Uncertainty-aware Self-ensembling Model for Semi-supervised 3D Left Atrium Segmentation

School of Electronic Information and Communications

Huazhong University of Science and Technology

Presenter: Zhanke Zhou

Date: 2020.12.13

Outline

- **Background**
- Introduction to WSL
- Introduction to UA-MT
- UA-MT
- Experiments
- Case Study
- Improvement
- Conclusion
- Discussion
- Q&A

Zhanke Zhou
zhankezhou@hust.edu.cn

Background

Information

- Paper-1: A Brief Introduction to Weakly Supervised Learning
- Authors: **Zhi-Hua Zhou**
- Publish: National Science Review
- Url: https://academic.oup.com/nsr/article/5/1/44/4093912

- Paper-2: Uncertainty-aware Self-ensembling Model for Semi-supervised 3D Left Atrium Segmentation
- Authors: **Lequan Yu,** Shujun Wang, Xiaomeng Li, Chi-Wing Fu, **Pheng-Ann Heng**
- Publish: MICCAI 2019
- Url: https://arxiv.org/abs/1907.07034
- Code: https://github.com/yulequan/UA-MT

Outline

- Background
- **Introduction to WSL**
- Introduction to UA-MT
- UA-MT
- Experiments
- Case Study
- Improvement
- Conclusion
- Discussion
- Q&A

Zhanke Zhou
zhankezhou@hust.edu.cn

Introduction to WSL

Weakly Supervised Learning (WSL)

- **Motivation**
  - In many tasks, it is difficult to get **strong supervision information**
  - like fully ground-truth labels
  - due to the **high cost** of data labeling process.

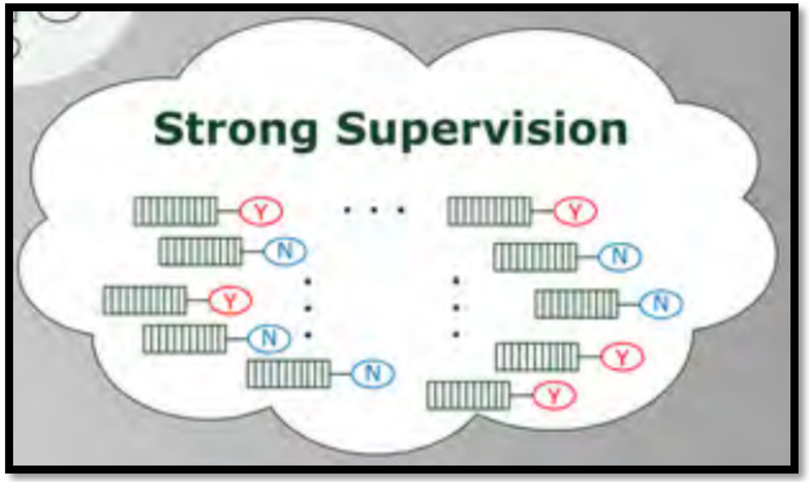
- **Taxonomy of WSL**
  - **Incomplete Supervision 不完全监督**
    - Lack of annotations (i.e., existing plenty of unlabeled data)
  - **Inexact Supervision 不确切监督**
    - Only coarse annotations available
  - **Inaccurate Supervision 不准确监督**
    - Learning with label noise

