



国防科技大学

National University of Defense Technology



AttENT: Domain-Adaptive Medical Image Segmentation via Attention-Aware Translation and Adversarial Entropy Minimization

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

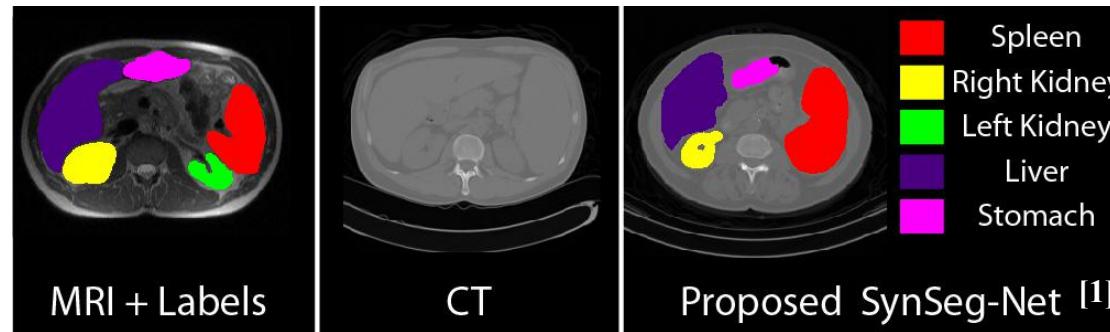
College of Computer, National University of Defense Technology, Changsha, China

† 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] Huo, Yuankai, et al. "Synseg-net: Synthetic segmentation without target modality ground truth." *IEEE transactions on medical imaging (TMI)* 38.4 (2018): 1016-1025.

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 labels of target samples are absent.

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

[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.

Motivation

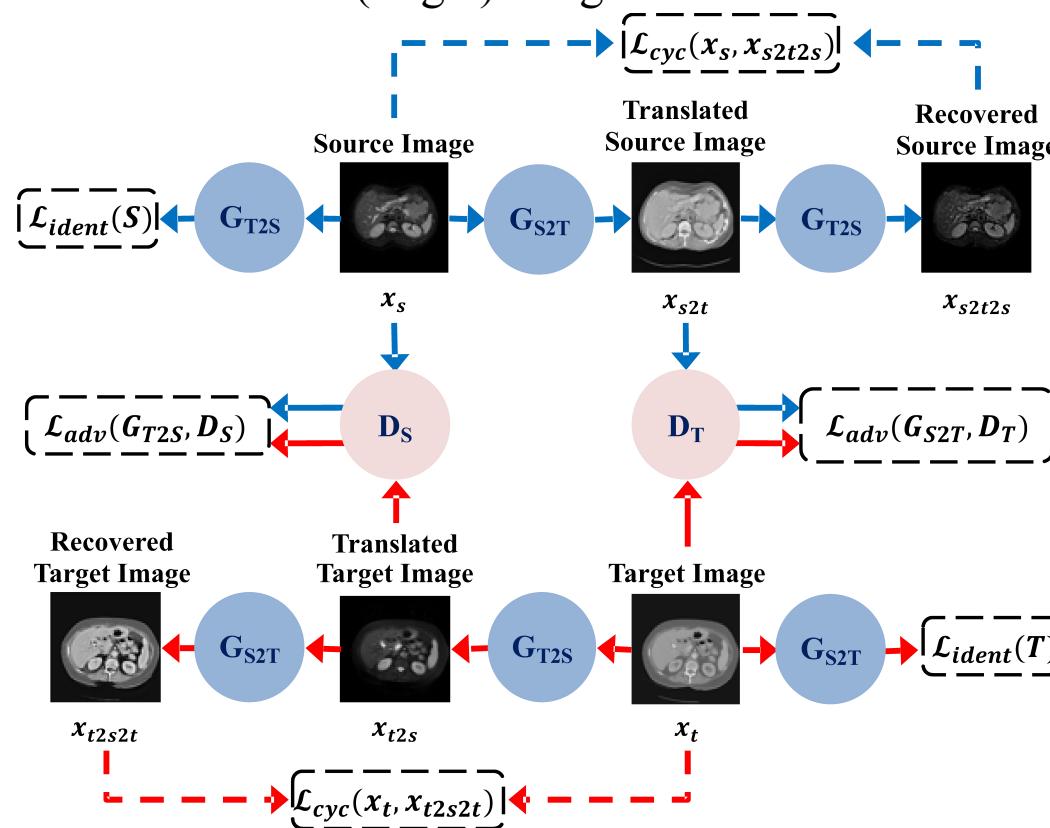
- There were two drawbacks that prevent the existing UDA methods from being directly applied, and our work optimized them in the two aspects.
 - ① The structural information of the objects may be lost during image alignment, due to the destruction of details caused by the down-sampling in the generators.
 - ② The domain-invariant representations extracted through direct adversarial training contain lots of information irrelevant to the task, which do not contribute to the actual downstream task.

