Semi-supervised Medical Image Segmentation via Dual-task Consistency

Xiangde Luo

University of Electronic Science and Technology of China

Background

Why need semi-supervised learning?

Problems

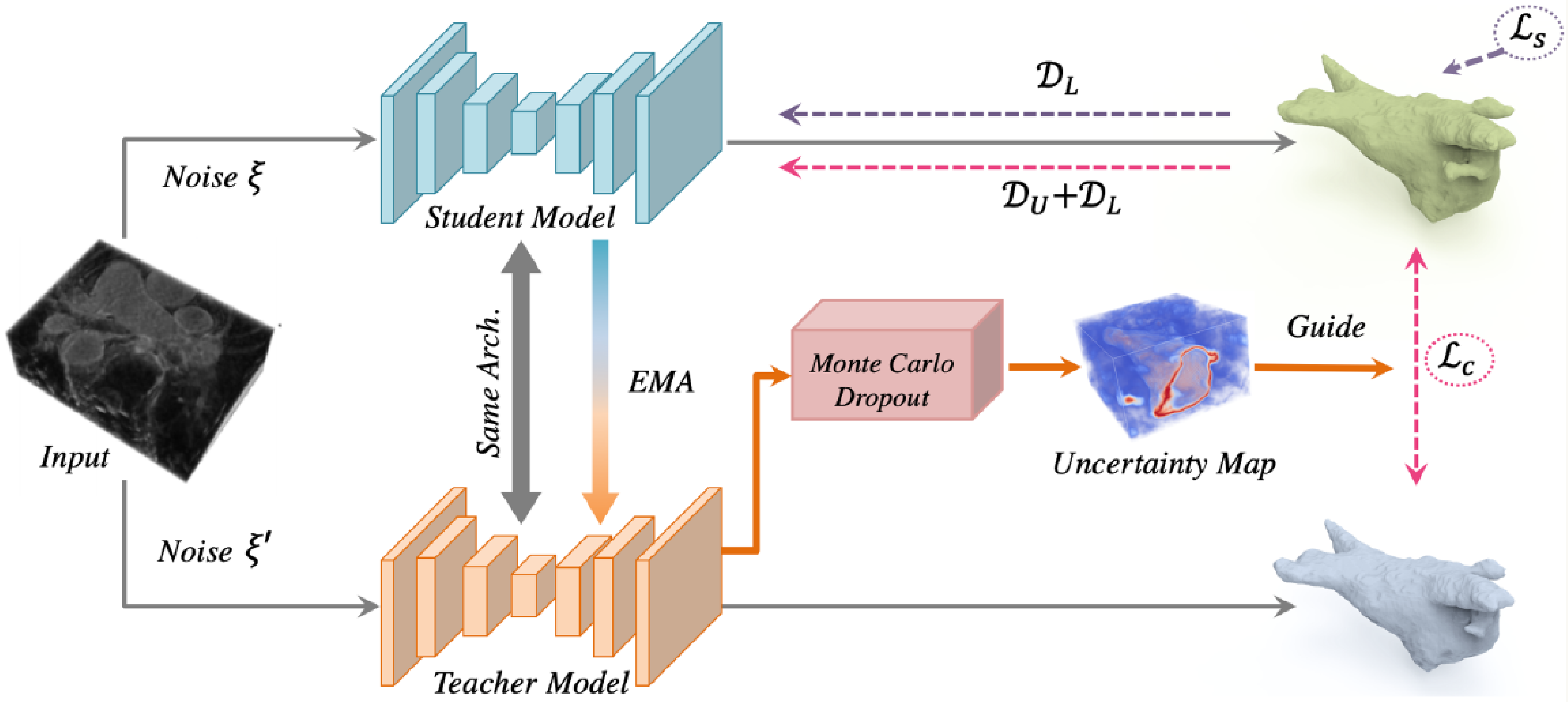
- deep learning with Convolutional Neural Networks (CNNs) highly depends on the availability of a large set of training images with manual annotations given by experts and lacks of generalization.
- a large set of manual annotations is expensive and labor-intensive, which has become the main obstacle for developing deep leaning models for medical image segmentation tasks.