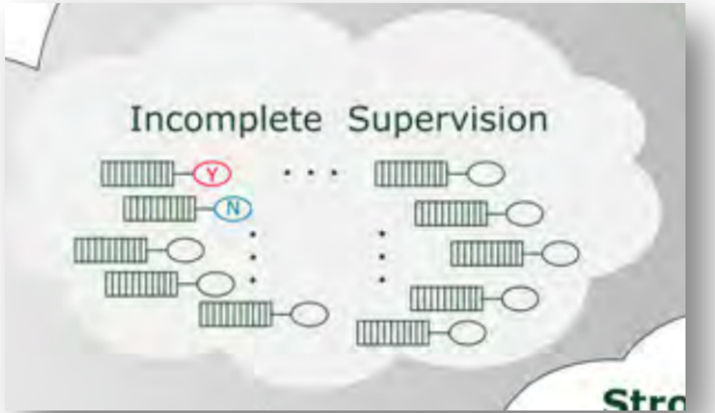
Zhanke Zhou
zhankezhou@hust.edu.cn

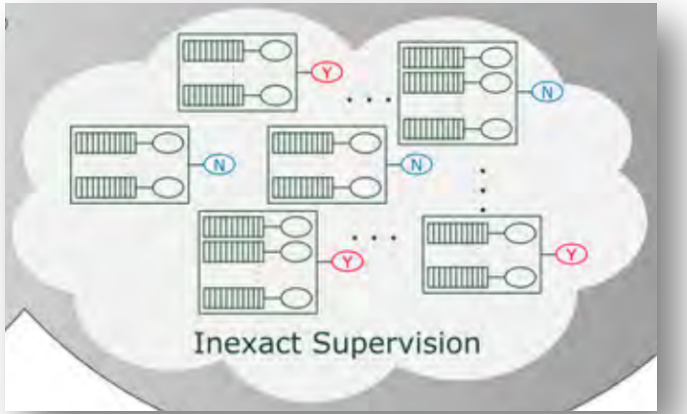
Introduction to WSL

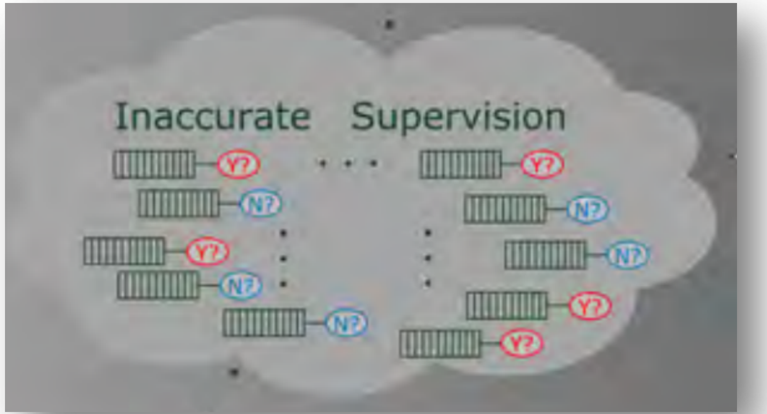
Weakly Supervised Learning (WSL)

- Taxonomy of WSL
  - **Incomplete** Supervision 不完全监督
  - **Inexact** Supervision 不确切监督
  - **Inaccurate** Supervision 不准确监督









Zhanke Zhou
zhankezhou@hust.edu.cn

Outline

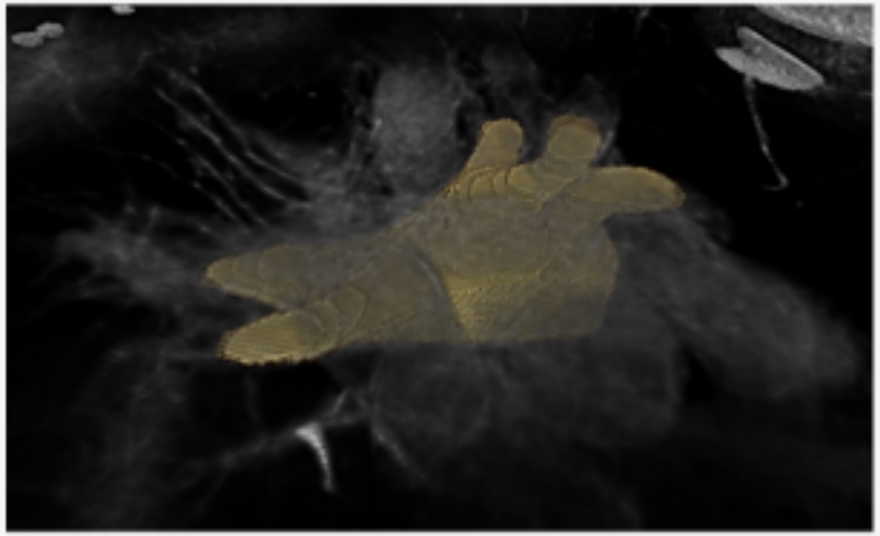
- Background
- Introduction to WSL
- **Introduction to UA-MT**
- UA-MT
- Experiments
- Case Study
- Improvement
- Conclusion
- Discussion
- Q&A