Motivation

- There were two drawbacks that prevent the existing UDA methods from being directly applied, and our work optimized them in the two aspects.
- ① The structural information of the objects may be lost during image alignment, due to the destruction of details caused by the down-sampling in the generators.
→ Introducing the attention mechanism and enhancing the learning of objects.
- ② The domain-invariant representations extracted through direct adversarial training contain lots of information irrelevant to the task, which do not contribute to the actual downstream task.

Attention-aware Image Alignment in the Pixel Space

- Based on the CycleGAN^[1], we introduce the representation map $G_{T2S}(G_{S2T})$ to translate target (source) images into source (target) style-like images.
- The adversarial discriminator $D_S(D_T)$ is designed to distinguish the generated source (target) images and the real source (target) images.



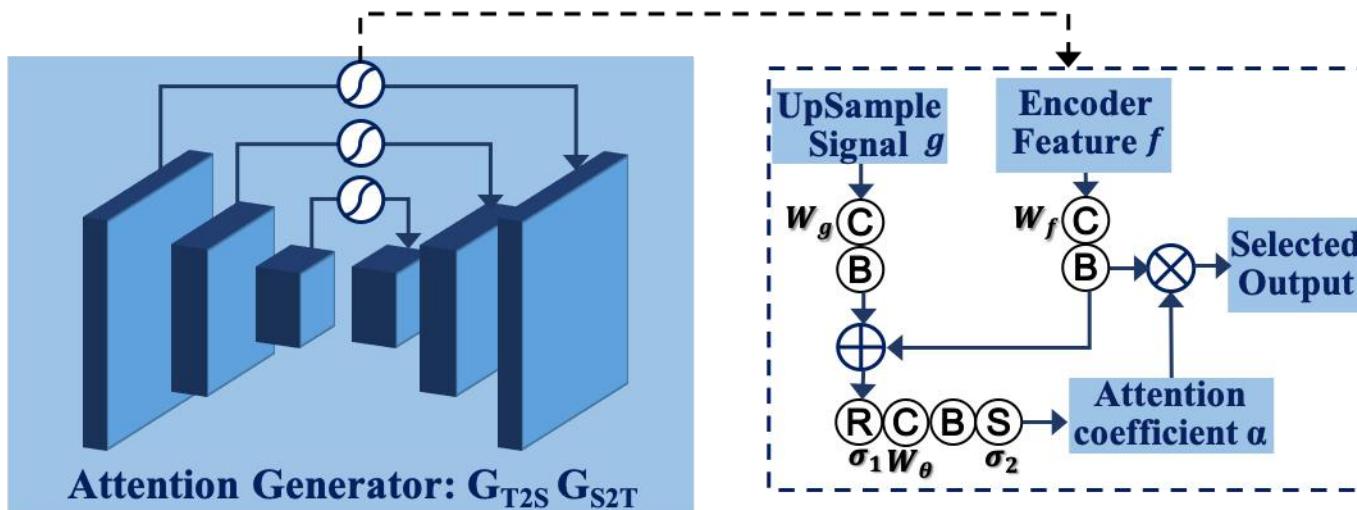
^[1] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." in IEEE international conference on computer vision (ICCV). 2017.

Attention-aware Image Alignment in the Pixel Space

- The AttENT bridges the attention gates between encoder and the decoder to enhance the learning toward task-related organs.
- The process of feature selection and the output of Attention Gate (AG) can be formulated as follows:

$$AG = \mathbf{f} \times \sigma_2 \{ W_\theta^T \times \sigma_1 \times [(W_f^T \times \mathbf{f} + b_f) + (W_g^T \times \mathbf{g} + b_g)] + b_\theta \}$$

where the f is encoder feature while the g means upsample signal. The (W, b) and σ represent the convolution and activation.



Attention-aware Image Alignment in the Pixel Space

- The above zero-sum game process can be represented by minimizing the following objectives:

$$\mathcal{L}_{adv} = \mathbb{E}_{x_s \sim \mathbb{X}_S} [\log D_S(x_s)] + \mathbb{E}_{x_t \sim \mathbb{X}_T} [\log (1 - D_S(G_{T2S}(x_t)))]$$

$$\mathcal{L}_{adv} = \mathbb{E}_{x_t \sim \mathbb{X}_T} [\log D_T(x_t)] + \mathbb{E}_{x_s \sim \mathbb{X}_S} [\log (1 - D_T(G_{S2T}(x_s)))]$$

- Besides, the cycle consistency loss function is also adopted to avoid contradiction between G_{T2S} and G_{S2T} in the conversion process, which is defined as follows:

