国防科技大学
National University of Defense Technology

IEEE BIBM 2021
Virtual Event • 9–12 December

# Tri-Directional Tasks Complementary Learning for Unsupervised Domain Adaptation of Cross-modality Medical Image Semantic Segmentation

**Chen Li†, Wei Chen†✉, Mingfei Wu,Xin Luo , Yulin He, Yusong Tan**

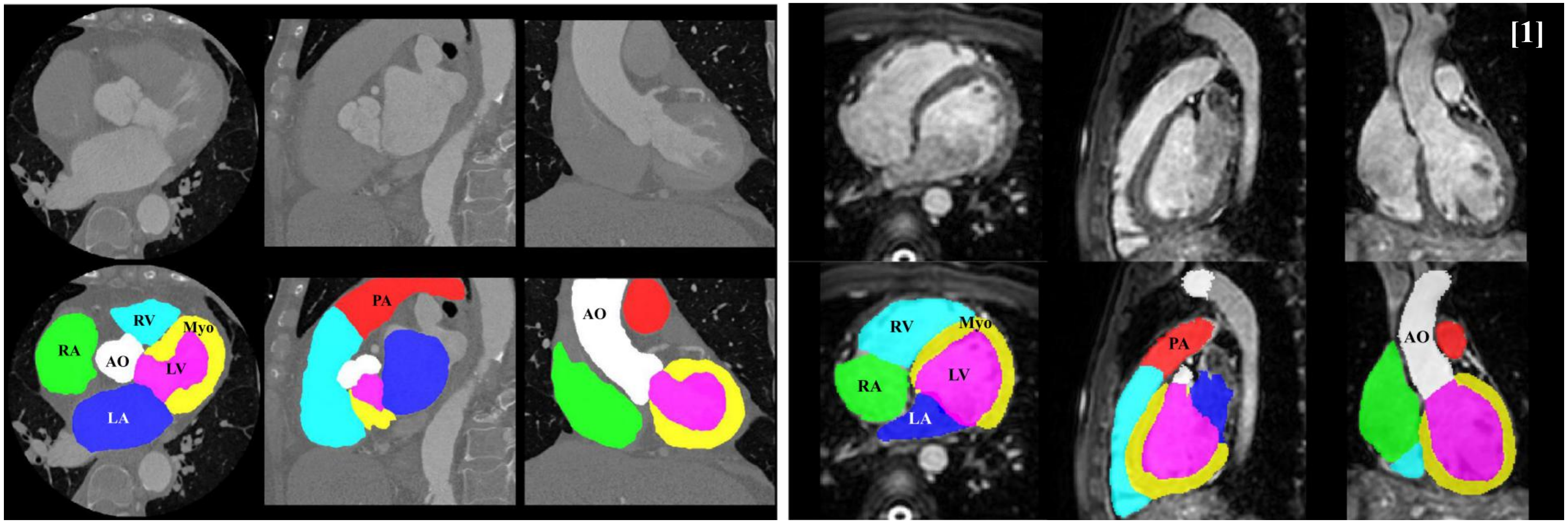
College of Computer, National University of Defense Technology, Changsha, China

† These authors contributed equally to this work

✉ chenwei@nudt.edu.edu

# Medical Images Segmentation (MIS)

- Medical image segmentation means classifying pixel-wise segments into different components from biomedical data (CT, MRI, Ultrasound, cells scan …… )




- Medical image segmentation is an essential step and plays a crucial role in many clinical applications, such as disease diagnosis and treatment planning.
- Segmentation from medical images is more challenging than natural image.

*[1]* Zhuang, Xiahai, et al. "Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge." Medical image analysis 58 (2019): 101537.