Methods

- weakly-supervised methods(e.g. image-level labels, sparse pixel-level annotations, bounding boxes)
- **semi-supervised methods (learning from limited labeled data and larger unlabeled data)**
- Self-supervised learning (using pretext tasks to learn in a fully-supervised manner)
- Intelligent interactive segmentation tools (combining user interaction with deep neural network to perform image segmentation interactively)

Background

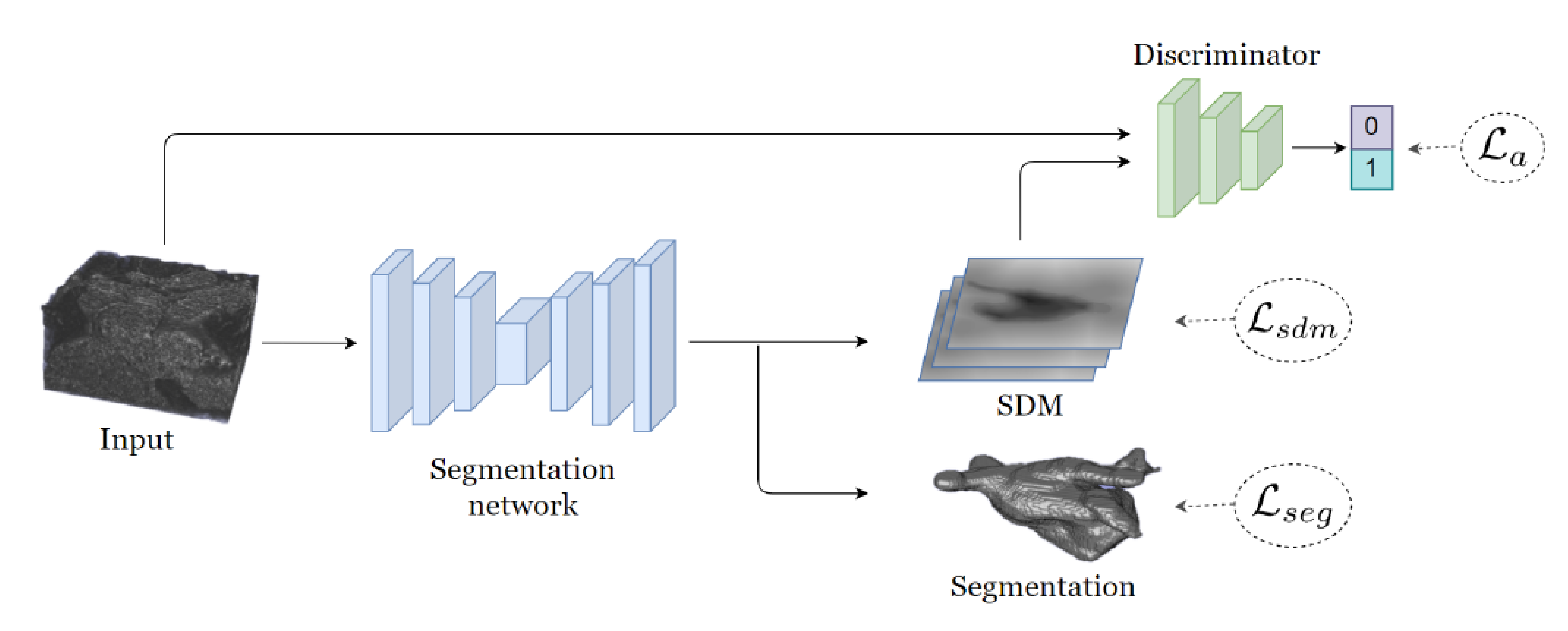
Existing-methods (1): Mean-teacher model[1]