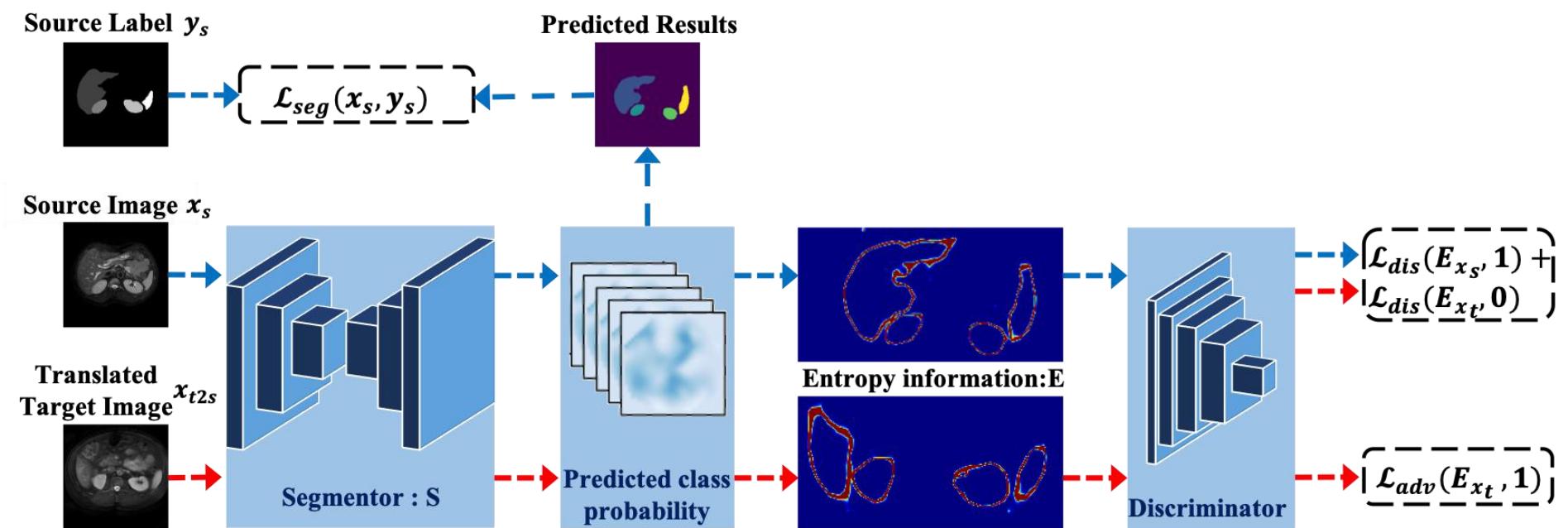
$$\mathcal{L}_{cyc} = \mathbb{E}_{x_s \sim \mathbb{X}_S} [\|G_{T2S}(G_{S2T}(x_s)) - x_s\|_1] + \mathbb{E}_{x_t \sim \mathbb{X}_T} [\|G_{S2T}(G_{T2S}(x_t)) - x_t\|_1]$$

Motivation

- There were two drawbacks that prevent the existing UDA methods from being directly applied, and our work optimized them in the two aspects.
- ① The structural information of the objects may be lost during image alignment, due to the destruction of details caused by the down-sampling in the generators.
→ Introducing the attention mechanism and enhancing the learning of objects.
- ② The domain-invariant representations extracted through direct adversarial training contain lots of information irrelevant to the task, which do not contribute to the actual downstream task.
→ Implementing the entropy information with segmentation

Adversarial Feature Alignment in the Entropy Space

- To reduce the entropy map discrepancy between target domain and the source domain by minimizing the adversarial entropy, and then alleviate domain discrepancy in entropy space.

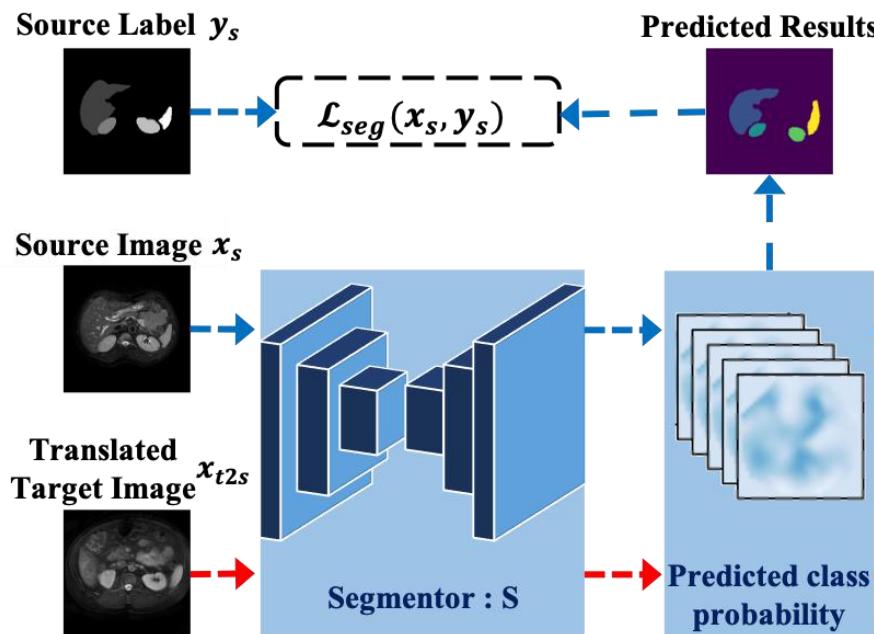


Adversarial Feature Alignment in the Entropy Space

- Firstly, we utilize the source image with annotation to train the source domain segmentor S in the supervised manner.

$$\mathcal{L}_{seg} = \sum_c^C [-y_s^c \log S(x_s)^c]$$

where C denotes the number of classes.