# Limitations in supervised MIS methods

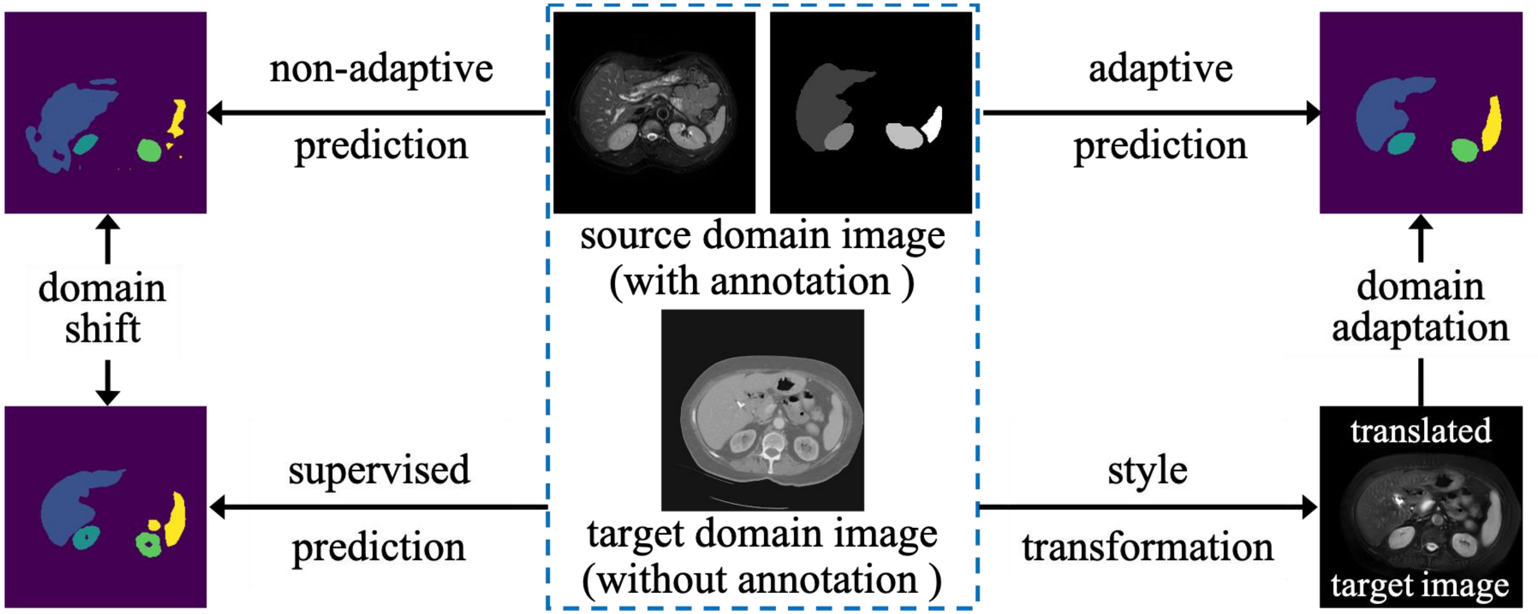
- Supervised methods have shown promising performances in various medical image segmentation tasks.

- Well-trained models often fail when deployed to real-world clinical scenarios, as medical images acquired with different acquisition parameters or modalities have very different characteristics.

- Such cross-modality domain shift would lead to severe performance degradation of deep networks.

# Unsupervised Domain Adaptation (UDA)

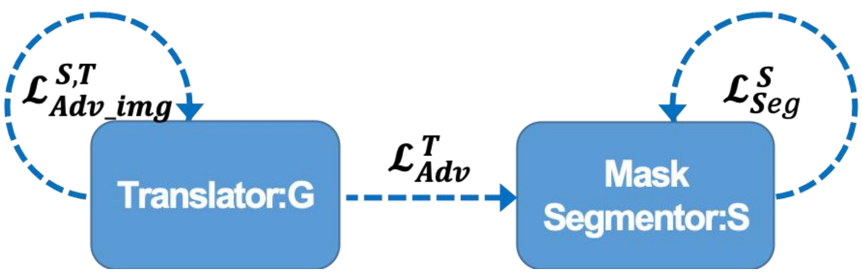
➢ The main idea of UDA is to extract domain-invariable representations and transfer them from source domain to target domain, where source samples are annotated and the the labels of target samples are absent.



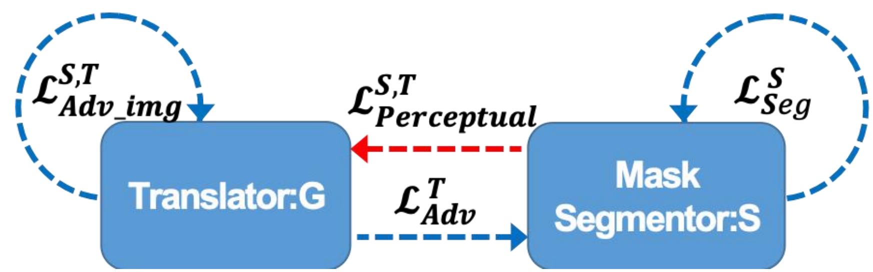
# Motivation

➢ Previous UDA methods achieve this goal by aligning domains, including image-level alignment[1], feature-level alignment[2] and model-level alignment[3].