[1] Yu et al. Uncertainty-aware Self-ensembling Model for Semi-supervised 3D Left Atrium Segmentation. In MICCAI 2019

Background

Existing-methods (2): Shape-aware semi-supervised model[2]



[2] Li et al. Shape-aware Semi-supervised 3D Semantic Segmentation for Medical Images. In MICCAI 2020

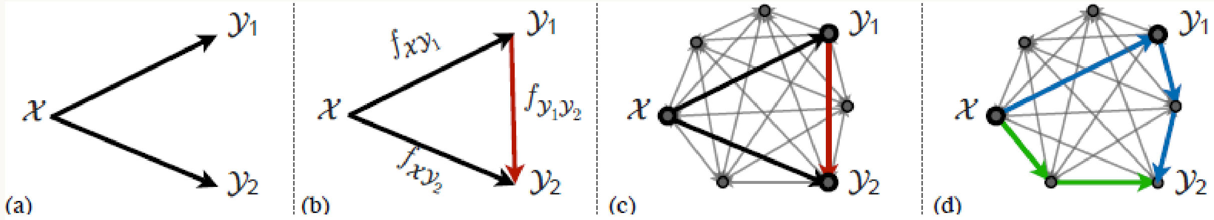
Background

Limitations of existing methods

- Computational-Cost (Training and Testing)
  - Time-consuming (performing forward many times)
  - GPU-consuming (computation graph needs to retain)
- Network
  - Very complex (one encoder, many decoders, or many encoders-decoders)
  - Parameters, not efficient
- Consistency[3]
  - data-level (transformation/augmentation)
  - model-level (perturbation/dropout)
  - task level?