Zhanke Zhou
zhankezhou@hust.edu.cn

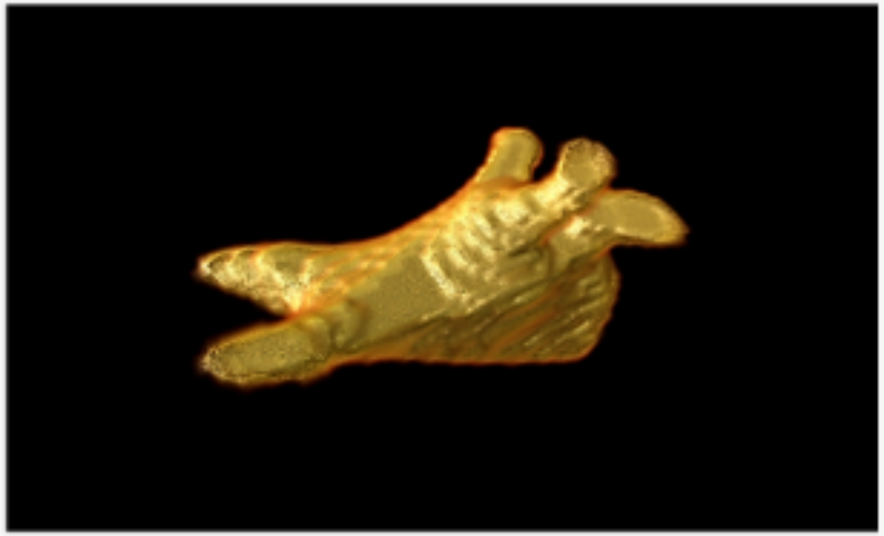
Introduction to UA-MT

3D Left Atrium(LA) Segmentation

- Pathology: Atrial Fibrillation
- Images: 3D Gadolinium-Enhanced Magnetic Resonance (MR) Imaging
- Labels : 3D Binary Masks of the Left Atrial Cavity



3D Left Atria Superimposed on LGE-MRI



3D Left Atria Visualization

Zhanke Zhou
zhankezhou@hust.edu.cn

2018 Atrial Segmentation Challenge

Introduction to UA-MT

Challenges

- Lack of labeled data
  - **Why?** High cost in collecting and annotating

Motivation

- **UA-MT: Uncertainty-Aware Mean Teacher Framework**
- **Three Key Points**
- **Semi-supervised Learning**
  - Leveraging abundant **unlabeled** data
- **Uncertainty-aware (UA)**
  - Considering the **reliability** of the targets
- **Mean Teacher (MT)**
  - To overcome the limitations of **Temporal Ensembling**

Zhanke Zhou
zhankezhou@hust.edu.cn

Outline

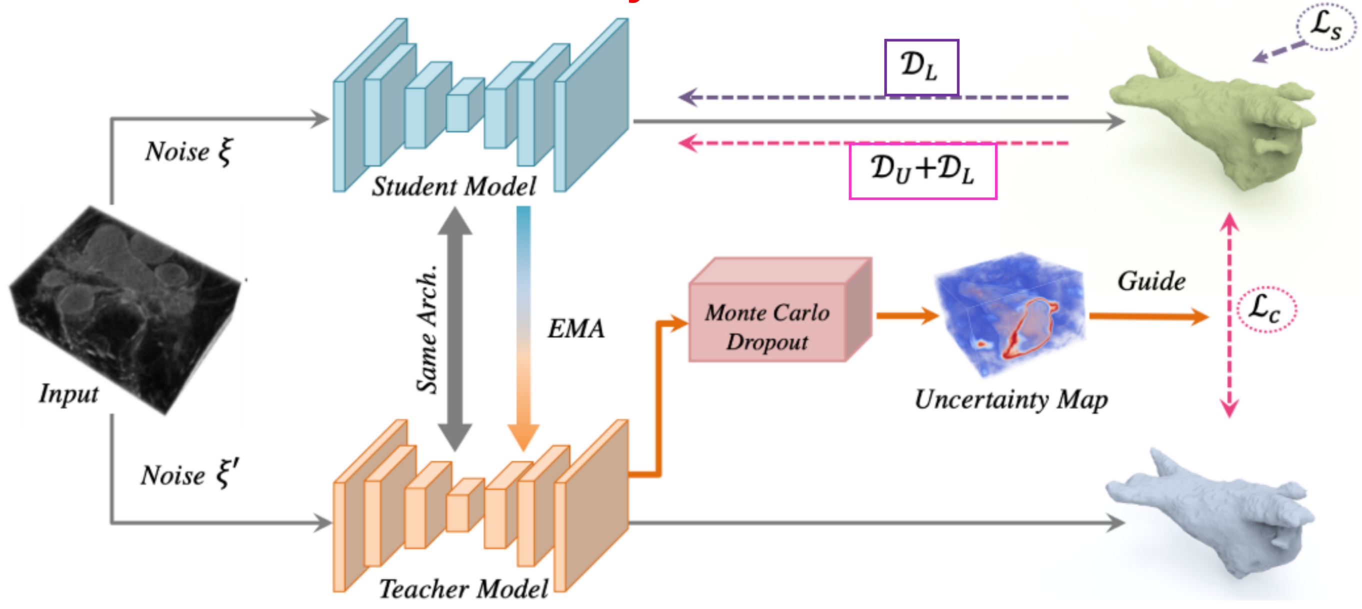
- Background
- Introduction to WSL
- Introduction to UA-MT
- **UA-MT**
- Experiments
- Case Study
- Improvement
- Conclusion
- Discussion
- Q&A

