Medical Vision Seminar

——Wei Lou

2022_01_12

(MICCAI2021)Modality-aware Mutual Learning for Multi-modal Medical Image Segmentation

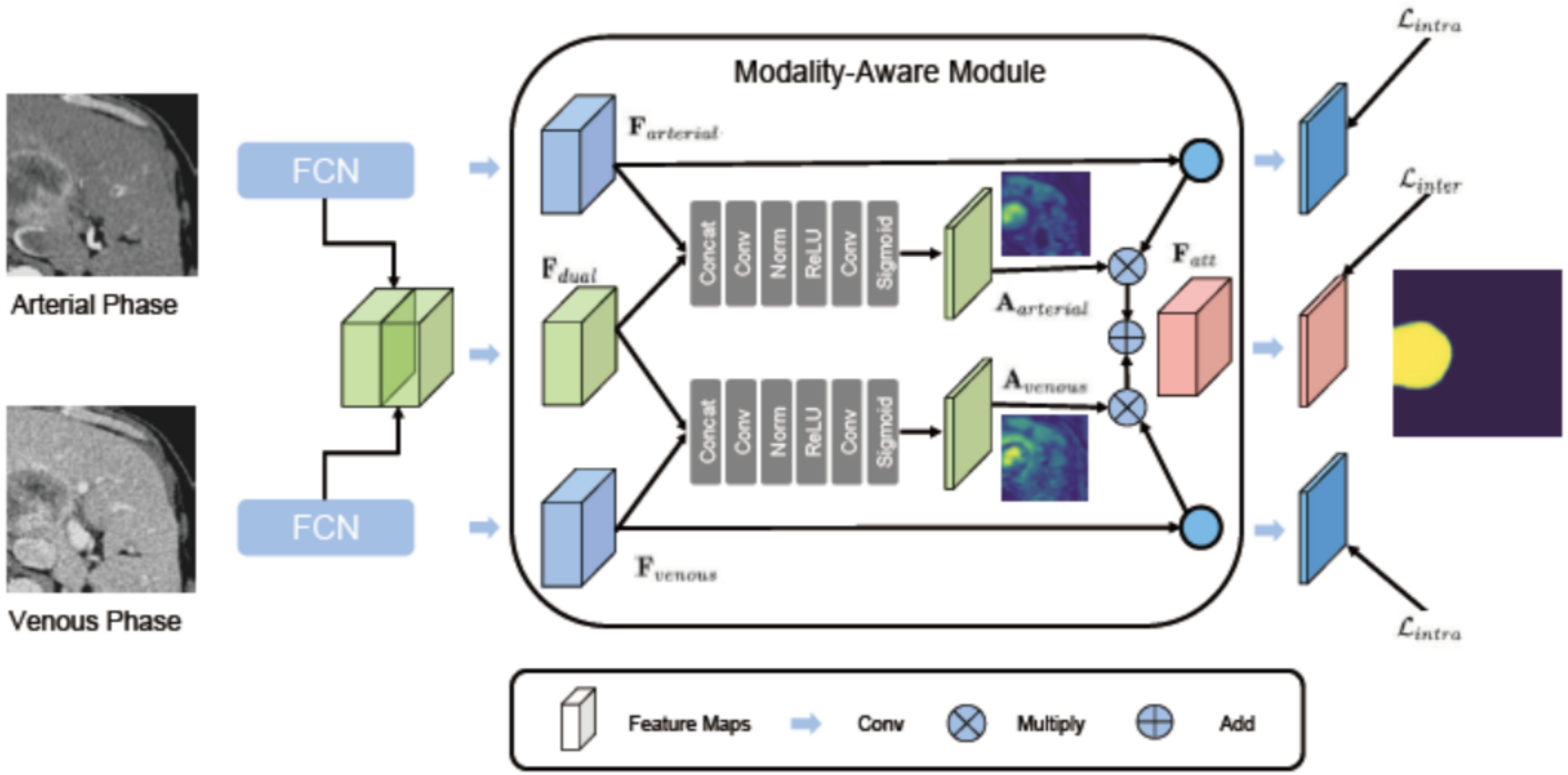
—— Yao Zhang, Jiawei Yang, Jiang Tian, Zhongchao Shi, Cheng Zhong,
Yang Zhang, and Zhiqiang He

1. Motivation

There exists two major issues applying FCNs in multi-modal segmentation.