Implicitly modeling

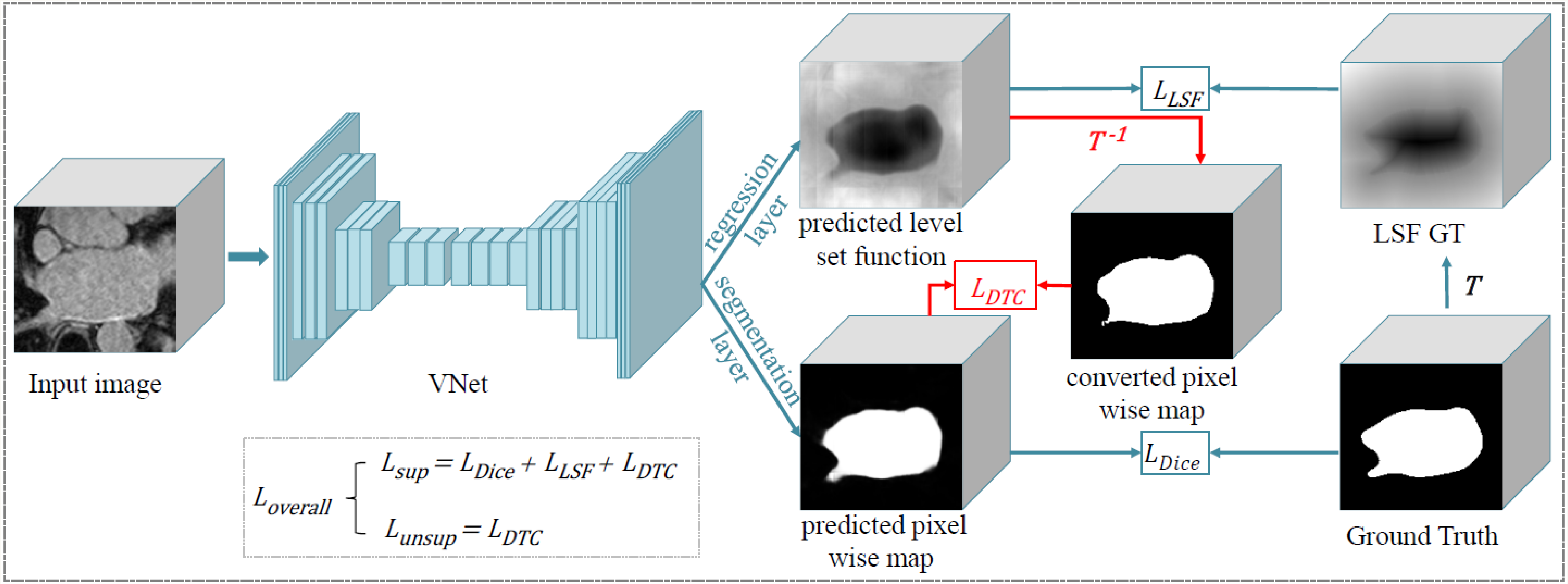
Explicitly modeling



[3] Zamir et al. Robust Learning Through Cross-Task Consistency. In CVPR 2020

Method

Overview of the proposed dual-task consistency framework



Method (1)

Dual-task-consistency

- Level set function
  - Geometric constraints and shape prior
  - Easy to implement

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

- Task transformation
  - Level set function to segmentation
  - Differentiable way

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

$$\frac{\partial \mathcal{T}^{-1}}{\mathrm{d}z} = \left(\frac{1}{1 + e^{-k \cdot z}}\right)' = k \cdot \frac{1}{1 + e^{-kz}} \cdot \left(1 - \frac{1}{1 + e^{-kz}}\right)$$

- Dual-task-consistency loss
  - Regularization of supervised learning
  - Learning from unlabeled data
  - Embedding shape and geometric constraints for semi-supervised learning

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

Method (2)

Semi-supervised learning through DTC

Algorithm 1 Semi-supervised training through Dual-task consistency,

Input: $\mathbf{x}_i \in \mathcal{D}_l + \mathcal{D}_n$, $\mathbf{y}_i \in \mathcal{D}_l$

Output: Dual-task model's parameter θ_1 for segmentation head, θ_2 for level-set function (LSF) head and θ for shared-weights backbone network

$f_1(x)$ = segmentation task branch with shared parameter θ and segmentation head's parameter θ_1

$f_2(x)$ = LSF task branch with shared parameter θ and LSF head's parameter θ_2

while stopping criterion not met: **do**

- Sample batch $b_l = (\mathbf{x}_i, \mathbf{y}_i) \in \mathcal{D}_l$ and $b = b_l + b_u$, where $b_u = \mathbf{x}_i \in \mathcal{D}_u$
- Generating LSF ground truth $\mathcal{T}(\mathbf{y}_i)$ according to Equation. 1
- Computing dual-task predictions $f_1(\mathbf{x}_i)$ and $f_2(\mathbf{x}_i)$, $i \in \{1, ..., N\}$ where N denotes the batch size
- Applying task transform layer $\mathcal{T}^{-1}(f_2(\mathbf{x}_i))$ according to Equation. 2
- $\mathcal{L}_{DTC}(\mathbf{x}) = \frac{1}{|b|}\sum_{\mathbf{x}_i \in b} \left\| f_1(\mathbf{x}_i) - \mathcal{T}^{-1}(f_2(\mathbf{x}_i)) \right\|^2$
- $\mathcal{L}_{LSF}(\mathbf{x}, \mathbf{y}) = \frac{1}{|b_l|}\sum_{\mathbf{x}_i, \mathbf{y}_i \in b_l} \| f_2(\mathbf{x}_i) - \mathcal{T}(\mathbf{y}_i) \|^2$
- $\mathcal{L}_{Seg}(\mathbf{x}, \mathbf{y}) = 1 - \frac{1}{|b_l|}\sum_{\mathbf{x}_i, \mathbf{y}_i \in b_l} 2\frac{\sum f_1(\mathbf{x}_i)\mathbf{y}_i}{\sum f_1(\mathbf{x}_i) + \sum \mathbf{y}_i}$
- $\mathcal{L}_{total} = \mathcal{L}_{Seg} + \mathcal{L}_{LSF} + \lambda_d \mathcal{L}_{DTC}$
- Computing gradient of loss function $\mathcal{L}_{total}$ and update network parameters θ_1, θ_2 and θ by back propagation.