➢ Try to align domains in a single direction or two directions, failing to take advantage of the complementary relationship between different directions and alignment tasks. <u>Ignoring the complement relationship between above alignment.</u>




Single-directional Learning[1,2]




Bio-directional Learning[4]

*[1] Hoffman, Judy, et al. "Cycada: Cycle-consistent adversarial domain adaptation." International conference on machine learning (ICML). PMLR, 2018.*

*[2] Chen, Cheng, et al. "Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation." IEEE transactions on medical imaging (TMI) 39.7 (2020): 2494-2505.*

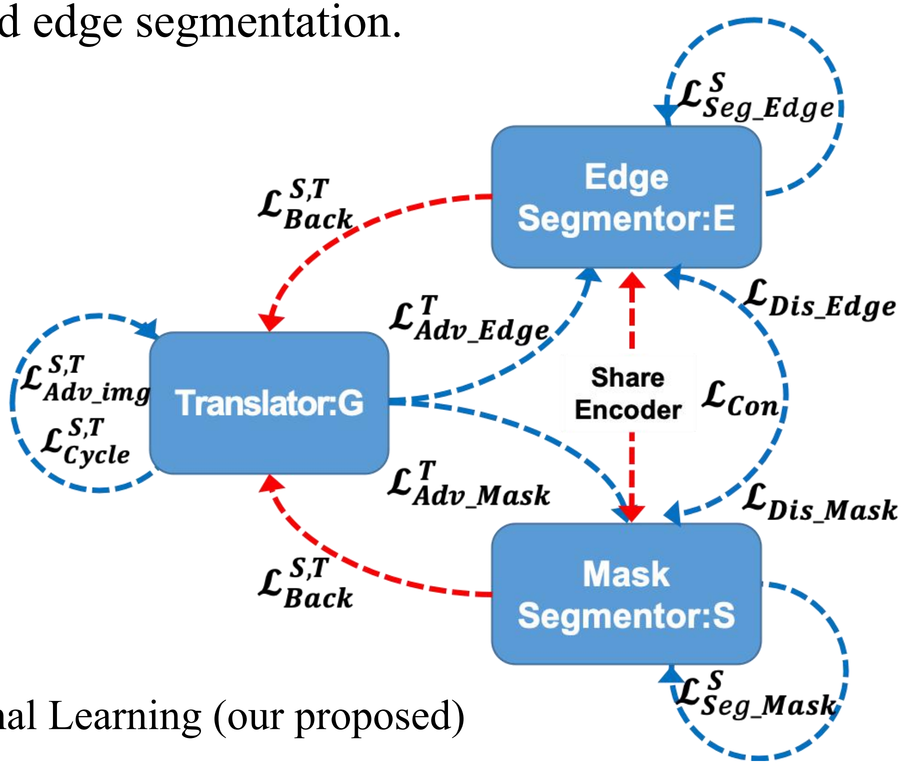
*[3] Li, Rui, et al. "Model adaptation: Unsupervised domain adaptation without source data." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020.*

*[4] Li, Yunsheng, Lu Yuan, and Nuno Vasconcelos. "Bidirectional learning for domain adaptation of semantic segmentation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019.*

# Motivation

➢ We proposed to solve unsupervised bio-medical cross-modality domain adaptation and carry out collaborative learning among three components ( image translator G, mask segmentor M and edge segmentor E) .

➢The above three components perform three tasks respectively, cross-modality style transformation, mask segmentation and edge segmentation.




Tri-directional Learning (our proposed)

# Tri-Tasks Learning for Unsupervised Domain Adaptation

- The first task is the ***<u>cross-modality style transformation</u>***.
- There are two generators ($G_{T2S}$ and $G_{S2T}$ ) and discriminators ($D_S$ and $D_T$ ) to adversarially generate image with new style.
- The goal of generators is to make synthetic images look similar to real images while discriminators aim to classify all images correctly.

$$\mathcal{L}^{S}_{Adv\_img} = \sum_{x_s \in \mathbb{X}_S} [\log D_S(x_s)] + \sum_{x_t \in \mathbb{X}_T} [\log (1 - D_S(G_{T2S}(x_t)))]$$

$$\mathcal{L}^{T}_{Adv\_img} = \sum_{x_t \in \mathbb{X}_T} [\log D_T(x_t)] + \sum_{x_s \in \mathbb{X}_S} [\log (1 - D_T(G_{S2T}(x_s)))]$$