Zhanke Zhou
zhankezhou@hust.edu.cn

UA-MT

Network Structure

- Consisting of a student model and a teacher model
  - **Teacher model**
    - Generates **targets** for the student model to learn from
    - Estimates the **uncertainty** of the targets
  - **Student model**
    - optimized by minimizing two kinds of loss
      - **Ls : supervised loss** on labeled data
      - **Lc : consistency loss** on both labeled and unlabeled data



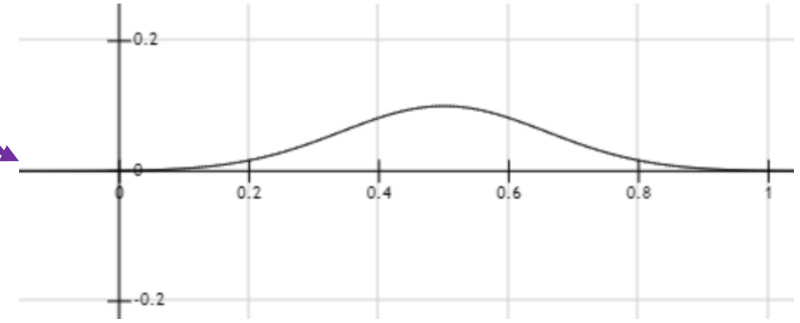
Labeled Data: $\mathcal{D}_L = \{(x_i, y_i)\}_{i=1}^{N}$

Unlabeled Data: $\mathcal{D}_U = \{x_i\}_{i=N+1}^{N+M}$

Objective function:

$$\boxed{\min_{\theta} \sum_{i=1}^{N} \mathcal{L}_s(f(x_i;\theta), y_i) + \boxed{\lambda} \sum_{i=1}^{N+M} \mathcal{L}_c(f(x_i;\theta', \xi'), f(x_i;\theta,\xi))}$$

$$\boxed{\lambda(t)} = 0.1 * e^{(-5(1-t/t_{max})^2)}$$



UA-MT

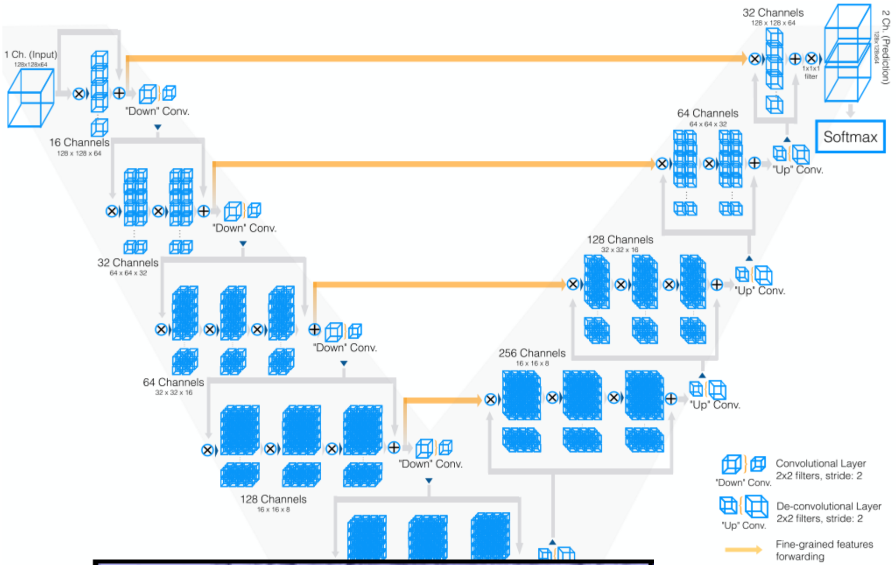
Backbone

- V-Net: a **volumetric, fully convolutional** neural network for 3D image segmentation
- Input shape: 128×128×64
- Output shape: **2**×128×128×64
  - **2 Classes :** Background (0) / Foreground (1)