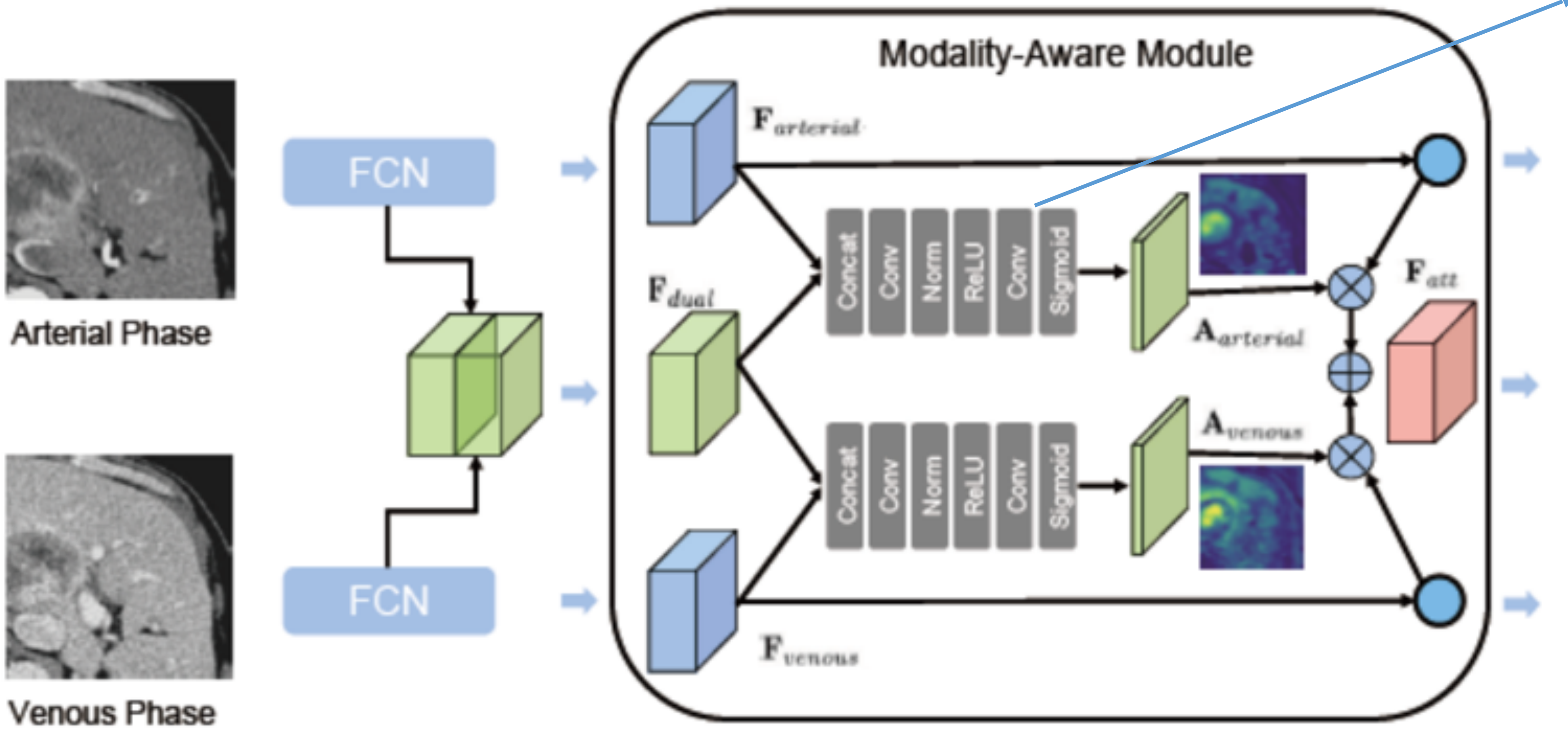
- How to integrate information from multi-modal medical images effectively.
- How to deal with the scenario of missing modalities that is common in practice.