end while

return θ_1, θ_2 and θ

$$\mathcal{L}_{Seg}(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{x}_i, \mathbf{y}_i \in \mathcal{D}_l} \mathcal{L}_{Dice}(\mathbf{x_i}, \mathbf{y}_i)$$

$$= \sum_{\mathbf{x}_i, \mathbf{y}_i \in \mathcal{D}_l} \left(1 - \frac{2\sum_{x_j \in \mathbf{x}_i, y_j \in \mathbf{y}_i} f_1(x_i)\, y_i}{\sum_{x_j \in \mathbf{x}_i, y_j \in \mathbf{y}_i} f_1(x_j) + \sum_{y_j \in \mathbf{y}_i} y_j}\right)$$

$$\mathcal{L}_{LSF}(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{x}_i, \mathbf{y}_i \in \mathcal{D}_l} \| f_2(\mathbf{x}_i) - \mathcal{T}(\mathbf{y}_i) \|^2$$

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

$$\lambda_d(t) = e^{(-5(1-\frac{t}{t_{max}})^2)}$$

Experiments

Dataset and Metrics

- Pancreas CT data
  - 82 cases
    - 62 volumes for training
    - 20 volumes for testing
  - Semi-supervised setting (20% cases with labeled, 80% without labeled)
- Left Atrium MRI data
  - 100 cases
    - 80 volumes for training
    - 20 volumes for testing
  - Semi-supervised setting (20% cases with labeled, 80% without labeled)
- Evaluation Metrics
  - Dice and Jaccard
  - ASD and HD_{95}

Experiments (1)