# Tri-Tasks Learning for Unsupervised Domain Adaptation

- The first task is the ***cross-modality style transformation***.
- Besides, our work refer the CycleGAN [1] and design cycle_x0002_consistent loss function to avoid contradiction between cross-modality generators when adversarial training.

$$\mathcal{L}_{Cycle} = \sum_{x_s \in \mathbb{X}_S} \left[ \| G_{T2S}(G_{S2T}(x_s)) - x_s \|_1 \right] + \sum_{x_t \in \mathbb{X}_T} \left[ \| G_{S2T}(G_{T2S}(x_t)) - x_t \|_1 \right].$$

*[1] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." in IEEE international conference on computer vision (ICCV). 2017.*

# Tri-Tasks Learning for Unsupervised Domain Adaptation

- The second task is ***mask segmentation with cross-domain adaptation***.
- Based on DeepLab V2 [1], the mask segmentor mainly consists of two components, i.e., feature extractor and mask generator.

$$\mathcal{L}_{Seg\_Mask} = \sum_{x_s \in \mathbb{X}_S, y_s \in \mathbb{Y}_S} [-y_s \log M(x_s)]$$

*[1] Chen, Liang-Chieh, et al. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs." IEEE transactions on pattern analysis and machine intelligence (T-PAMI) 40.4 (2017): 834-848.*

# Tri-Tasks Learning for Unsupervised Domain Adaptation

- The third task is ***edge segmentationwith cross-domain adaptation***.
- Similiar to the mask segmentation, the edge segmentor consists of two components,

i.e., feature extractor and edge generator.

$$\mathcal{L}_{Seg\_Edge} = \sum_{x_s \in \mathbb{X}_S, y_s \in \mathbb{Y}_S} [-\xi(y_s) \log E(x_s)]$$

# Tri-Tasks Learning for Unsupervised Domain Adaptation

- The third task is ***edge segmentationwith cross-domain adaptation***.
- Discriminators $D_M$ and $D_E$ are used to distinguish the mask and edge respectively

$$\mathcal{L}_{Dis\_Mask} = \sum_{x_s \in \mathbb{X}_S} \mathcal{L}_{bce}\{D_M[M(x_s)], M\} + \sum_{x_t \in \mathbb{X}_T} \mathcal{L}_{bce}\{D_M[M(G_{T2S}(x_t))], M\},$$

$$\mathcal{L}_{Dis\_Edge} = \sum_{x_s \in \mathbb{X}_S} \mathcal{L}_{bce}\{D_E[E(x_s)], E\} + \sum_{x_t \in \mathbb{X}_T} \mathcal{L}_{bce}\{D_E[E(G_{T2S}(x_t))], E\}.$$

# Tri-directional Collaborative Learning

- The first direction is the ***translator boosts the mask segmentor and edge segmentor***.
- After image style transformation, we utilize the labeled source samples and unlabeled translated target samples to train the segmentors M and E.

$$\mathcal{L}_{Adv\_Mask} = \sum_{x_t \in \mathbb{X}_T} \mathcal{L}_{bce}\{D_M[M(G_{T2S}(x_t))], S\}$$

$$\mathcal{L}_{Adv\_Edge} = \sum_{x_t \in \mathbb{X}_T} \mathcal{L}_{bce}\{D_E[E(G_{T2S}(x_t))], S\}.$$

- In summary, we can derive the optimization problem to boost M and E with the help of transformation G:

$$\min_{\theta_M} \max_{\theta_{D_M}} \left[\mathcal{L}_{Seg\_Mask} + \lambda_{Adv}\mathcal{L}_{Adv\_Mask}\right],$$

$$\min_{\theta_E} \max_{\theta_{D_E}} \left[\mathcal{L}_{Seg\_Edge} + \lambda_{Adv}\mathcal{L}_{Adv\_Edge}\right].$$

# Tri-directional Collaborative Learning

- The second direction is the mask segmentor and edge segmentor work collaboratively with each other.
- In order to obey the fact that the prediction of edge and the boundary of mask should keep consistent, we propose to align masks and edges in the self-supervised manner.