2. Methods (Liver CT)



2.1 Modality-Aware Module

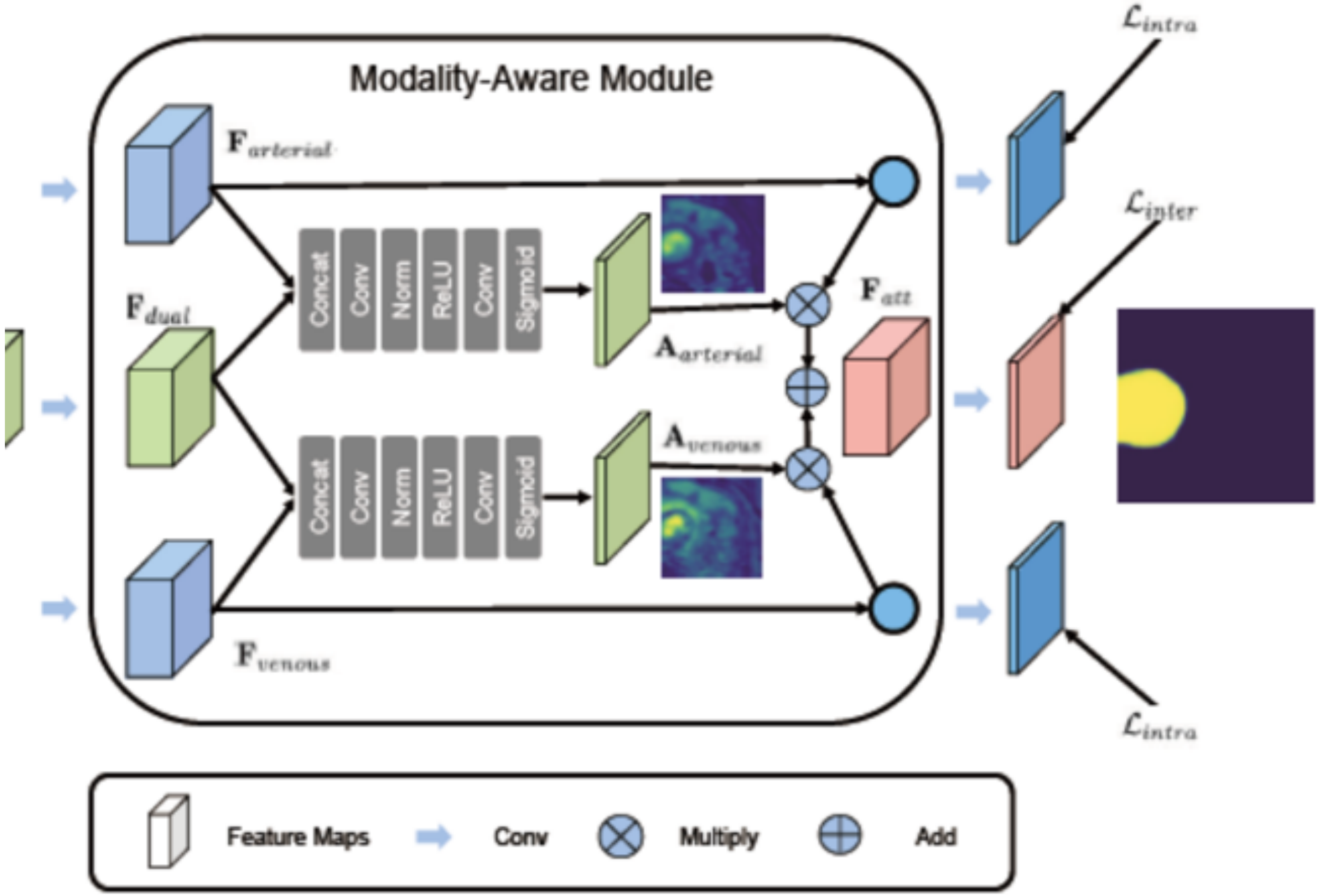
- Cross-model attention to highlight the target features.
- Although F_{dual} encodes both phases, it introduces noise from each modality. So the MA module is to adaptively measure each phase's contribution.



Conv 3*3*3 + conv 1*1*1

$$\mathbf{F}_{att} = \sum_{i \in \{AP, VP\}} \mathbf{A}_i * \mathbf{F}_i.$$

2.2 Mutual Learning Strategy



- Each modality-specific model interacts as a teacher and a student mutually.
- The venous model not only learn from the venous phase but also learn from arterial model.

Achieved by intra-phase losses and a joint loss.

$$\mathcal{L} = \lambda \sum_{i \in \{AP, VP\}} \mathcal{L}_{intra}(Y|X_i; W_i) + \mathcal{L}_{joint}(Y|X; W),$$

4. Experiments

4.1 Dataset

1. 654 contrast-enhanced Liver CT volume with arterial and venous phases obtained from Chinese PLA General Hospital.
2. BraTS 2018 dataset contains MR scans from 285 patients with four modalities: T1, T2, T1 contrasted-enhanced (T1ce) and Flair.

Experimental setting:

- Truncate the raw intensity values within the range 0:5%-99:5% of the initial HU value and normalize each raw CT case to have zero mean and unit variance.
- Input size: 128*128*128

4.2 Effectiveness of Multi-modal Modeling.

Table 1. Results on multi-modal liver tumor segmentation. Best results are highlighted with bold.

Methods	Dice [%] ↑	ASSD [voxel] ↓
nnUNet [8]	78.76 ± 18.91	8.02 ± 20.21
OctopusNet [3]	78.89 ± 18.65	12.67 ± 42.43
MS+Ensemble	78.96 ± 19.37	5.88 ± 10.73
MS+MA	80.98 ± 18.58	5.38 ± 9.20
MAML	**81.25 ± 17.02**	**4.71 ± 6.13**

MS+Ensemble: Straightforward average of the outputs of modality-specific models.

MS+MA: No mutual learning, only use F_{atten}.