The Effects of Different Tasks for Full Supervised Learning

Comparison of different tasks on Pancreas-CT dataset

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

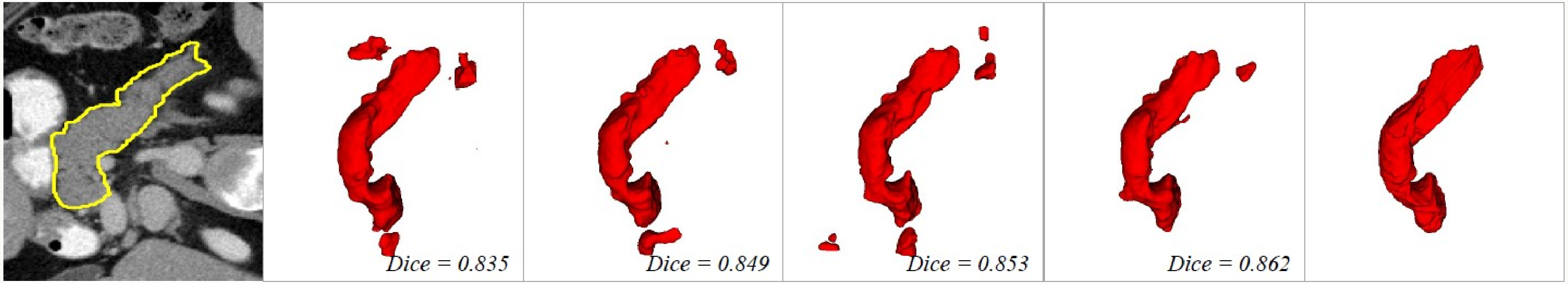
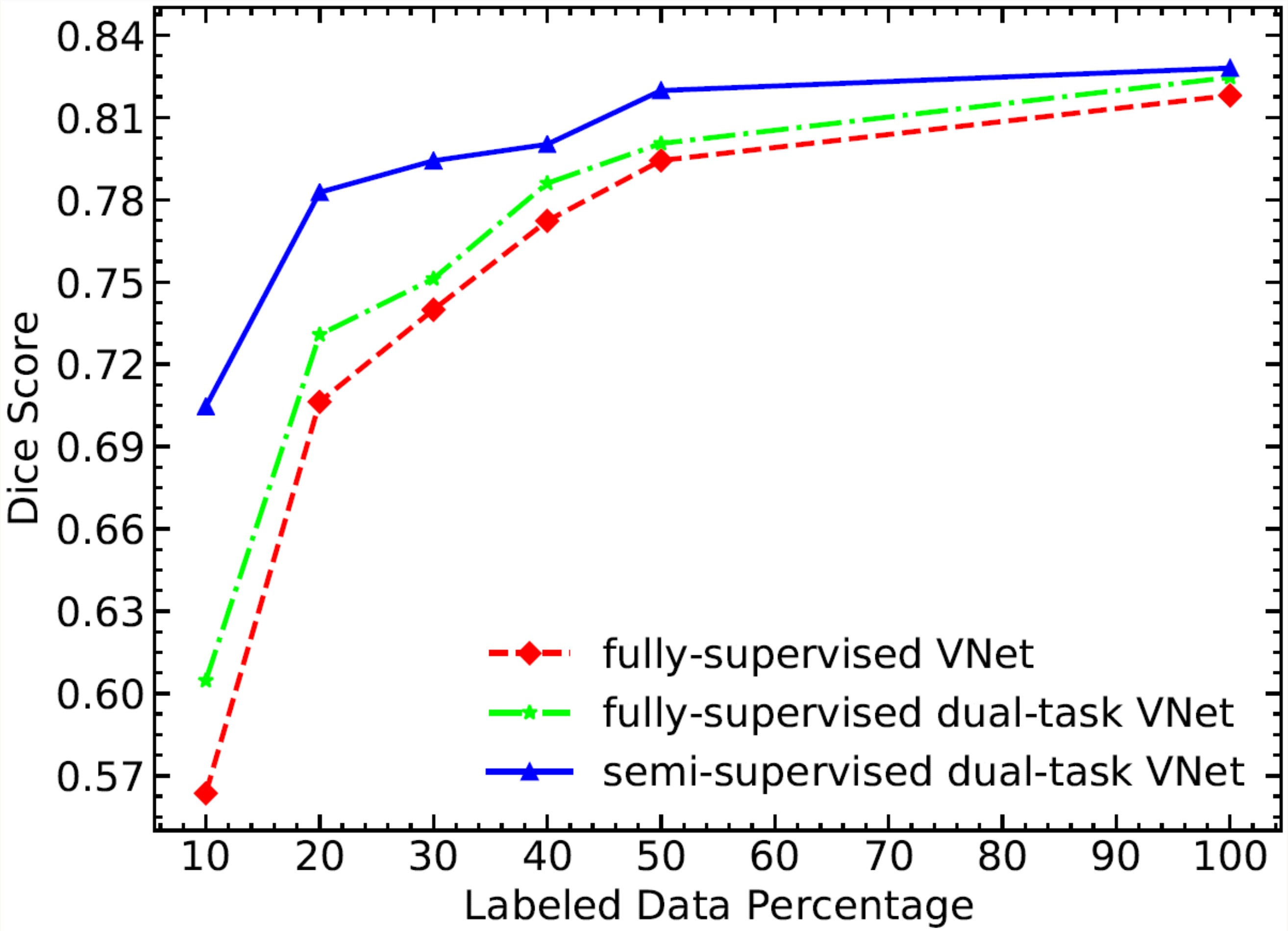


Image with GT | Seg | LSF | Seg + LSF | Seg + LSF +DTC | GT

Experiments (2)