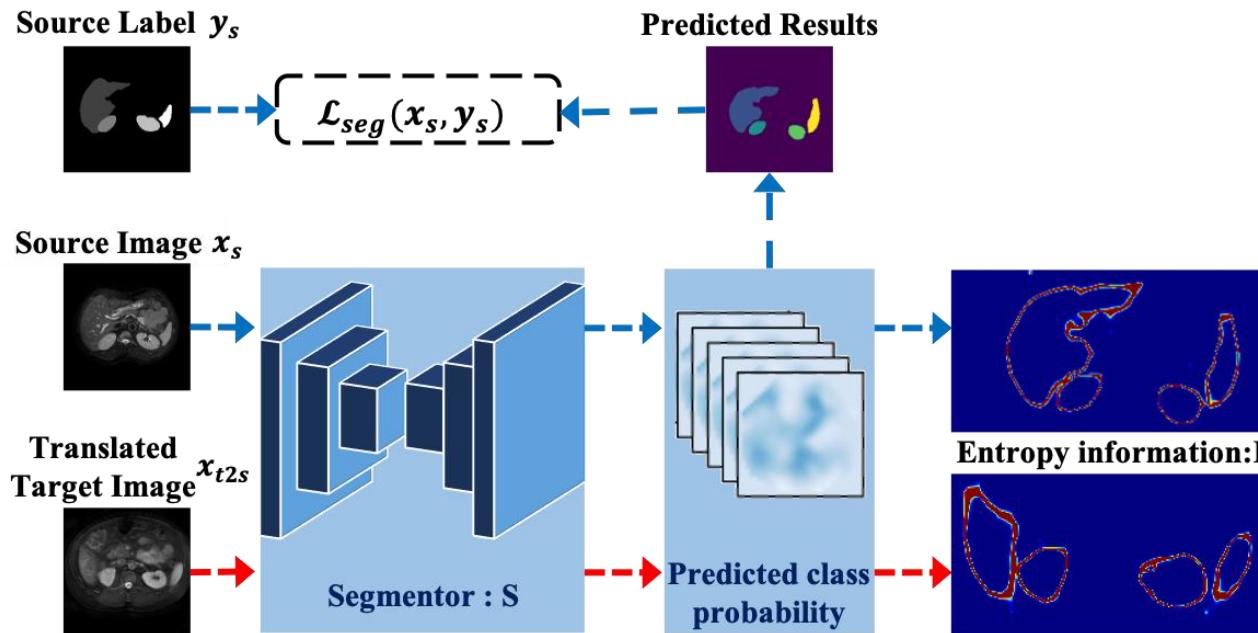


Adversarial Feature Alignment in the Entropy Space

- Secondly, we calculate the normalized entropy map E_{x_s} based on the predicted result $S(x_s)$. The same is true for the entropy map of translated target image.