4.3 Interpretable Fusion.

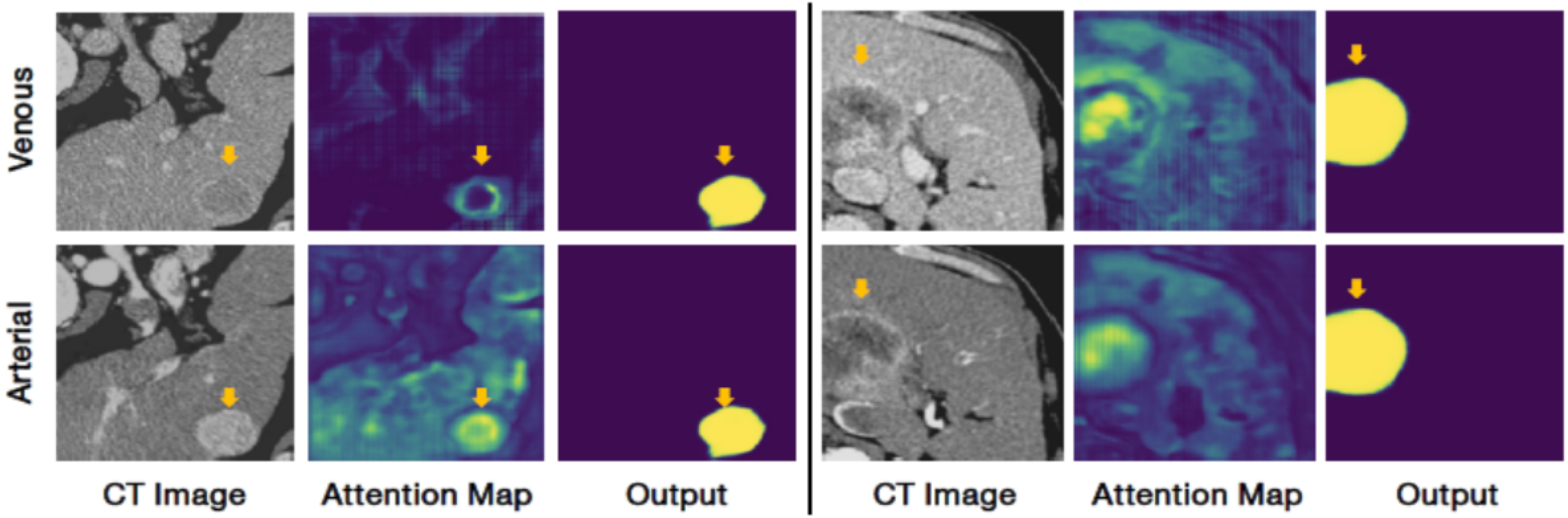


Fig. 2. Attention maps produced by Modality-Aware Module are able to capture enhanced part (left) as well as bleeding part and pseudo capsule (right) of the tumor.

4.4 Handling Missing Modalities.

Table 2. Results on handling missing modalities for liver tumor segmentation. Best results are highlighted with bold.

Methods		Dice [%] ↑	ASSD [voxel] ↓
Arterial Phase	nnUNet [8]	71.21 ± 25.87	9.51 ± 28.34
	MAML	$\mathbf{79.55 \pm 19.06}$	$\mathbf{6.38 \pm 12.00}$
Venous Phase	nnUNet [8]	75.10 ± 20.65	9.26 ± 30.82
	MAML	$\mathbf{79.81 \pm 18.42}$	$\mathbf{6.35 \pm 12.03}$

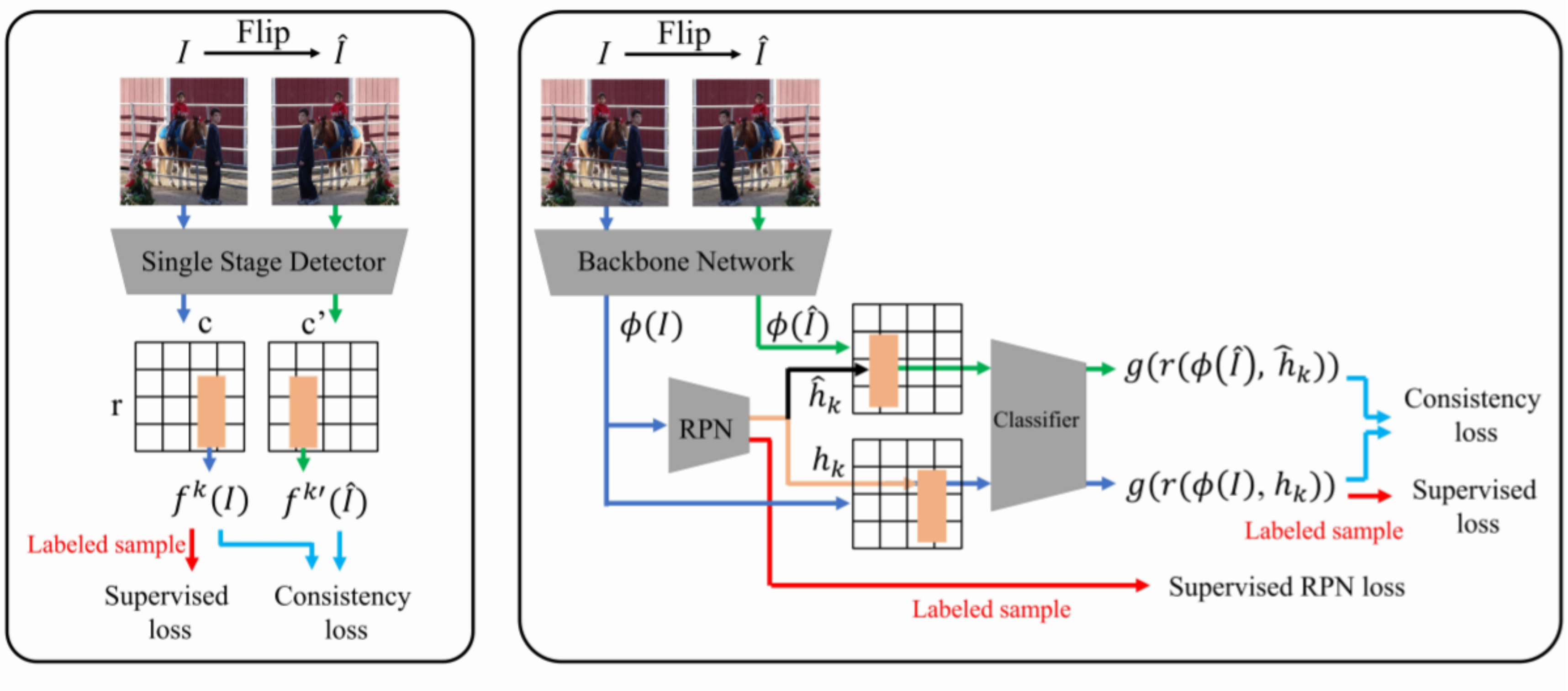
Methods	Enhanced Tumor	Tumor Core	Whole Tumor
HeMIS [6]	60.8	58.5	58.5
U-HVED [4]	65.5	66.7	62.4
KD-Net [7]	71.67 ± 1.22	81.45 ± 1.25	76.98 ± 1.54
MAML	$\mathbf{73.42 \pm 1.10}$	$\mathbf{83.36 \pm 1.23}$	$\mathbf{78.32 \pm 1.41}$