$$\mathcal{L}_{Con} = \sum_{x_t \in \mathbb{X}_T} \mathcal{L}_{bce} \{\xi(M[G_{T2S}(x_t)]), E[G_{T2S}(x_t)]\}$$

- In summary, we can achieve optimization through self-supervised collaboration between M and E:

$$\min_{\theta_{M,E}} \max_{\theta_{D_M, D_E}} [\mathcal{L}_{Dis\_Mask} + \mathcal{L}_{Dis\_Edge} + \lambda_{Con}\mathcal{L}_{Con}]$$

# Tri-directional Collaborative Learning

- The third direction is the well-trained mask segmentor and edge segmentor promote the translator in return.
- we reconstruct the translated image again to optimize style transformation and maintain the semantic consistency between original sample and translated sample at the same time.

$$\begin{aligned}\mathcal{L}_{Back}^{S} = \textstyle\sum_{x_s \in \mathbb{X}_S} [ & \|M(G_{S2T}(x_s)) - M(x_s)\|_1 \\ & + \|M(G_{T2S}(G_{S2T}(x_s))) - M(x_s)\|_1 \\ & + \|E(G_{S2T}(x_s)) - E(x_s)\|_1 \\ & + \|E(G_{T2S}(G_{S2T}(x_s))) - E(x_s)\|_1].\end{aligned}$$

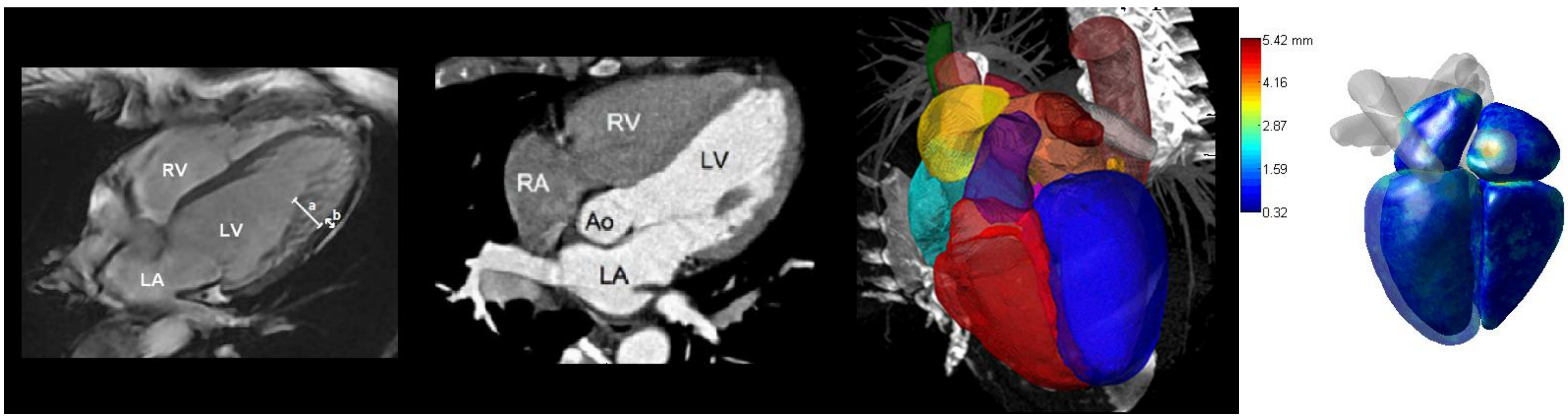
$$\begin{aligned}\mathcal{L}_{Back}^{T} = \textstyle\sum_{x_t \in \mathbb{X}_T} [ & \|M(G_{T2S}(x_t)) - M(x_t)\|_1 \\ & + \|M(G_{S2T}(G_{T2S}(x_t))) - M(x_t)\|_1 \\ & + \|E(G_{T2S}(x_t)) - E(x_t)\|_1 \\ & + \|E(G_{S2T}(G_{T2S}(x_t))) - E(x_t)\|_1].\end{aligned}$$

- In summary, with the collaboration of segmentors M and E, the optimization of translator is updated:

$$\begin{aligned}\min_{\theta_{G_{S2T}, G_{T2S}}} \max_{\theta_{D_S, D_T}} & [\mathcal{L}_{Adv\_img}^{S} + \mathcal{L}_{Adv\_img}^{T} + \mathcal{L}_{Cycle} \\ & + \lambda_{Back}(\mathcal{L}_{Back}^{S} + \mathcal{L}_{Back}^{T})].\end{aligned}$$