$$E_{x_s} = -\frac{1}{\log C} \sum_c^C S(x_s)^c \log S(x_s)^c$$

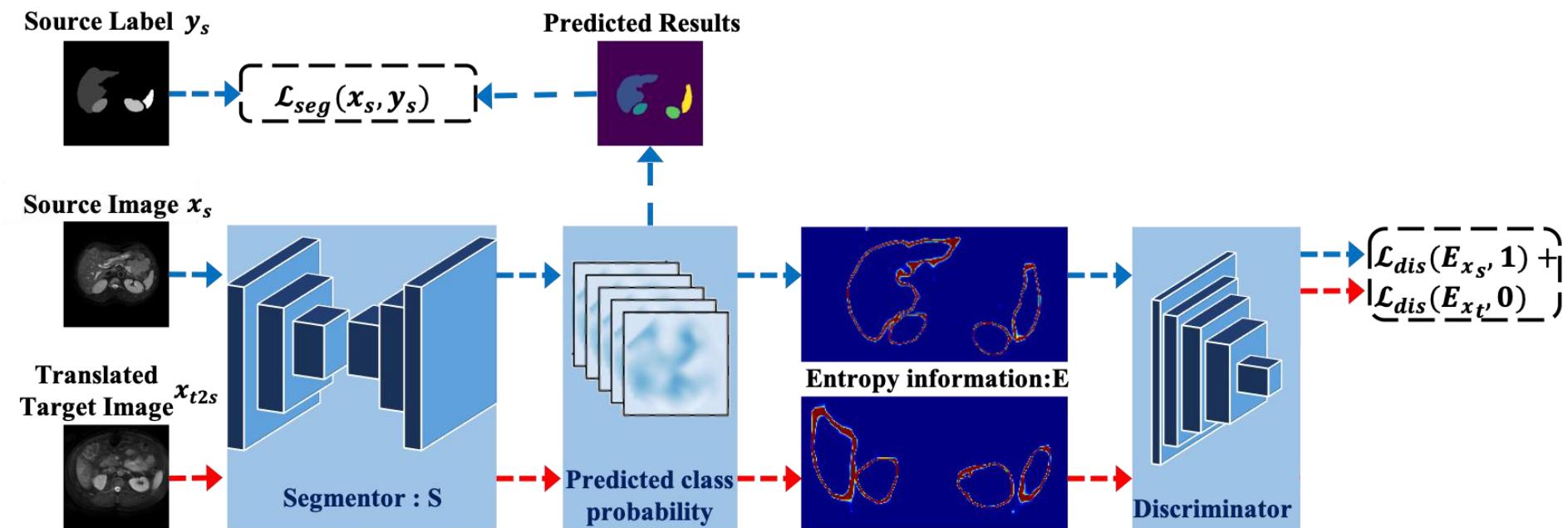
$$E_{x_t} = -\frac{1}{\log C} \sum_c^C S(x_{t2s})^c \log S(x_{t2s})^c$$



Adversarial Feature Alignment in the Entropy Space

- After that, the discriminator D_E is constructed to distinguish entropy map of the source image and the translated target image.
- D_E is trained by discriminative loss to minimize the features distribution between source domain and target domain in the entropy space.

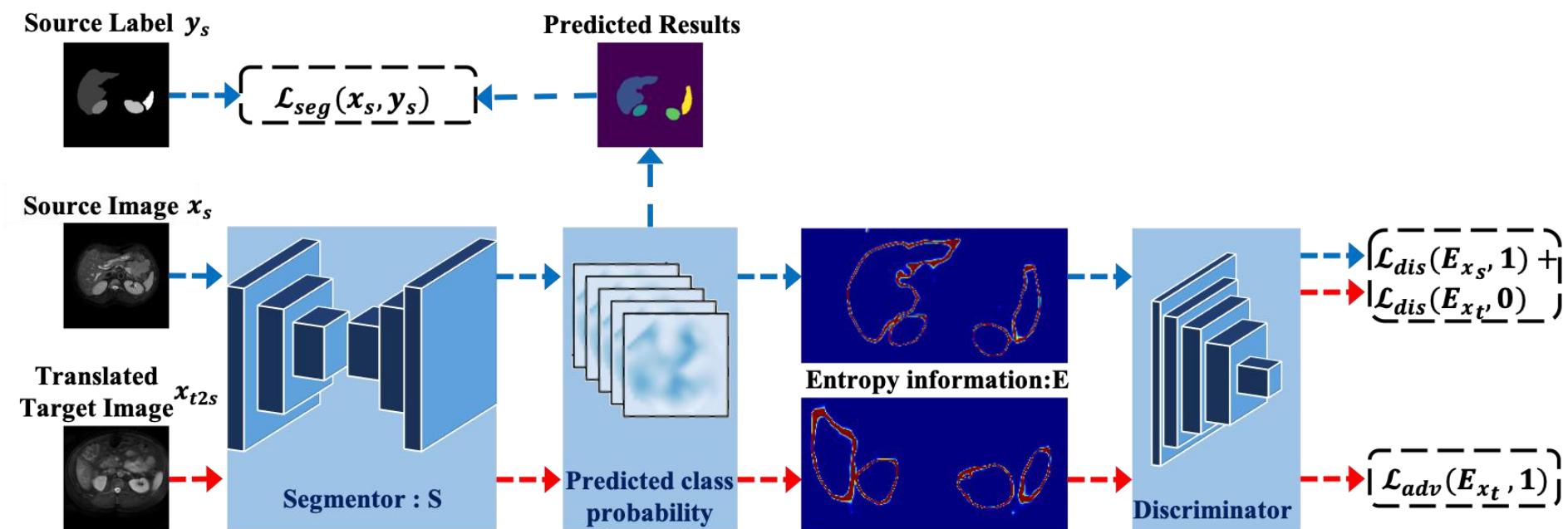
$$\mathcal{L}_{dis} = \sum_{x_s}^{\mathbb{X}_S} \mathcal{L}_{D_E}(E_{x_s}, 1) + \sum_{x_t}^{\mathbb{X}_T} \mathcal{L}_{D_E}(E_{x_t}, 0).$$