（AAAI2021） Semi-supervised Medical Image Segmentation through Dual-task Consistency

—— Xiangde Luo, Jieneng Chen, Tao Song, Guotai Wang

1. Motivation

Can we explicitly build task-level regularization rather than implicitly constructing networks and/or data-level perturbation and then regularization for SSL?

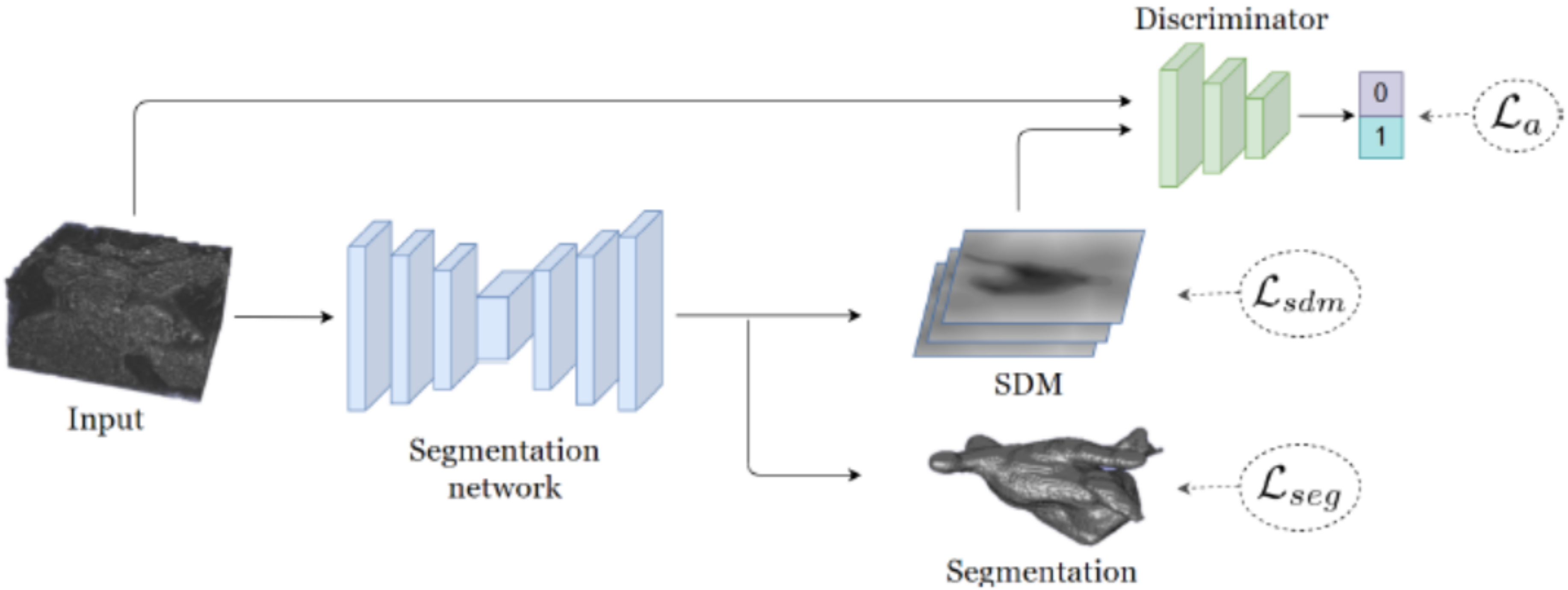


(a) single stage detector (b) two stage detector

Reprinted from Consistency-based Semi-supervised Learning for Object Detection (NIPS2019)

2. Related work – SASSnet

The SDM(signed distance map) assigns each pixel a value indicating its signed distance to the nearest boundary of target object, which provides a shape-aware representation that encodes richer features of object shape and surface.
Design a discriminator to regularize the network training.