Supervised Loss

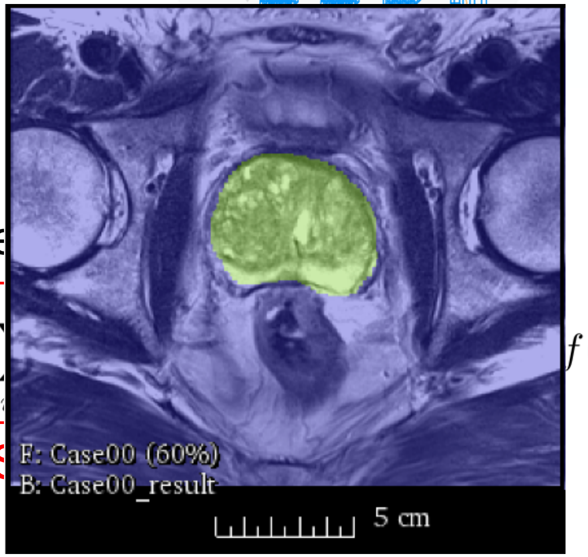
- Cross entropy loss + Dice loss
  - **Why dice loss?**
    - Foreground occupies only a small region
    - Give more importance to foreground
  - Dice loss = 1 - D (dice coefficient)

$$D = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2}$$



Obje

$\min_\theta$ ... $f(x_i;\theta',\xi'), f(x_i;\theta,\xi))$



Zhanke Zhou
zhankezhou@hust.edu.cn

UA-MT

Objective function:

Consistency Loss

$$\min_{\theta} \sum_{i=1}^{N} \mathcal{L}_s(f(x_i;\theta), y_i) + \lambda \boxed{\sum_{i=1}^{N+M} \mathcal{L}_c(f(x_i;\theta',\xi'), f(x_i;\theta,\xi))}$$

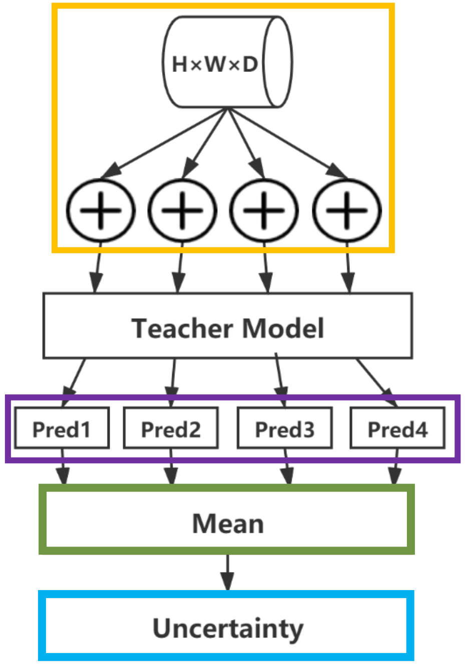
Uncertainty Estimation

- **Why?**
  - The **predicted targets** from the teacher model may be **unreliable and noisy**
  - Previous works do not consider the **reliability** of the targets
- Estimate the uncertainty with
  - Monte Carlo Sampling
  - Test Data Augmentation
- **Predictive Entropy**
  - The metric to approximate the uncertainty
    - obtain a set of softmax probability vector: $\{\mathbf{p}_t\}_{t=1}^{T}$

$$\mu_c = \frac{1}{T}\sum_t \mathbf{p}_t^c \quad \text{and} \quad u = -\sum_c \mu_c \log \mu_c$$

- **Uncertainty-Aware Consistency Loss**
  - voxel-level mean squared error (MSE) loss

$$\mathcal{L}_c(f', f) = \frac{\sum_v \mathbb{I}(u_v < H)\,\|f'_v - f_v\|^2}{\sum_v \mathbb{I}(u_v < H)}$$



T=4	Fore	Back
Sample 1	0.9	0.1
Sample 2	0.91	0.09
Sample 3	0.92	0.08
Sample 4	0.89	0.11
Mean	0.905	0.095
Uncertainty	0.453	