Adversarial Feature Alignment in the Entropy Space

- Finally, we use the adversarial loss \mathcal{L}_{adv} to train segmentor S again and recover the degraded performance due to domain shift, which is formulated as follows:

$$\mathcal{L}_{adv} = \sum_{x_t}^{\mathbb{X}_T} \mathcal{L}_{DE}(E_{x_t}, 1)$$



Dataset & Metric

- Combined (CT-MR) Healthy Abdominal Organ Segmentation (CHAOS)
 - 120 DICOM data sets from two different MRI sequences
 - supported by ISBI 2019
 - <https://chaos.grand-challenge.org/Data/>
- Multi-Atlas Labeling Beyond the Cranial Vault(MALBCV)
 - 30 labeled volumes from the CT training dataset.
 - supported by MICCAI 2015
 - <https://www.synapse.org/Synapse:syn3193805/wiki/217789>
- Evaluation Metrics:
 - Dice coefficient and Average symmetric surface distance (ASSD)

Ablation study & Significance tests

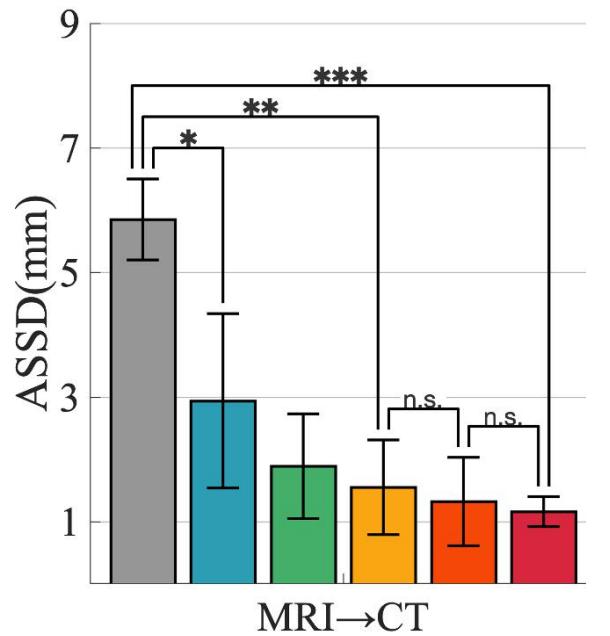
- Table reported the performance discrepancy between different components of AttENT on multi-modalities abdominal organs segmentation.

MRI \Rightarrow CT						
Methods ASSD (Mean \pm Std.)	W/o Adaptation (5.85 \pm 0.65)	CycleGAN (1.55 \pm 0.76)	Pixel Space Alignment (1.32 \pm 0.71)	Feature Space Alignment (2.94 \pm 1.40)	Entropy Space Alignment (1.89 \pm 0.84)	AttENT (1.16 \pm 0.24)
W/o Adaptation (5.85 \pm 0.65)	\	0.0018	\	0.0281	\	0.0003
CycleGAN (1.55 \pm 0.76)	0.0018	\	0.3827	\	\	\
Pixel Space Alignment (1.32 \pm 0.71)	\	0.3827	\	\	\	0.3895
Feature Space Alignment (2.94 \pm 1.40)	0.0281	\	\	\	0.2063	\
Entropy Space Alignment (1.89 \pm 0.84)	\	\	\	0.2063	\	0.1521
AttENT (1.16 \pm 0.24)	0.0003	\	0.3895	\	0.1521	\

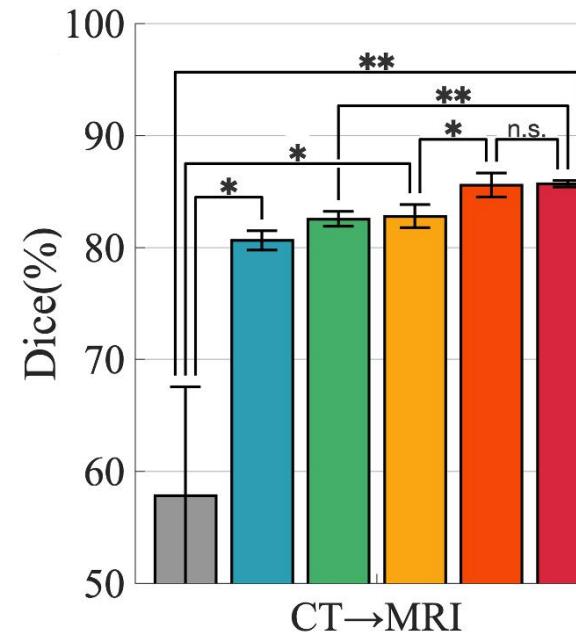
CT \Rightarrow MRI						
Methods Dice (Mean \pm Std.)	W/o adaptation (57.73 \pm 9.72)	CycleGAN (83.05 \pm 1.03)	Pixel Space Alignment (85.57 \pm 1.07)	Feature Space Alignment (80.62 \pm 0.86)	Entropy Space Alignment (82.54 \pm 0.67)	AttENT (85.67 \pm 0.29)
W/o Adaptation (57.73 \pm 9.72)	\	0.0113	\	0.0149	\	0.0078
CycleGAN (83.05 \pm 1.03)	0.0113	\	0.0283	\	\	\
Pixel Space Alignment (85.57 \pm 1.07)	\	0.0283	\	\	\	0.4524
Feature Space Alignment (80.62 \pm 0.86)	0.0149	\	\	\	0.0337	\
Entropy Space Alignment (82.54 \pm 0.67)	\	\	\	0.0337	\	0.0019
AttENT (85.67 \pm 0.29)	0.0078	\	0.4524	\	0.0019	\