2. Related work – Level set function

Level-set function is a traditional task that captures geometric active contours and distance information. (Segmentation map -> Level set map)

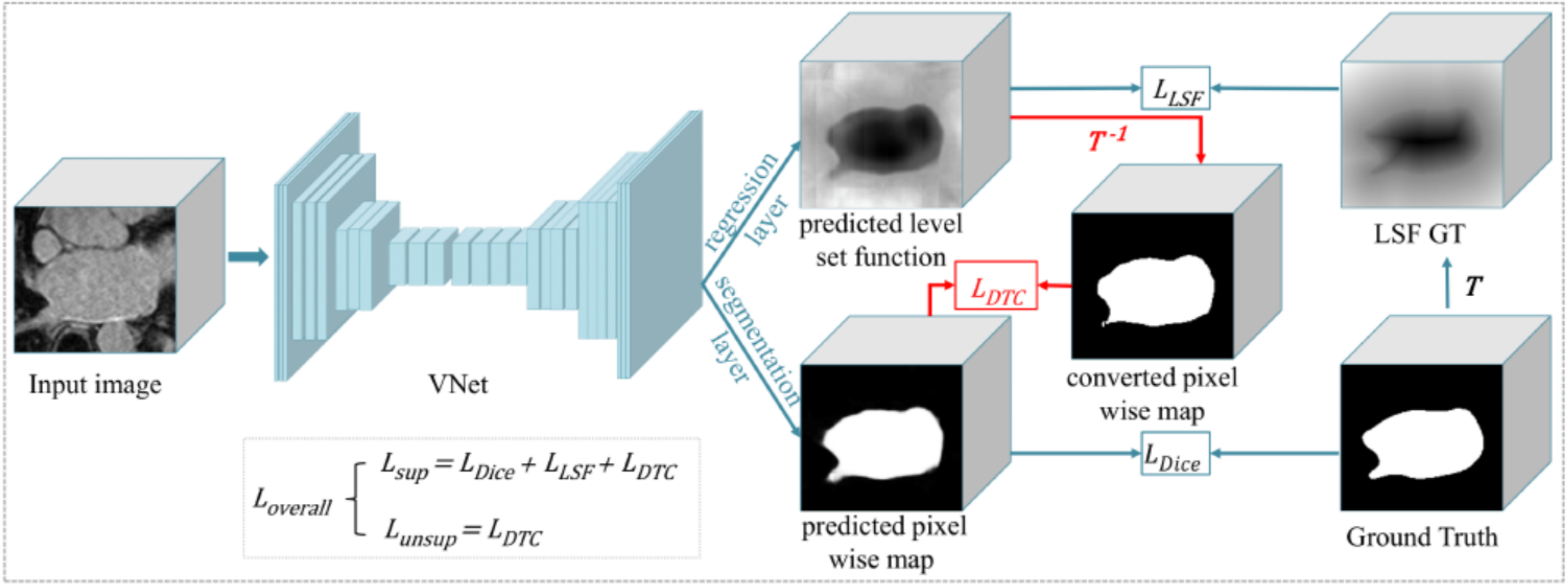
$$\mathcal{T}(x) = \begin{cases} -\inf\limits_{y \in \partial\mathcal{S}} \|x - y\|_2, & x \in \mathcal{S}_{\text{in}} \\ 0, & x \in \partial\mathcal{S} \\ +\inf\limits_{y \in \partial\mathcal{S}} \|x - y\|_2, & x \in \mathcal{S}_{\text{out}} \end{cases}$$

δS: the contour of the target object.

S_{in}, S_{out}: inside and outside region of the target object.

3. **Methods** (task-level regularization)

Two tasks: 1. pixel-wise classification head; 2. level set function regression head
Supervised: $L_{Dice}, L_{LSF}, L_{DTC}$. Dice loss, Level set function; Dual task consistency;
Unsupervised: L_{DTC}