# Dataset & Metric

- Multi-modality Whole Heart Segmentation Challenge (MM-WHS 2017)
  - 20 CT volumes and 20 MRI volumes
  - seven cardiac structures with pixel-level annotation.
  - supported by MICCAI 2017
  - http://www.sdspeople.fudan.edu.cn/zhuangxiahai/0/mmwhs/




- Evaluation Metrics:
  - Dice coefficient and Average symmetric surface distance (ASSD)

# Ablation study

- Firstly, we trained the mask segmentor without any domain adaptation technique and regarded the results as the baseline.
- Secondly, we delved into the key components of TriDL and divided them step by step for further proof.
- In the end, we merged all components to get the best TriDL model for final validation.

| Methods (Mean $\pm$ Std.) | AVG. Dice (%) | AVG. ASSD (voxel) |
|---|---|---|
| Without Adapation | 25.70$\pm$0.78 | 26.23$\pm$6.47 |
| Only Style Transformation $G_{(0)}$ | 62.91$\pm$10.81 | 4.97$\pm$1.57 |
| $G_{(0)}$ + Mask Segmentation $M_{(0)}$ | 64.15$\pm$6.49 | 4.97$\pm$1.69 |
| $G_{(0)}$+ $M_{(0)}$+ Edge Segmentation $E_{(0)}$ | 69.77$\pm$7.08 | 4.23$\pm$1.32 |
| Updated $G_{(1)}$ + $M_{(0)}$ + $E_{(0)}$ | 72.00$\pm$7.46 | 3.28$\pm$0.81 |
| $G_{(1)}$ + Updated $M_{(1)}$ + $E_{(0)}$ | 74.52$\pm$6.47 | 3.23$\pm$0.44 |
| $G_{(1)}$ + $M_{(1)}$ + Updated $E_{(1)}$ | 77.16$\pm$6.70 | 2.80$\pm$0.64 |

# Quantitatively measure the domain shift

➢ In addition to quantitative representation, we also conducted significance tests for some key components pair.

| Methods (Dice) | Ascending Aorta (AA) | Left Atrium blood Cavity (LAC) | Left Ventricle blood Cavity (LVC) | Myocardium of left ventricle (MYO) | Average |
|---|---|---|---|---|---|
| Without adapation | 27.57±16.37 | 27.63±14.29 | 33.79±8.81 | 13.81±9.44 | 25.70±0.78 |
| Supervised learning | 92.43±2.31 | 83.84±8.87 | 91.09±4.72 | 85.39±6.75 | 88.18±4.47 |

# Quantitatively measure the domain shift

- In addition to quantitative representation, we also conducted significance tests for some key components pair.

| Methods (ASSD) | Ascending Aorta (AA) | Left Atrium blood Cavity (LAC) | Left Ventricle blood Cavity (LVC) | Myocardium of left ventricle (MYO) | Average |
|---|---|---|---|---|---|
| Without Adaptation | 43.00±23.27 | 21.14±10.24 | 14.22± 4.15 | 26.54±9.63 | 26.23±6.47 |
| Supervised learning | 1.10±0.26 | 1.80±0.69 | 0.96±0.48 | 1.14±0.60 | 1.25±0.45 |

# Comparison with the State-of-the-art Methods

➢ The quantitative results of six SOTA UDA methods and proposed method demostate the superior of our work in domain-adaptive segmentation of cardiac organs.