Ablation study & Significance tests

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



■ W/o adaptation
 ■ Feature-space Alignment
 ■ Entropy-space Alignment
 n.s.: No Significance
 *: p-value<0.05



■ CycleGAN
 ■ Pixel-space Alignment
 ■ AttENT (Ours)
 **: p-value<0.01
 ***: p-value<0.001

Quantitatively measure the domain shift

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

Methods	MRI \Rightarrow CT									
	ASSD (voxel)					Dice (%)				
	Liver	Right Kidney	Left Kidney	Spleen	Avg	Liver	Right Kidney	Left Kidney	Spleen	Avg
Supervised	1.0	1.8	0.9	1.2	1.2	92.8	86.4	87.4	88.2	88.7

W/o Adapt	2.9	5.6	7.7	7.4	5.9	73.1	47.3	57.3	55.1	58.2
-----------	-----	-----	-----	-----	-----	------	------	------	------	------

Methods	CT \Rightarrow MRI									
	ASSD (voxel)					Dice (%)				
	Liver	Right Kidney	Left Kidney	Spleen	Avg	Liver	Right Kidney	Left Kidney	Spleen	Avg
Supervised	1.3	2.0	1.5	1.3	1.5	92.0	91.1	80.6	85.7	87.3

W/o Adapt	4.5	12.3	6.8	4.5	7.0	48.9	50.9	65.3	65.7	57.7
-----------	-----	------	-----	-----	-----	------	------	------	------	------

Comparison with the State-of-the-art Methods

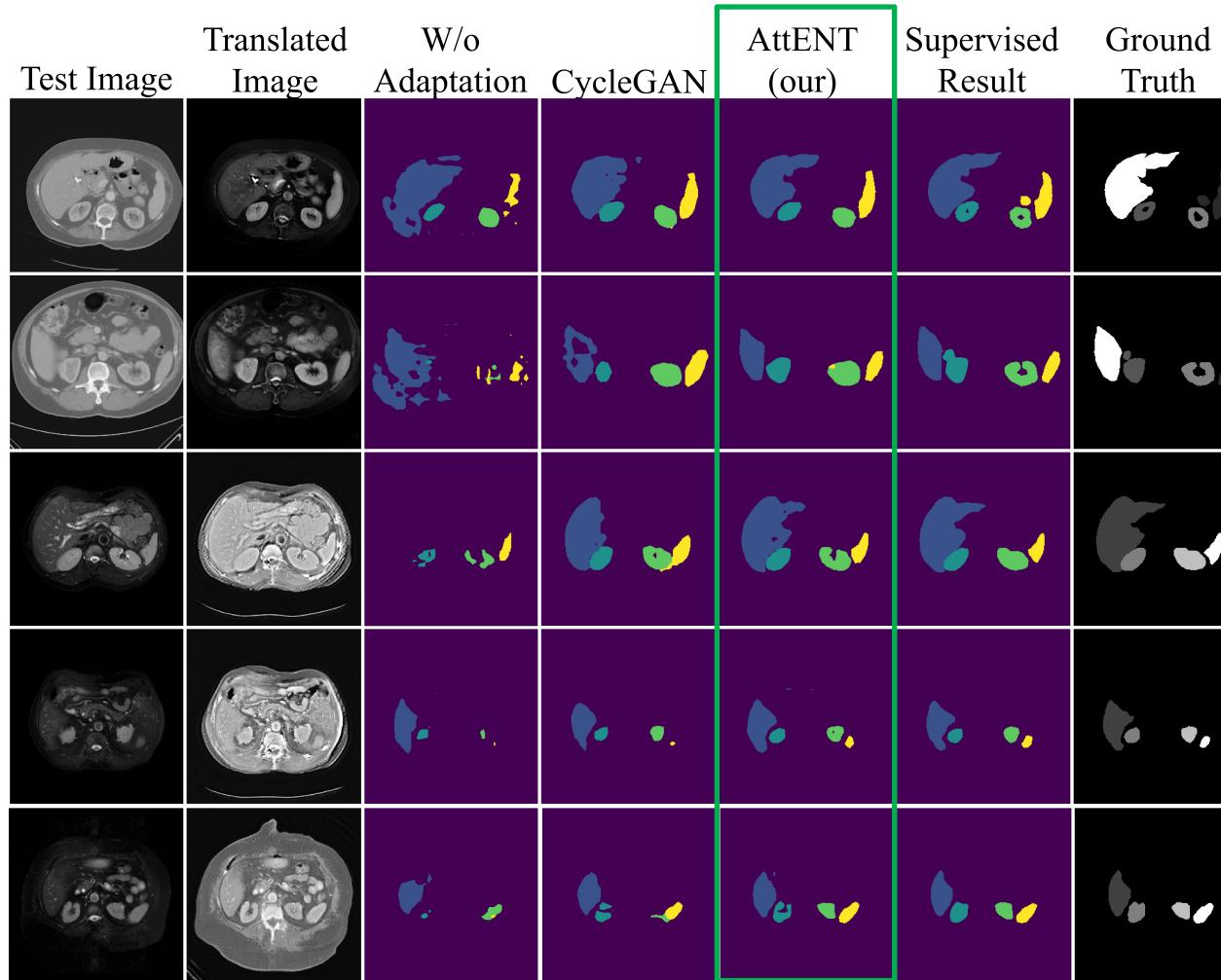
- The quantitative results of six SOTA UDA methods and proposed method demostate the superior of our work in domain-adaptive segmentation of abdominal organs.