3.1 Dual-task Consistency

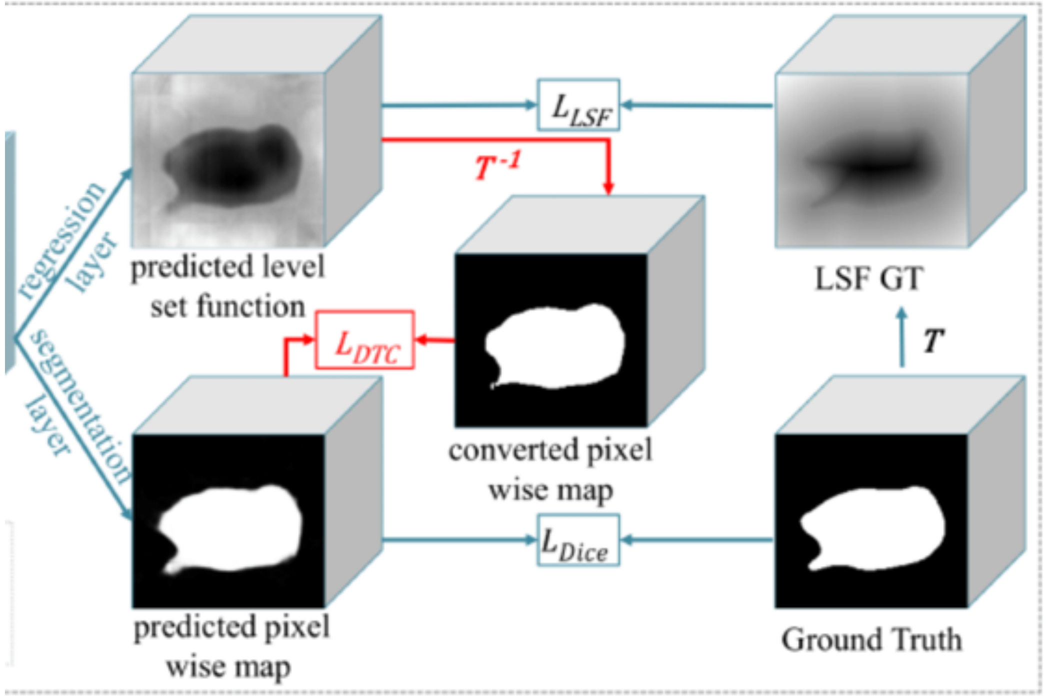
- **Smooth approximation to the inverse transform of level-set function** (Level set map -> Segmentation map)

$$\mathcal{T}^{-1}(z) = \frac{1}{1+e^{-k\cdot z}} = \sigma(k\cdot z)$$

z: level set value; k scale factor

- **Dual-task consistency loss (Pixel level reasoning ; Geometric structure information)**

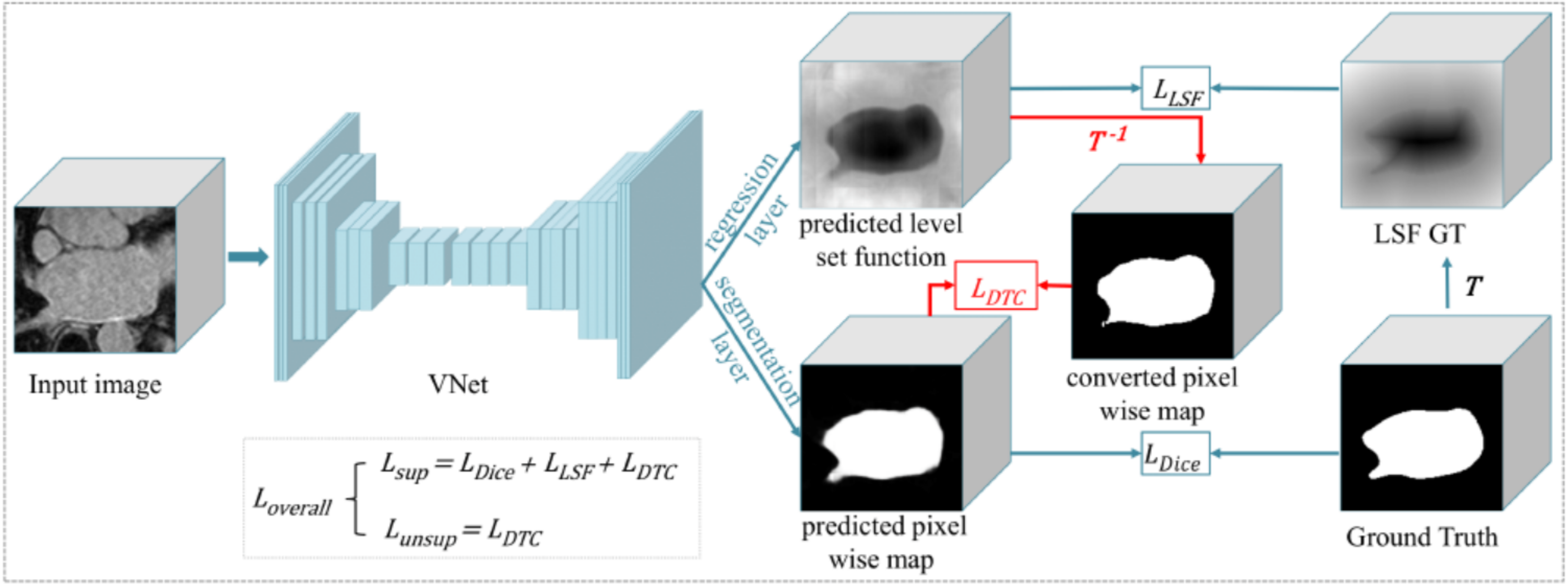
$$\begin{aligned}\mathcal{L}_{DTC}(\mathrm{x}) &= \sum_{\mathrm{x}_i\in\mathcal{D}} \left\|f_1(\mathrm{x}_i) - \mathcal{T}^{-1}\left(f_2(\mathrm{x}_i)\right)\right\|^2 \\ &= \sum_{\mathrm{x}_i\in\mathcal{D}} \left\|f_1(\mathrm{x}_i) - \sigma\left(k\cdot f_2(\mathrm{x}_i)\right)\right\|^2\end{aligned}$$