| Methods (Dice) | Ascending Aorta (AA) | Left Atrium blood Cavity (LAC) | Left Ventricle blood Cavity (LVC) | Myocardium of left ventricle (MYO) | Average |
|---|---|---|---|---|---|
| Without adapation | 27.57±16.37 | 27.63±14.29 | 33.79±8.81 | 13.81±9.44 | 25.70±0.78 |
| ADDA [6] | 47.60 | 60.90 | 11.20 | 29.20 | 37.20 |
| DANN [7] | 39.00 | 45.10 | 28.30 | 25.70 | 34.50 |
| Pnp-AdaNet [8] | 74.00±7.30 | 68.90±5.20 | 61.90±10.70 | 50.80±7.00 | 63.90±7.50 |
| SynSeg-Net [9] | 71.60 | 69.00 | 51.60 | 40.80 | 58.20 |
| CycleGAN [2] | 73.80 | 75.70 | 52.30 | 28.70 | 57.60 |
| CyCADA [10] | 72.90 | 77.00 | 62.40 | 45.30 | 64.40 |
| BEAL [11] | 75.47±9.67 | 62.77±9.47 | 68.49±11.23 | 57.93±7.59 | 66.17±10.25 |
| Cascaded U-Net [12] | 77.36±6.28 | 65.28±10.16 | 70.05±9.60 | 60.66±9.74 | 68.34±9.31 |
| SIFA-v1 [13] | 81.10 | 76.40 | 75.70 | 58.70 | 73.00 |
| SIFA-v2 [14] | 81.30 | 79.50 | 73.80 | 61.60 | 74.10 |
| DualHierNet [15] | 84.70±6.41 | 74.61±10.01 | **83.42±7.46** | **65.19±6.33** | 76.98±7.84 |
| BDL [16] | 84.34±6.76 | 71.31±18.70 | 77.04±3.50 | 60.36±13.73 | 73.26±10.31 |
| DSFN [17] | 84.70 | 76.90 | 79.10 | 62.40 | 75.80 |
| **TriDL** (ours) | **88.87±6.70** | **78.85±15.64** | 78.64±2.70 | 62.27±9.69 | **77.16±6.70** |
| Supervised learning | 92.43±2.31 | 83.84±8.87 | 91.09±4.72 | 85.39±6.75 | 88.18±4.47 |

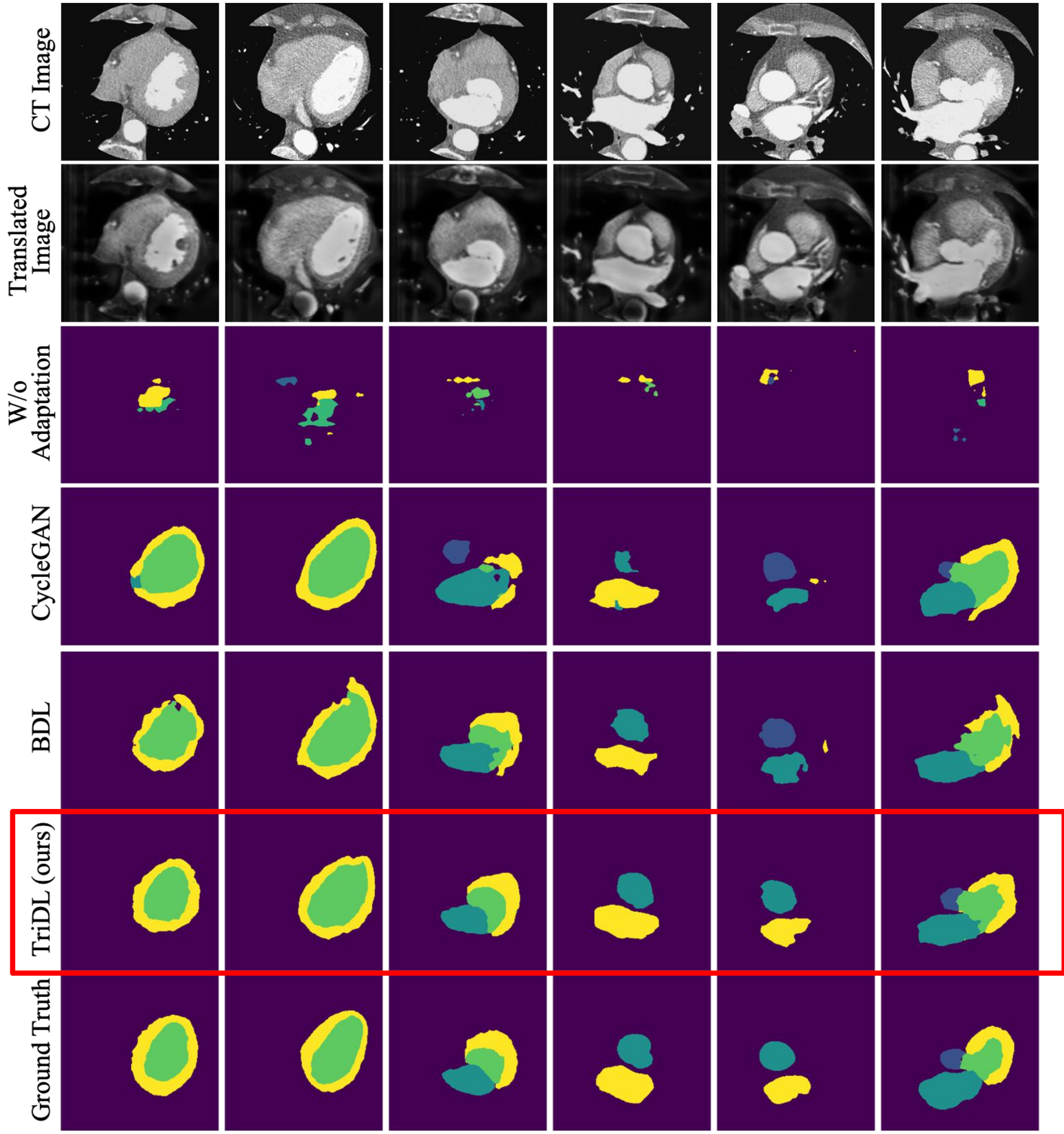
# Comparison with the State-of-the-art Methods

