task self-supervision.

4   Conclusion

We propose a dual-task collaboration framework for target self-supervision with low-level hierarchical adversarial learning for cross-modality image segmentation. We develop a novel self-supervision by constructing an auxiliary task to generate edges to assist segmentation task, and we also design a hierarchical adversarial mechanism according to the content of domain. Our framework outperforms several adaptation methods on cross-modality datasets and the proposed dual-task architecture even achieves promising performance in supervised setting.

References

1. Chen, C., et al.: Unsupervised multi-modal style transfer for cardiac MR segmentation. In: Pop, M., et al. (eds.) STACOM 2019. LNCS, vol. 12009, pp. 209–219. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-39074-7_22
2. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Synergistic image and feature adaptation: towards cross-modality domain adaptation for medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 865–872 (2019)
3. Chen, Y.C., Lin, Y.Y., Yang, M.H., Huang, J.B.: Crdoco: pixel-level domain transfer with cross-domain consistency. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1791–1800 (2019)
4. Dou, Q., et al.: PNP-adanet: plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation. IEEE Access (2019)
5. Hoffman, J., et al.: Cycada: Cycle-consistent adversarial domain adaptation. In: International Conference on Machine Learning, pp. 1989–1998 (2018)
6. Huang, X., Liu, M.Y., Belongie, S., Kautz, J.: Multimodal unsupervised image-to-image translation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 172–189 (2018)
7. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134 (2017)

8. Jiang, J., et al.: Cross-modality (CT-MRI) prior augmented deep learning for robust lung tumor segmentation from small MR datasets. Med. Phys. **46**(10), 4392–4404 (2019)
9. Ouyang, C., Kamnitsas, K., Biffi, C., Duan, J., Rueckert, D.: Data efficient unsupervised domain adaptation for cross-modality image segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 669–677. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_74
10. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
11. Tsai, Y.H., Hung, W.C., Schulter, S., Sohn, K., Yang, M.H., Chandraker, M.: Learning to adapt structured output space for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7472–7481 (2018)
12. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: adversarial entropy minimization for domain adaptation in semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2517–2526 (2019)
13. Wang, J., Huang, H., Chen, C., Ma, W., Huang, Y., Ding, X.: Multi-sequence cardiac MR segmentation with adversarial domain adaptation network. In: Pop, M., et al. (eds.) STACOM 2019. LNCS, vol. 12009, pp. 254–262. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-39074-7_27
14. Wang, S., Yu, L., Li, K., Yang, X., Fu, C.-W., Heng, P.-A.: Boundary and entropy-driven adversarial learning for fundus image segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 102–110. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_12
15. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)
16. Zhuang, X.: Multivariate mixture model for myocardial segmentation combining multi-source images. IEEE Trans. Pattern Anal. Mach. Intell. **41**(12), 2933–2946 (2018)
17. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Med. Image Anal. **31**, 77–87 (2016)