Methods	MRI \Rightarrow CT									
	ASSD (voxel)					Dice (%)				
	Liver	Right Kidney	Left Kidney	Spleen	Avg	Liver	Right Kidney	Left Kidney	Spleen	Avg
Supervised	1.0	1.8	0.9	1.2	1.2	92.8	86.4	87.4	88.2	88.7
SynSegNet [21]	2.2	1.3	2.1	2.0	1.9	85.0	82.1	72.7	81.0	80.2
AdaOutput [20]	1.7	1.2	1.8	1.6	1.6	85.4	79.7	79.7	81.7	81.6
CycleGAN [10]	1.8	1.3	1.2	1.9	1.6	83.4	79.3	79.4	77.3	79.9
CyCADA [15]	2.6	1.4	1.3	1.9	1.8	84.5	78.6	80.3	76.9	80.1
SIFA-v1 [18]	2.1	1.1	1.6	1.8	1.6	87.9	83.7	80.1	80.5	83.1
SIFA-v2 [28]	1.2	1.0	1.5	1.6	1.3	88.0	83.3	80.9	82.6	83.7
AttENT	0.68	1.31	1.43	1.21	1.16	88.56	80.66	85.59	86.34	85.29
W/o Adapt	2.9	5.6	7.7	7.4	5.9	73.1	47.3	57.3	55.1	58.2

Methods	CT \Rightarrow MRI									
	ASSD (voxel)					Dice (%)				
	Liver	Right Kidney	Left Kidney	Spleen	Avg	Liver	Right Kidney	Left Kidney	Spleen	Avg
Supervised	1.3	2.0	1.5	1.3	1.5	92.0	91.1	80.6	85.7	87.3
SynSegNet [21]	2.8	0.7	4.8	2.5	2.7	87.2	90.2	76.6	79.6	83.4
AdaOutput [29]	1.9	1.4	3.0	1.8	2.1	85.8	89.7	76.3	82.2	83.5
CycleGAN [10]	2.0	3.2	1.9	2.6	2.4	88.8	87.3	76.8	79.4	83.1
CyCADA [15]	1.5	1.7	1.3	1.6	1.5	88.7	89.3	78.1	80.2	84.1
SIFA-v1 [18]	2.3	0.9	1.4	2.4	1.7	88.5	90.0	79.7	81.3	84.9
SIFA-v2 [28]	1.5	0.6	1.5	2.4	1.5	90.0	89.1	80.2	82.3	85.4
AttENT	0.99	1.03	1.26	1.12	1.10	91.05	81.38	80.51	89.75	85.67
W/o Adapt	4.5	12.3	6.8	4.5	7.0	48.9	50.9	65.3	65.7	57.7

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 proposed a novel unsupervised domain-adaptive framework to recover performance degradation from the domain shift in cross-modality medical image segmentation.
 - Our framework is able to improve the image alignment by introducing the attention mechanism into CycleGAN in the pixel space.
 - The principle of entropy minimization is utilized to align features in the entropy space.

Summary

- We proposed a novel unsupervised domain-adaptive framework to recover performance degradation from the domain shift in cross-modality medical image segmentation.
 - Our framework is able to improve the image alignment by introducing the attention mechanism into CycleGAN in the pixel space.
 - The principle of entropy minimization is utilized to align features in the entropy space.
- We would like to thank:
 - National Key Research and Development Program of China (No. 2018YFB0204301)
- Related material:
 - <https://github.com/lichen14/AttENT>
- We are grateful for corrections and discussions!



国防科技大学

National University of Defense Technology



AttENT:

Domain-Adaptive Medical Image Segmentation

via Attention-Aware Translation and

Adversarial Entropy Minimization

Thank You!