➢ The quantitative results of six SOTA UDA methods and proposed method demostate the superior of our work in domain-adaptive segmentation of cardiac organs.

| Methods (ASSD) | Ascending Aorta (AA) | Left Atrium blood Cavity (LAC) | Left Ventricle blood Cavity (LVC) | Myocardium of left ventricle (MYO) | Average |
|---|---|---|---|---|---|
| Without Adaptation | 43.00±23.27 | 21.14±10.24 | 14.22± 4.15 | 26.54±9.63 | 26.23±6.47 |
| ADDA [16] | 13.80 | 10.20 | N/A | 13.40 | N/A |
| DANN [17] | 16.20 | 9.20 | 12.10 | 10.10 | 11.90 |
| Pnp-AdaNet [26] | 12.80±3.20 | 6.30±2.30 | 17.40±7.00 | 14.70±4.80 | 12.80±4.30 |
| SynSeg-Net [27] | 11.70 | 7.80 | 7.00 | 9.20 | 8.90 |
| CycleGAN [3] | 11.50 | 13.60 | 9.20 | 8.80 | 10.80 |
| CyCADA [4] | 9.60 | 8.00 | 9.60 | 10.50 | 9.40 |
| BEAL [9] | 7.70±5.20 | 7.00±3.40 | 9.80±5.00 | 8.90±4.90 | 8.40±4.90 |
| Cascaded U-Net [28] | 7.60±3.40 | 6.80±4.90 | 9.20±4.20 | 8.20±4.90 | 8.00±4.80 |
| SIFA-v1 [6] | 10.60 | 7.40 | 6.70 | 7.80 | 8.10 |
| SIFA-v2 [18] | 7.90 | 6.20 | 5.50 | 8.50 | 7.00 |
| DualHierNet [19] | 4.50±2.80 | 5.30±2.00 | 3.60±1.70 | 4.90±2.20 | 4.60±2.30 |
| BDL [29] | 1.89±0.53 | 4.01±1.82 | 3.64±0.73 | 3.69±1.78 | 3.31±1.18 |
| **TriDL** (ours) | **1.55±0.46** | **3.06±1.54** | **3.61±0.95** | **2.98±0.43** | **2.80±0.64** |
| Supervised learning | 1.10±0.26 | 1.80±0.69 | 0.96±0.48 | 1.14±0.60 | 1.25±0.45 |

# Comparison with the State-of-the-art Methods

- The qualitative segmentation results also showed our method can successfully locate the four organs and generate semantically meaningful mask.



# Summary

- We propose an unsupervised domain-adaptive framework (TriDL) for cross-modality medical image semantic segmentation.
  - Our framework is able to synergize cross-modality style transformation and mask segmentation and edge segmentation.
  - These tasks collaborately work to learn the domain-adaptive representations and effectively promote each other.

# Summary

- We propose an unsupervised domain-adaptive framework (TriDL) for cross-modality medical image semantic segmentation.
    - Our framework is able to synergize cross-modality style transformation and mask segmentation and edge segmentation.
    - These tasks collaborately work to learn the domain-adaptive representations and effectively promote each other.

- We would like to thank:
    - National Key Research and Development Program of China (No. 2018YFB0204301)
- Related material:
    - *https://github.com/lichen14/TriDL*

- We are grateful for corrections and discussions!

# Tri-Directional Tasks Complementary Learning for Unsupervised Domain Adaptation of Cross-modality Medical Image Semantic Segmentation

# Thank You!