The Effectiveness of DTC for Semi-Supervised Learning



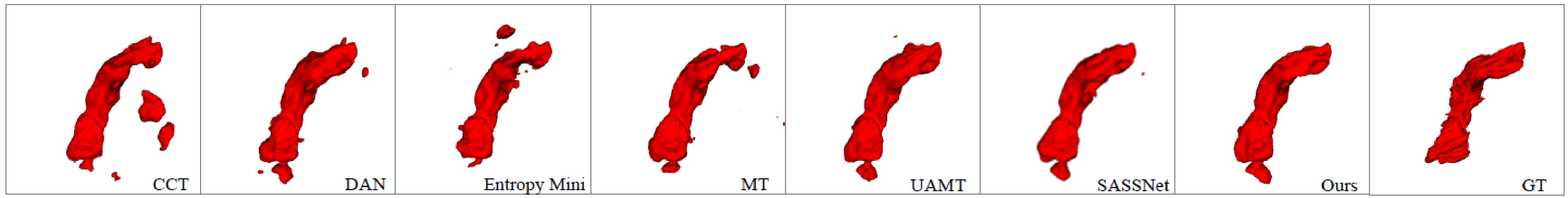
- Effectiveness
  - Utilizing the unlabeled data and brings performance gains.
  - Outperforms fully-supervised baseline.
  - Promising potential for further clinical use

Experiments (3)

Comparison with state-of-the-art methods

Comparison of different tasks on Pancreas-CT dataset

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

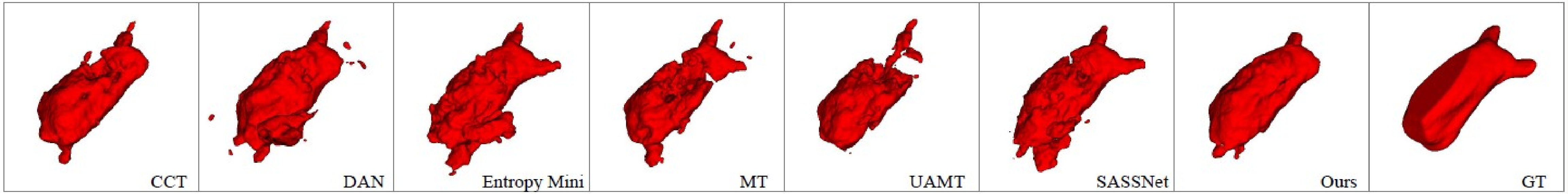


Experiments (3)

Comparison with state-of-the-art methods

Comparison of different tasks on Left-Atrium-MRI dataset

Method	Scans used		Metrics				Cost	
	Labeled	Unlabeled	Dice (%)	Jaccard (%)	ASD (voxel)	95HD (voxel)	Params (M)	Training time (h)
VNet	16	0	86.03	73.26	5.75	17.93	9.44	**1.8**
VNet	80	0	91.14	83.32	1.52	5.75	9.44	2.0
MT(NeurIPS'17)	16	64	88.23	79.29	2.73	10.64	9.44	3.2
DAN (MICCAI'17)	16	64	87.52	78.29	2.42	9.01	12.09	3.7
Entropy Mini (CVPR'19)	16	64	88.45	79.51	3.72	14.14	9.44	**1.9**
UA-MT (MICCAI'19)	16	64	88.88	80.21	2.26	7.32	9.44	3.6
CCT (CVPR'20)	16	64	88.83	80.06	2.49	8.44	15.65	3.9
SASSNet (MICCAI'20)	16	64	89.27	80.82	3.13	8.83	20.46	4.4
Ours	16	64	**89.42**	**80.98**	**2.10**	**7.32**	**9.44**	2.2



Discussion & Conclusion

- Why semi-supervised learning
  - Hard to obtain large-scale labeled data
  - Unlabeled data easy to collect
- Proposed framework
  - Dual-Task-Consistency improves full supervised learning performance
  - Dual-Task-Consistency is useful for semi-supervised learning
  - Task-level consistency is a explicitly modeling method

Future works

- Consistency
  - Data-level (many works have done)
  - Model-level (some works can be found, also can do (single model and single forward))
  - Task-level (worth to explore)
  - Can be used to solve domain generalization
- How to design consistency task
  - Task-transformation must be Differentiable!!!
  - Task-specific (depends on your data, key-points and edge, and others)
  - Relationship (can jointly learn)

Thanks for your attention

Q&A