3.2 Semi-supervised training through Dual-Task-Consistency

3.2.1 For labelled data:

$$\mathcal{L}_{total} = \mathcal{L}_{Seg} + \mathcal{L}_{LSF} + \lambda_d \mathcal{L}_{DTC}$$



4. Experiments

Datasets: atrial(心室) dataset (100 3D gadolinium-enhanced MR images); pancreas(胰腺) dataset(82 abdomen CT images)

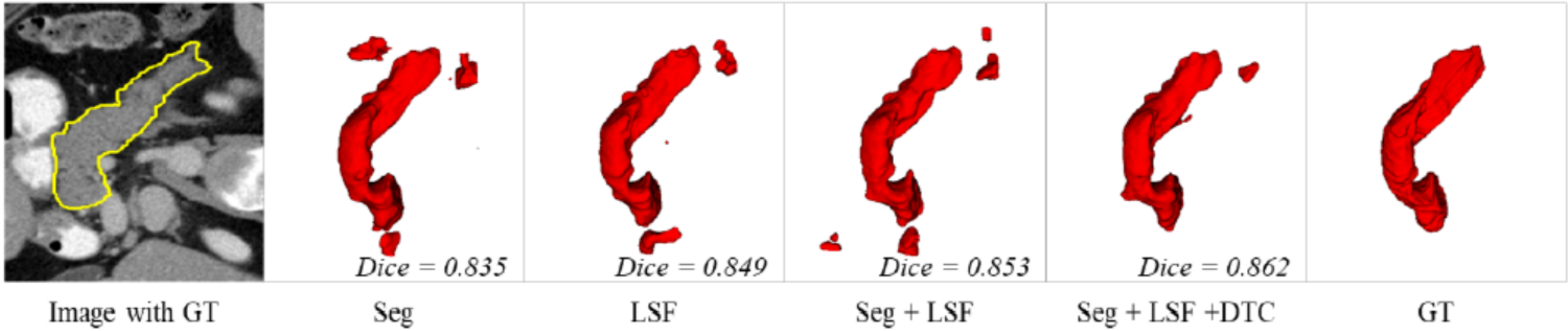


Figure 2: 3D Visualization of different training methods for pancreas segmentation. 12 annotated images without unannotated images were used for training. GT: ground truth. (best viewed in color)

4.1 Compare with SOTA

Method	Scans used		Metrics				Cost	
	Labeled	Unlabeled	Dice (%)	Jaccard (%)	ASD (voxel)	95HD (voxel)	Params (M)	Training time (h)
VNet	12	0	70.63	56.72	6.29	22.54	9.44	**2.1**
VNet	62	0	81.78	69.65	1.34	5.13	9.44	2.3
MT (NeurIPS'17)	12	50	75.85	61.98	3.40	12.59	9.44	2.9
DAN (MICCAI'17)	12	50	76.74	63.29	2.97	11.13	12.09	3.3
Entropy Mini (CVPR'19)	12	50	75.31	61.73	3.88	11.72	9.44	**2.2**
UA-MT (MICCAI'19)	12	50	77.26	63.82	3.06	11.90	9.44	3.9
CCT (CVPR'20)	12	50	76.58	62.76	3.69	12.92	15.65	4.1
SASSNet (MICCAI'20)	12	50	77.66	64.08	3.05	10.93	20.46	3.9
Ours	12	50	**78.27**	**64.75**	**2.25**	**8.36**	**9.44**	2.5

Table 2: Quantitative comparison between our methods and other semi-supervised methods on the Pancreas CT dataset. The first and second row are our fully supervised baseline, the last row is our proposed method, others are previous methods.

4.2 Ablation study

Method	Scans used		Metrics				Cost	
	Labeled	Unlabeled	Dice (%)	Jaccard (%)	ASD (voxel)	95HD (voxel)	Params (M)	Training time (h)
Seg	12	0	70.63	56.72	6.29	22.54	9.44	**2.1**
LSF	12	0	71.78	57.55	6.31	20.74	9.44	2.1
Seg + LSF	12	0	73.08	58.65	4.47	18.04	9.44	2.2
Seg + LSF + DTC	12	0	**74.84**	**60.78**	**2.17**	**9.34**	9.44	2.3
Seg	62	0	81.78	69.65	1.34	5.13	9.44	**2.3**
LSF	62	0	82.25	70.23	**1.18**	5.19	9.44	2.5
Seg + LSF	62	0	82.46	70.61	1.22	4.97	9.44	2.5
Seg + LSF + DTC	62	0	**82.80**	**71.05**	1.45	**4.67**	9.44	2.5

Table 1: Ablation study of our dual-task consistency method on the Pancreas CT dataset.