T=4	Fore	Back
Sample 1	0.9	0.1
Sample 2	0.8	0.2
Sample 3	0.95	0.05
Sample 4	0.75	0.25
Mean	0.85	0.15
Uncertainty	0.609	

Zhanke Zhou
zhankezhou@hust.edu.cn

UA-MT

Mean Teacher (MT)

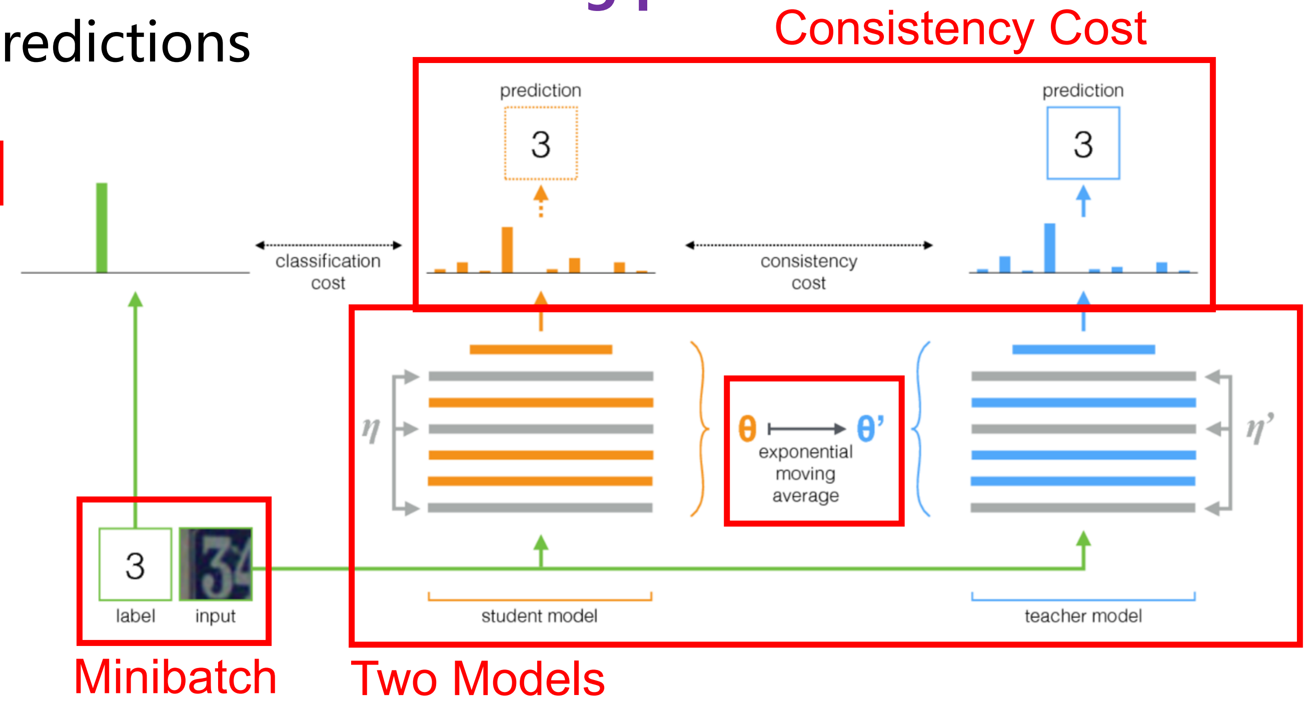
- **Why called Mean Teacher?**
  - The teacher model is an **average** of consecutive student models
- **Why using Mean Teacher?**
  - To overcome the limitations of Temporal Ensembling
  - Ensemble predictions of the network **at different training process** can improve the **quality** of the predictions

Exponential Moving Average (EMA)

- Updating the teacher's weights θ'_t at training step t as:

$$\theta'_t = \alpha\theta'_{t-1} + (1-\alpha)\theta_t$$

$$\alpha = \min(1 - 1/(t+1), \alpha)$$



Zhanke Zhou
zhankezhou@hust.edu.cn

Outline

- Background
- Introduction to WSL
- Introduction to UA-MT
- UA-MT
- **Experiments**
- Case Study
- Improvement
- Conclusion
- Discussion
- Q&A

Zhanke Zhou
zhankezhou@hust.edu.cn

Experiments

Atrial Segmentation Challenge dataset

- Distribution
  - **100 scans in total**
    - 80 for training
    - 20 for evaluation

Table 1: Comparison between our method and various methods.

Method	# scans used		Metrics			
	Labeled	Unlabeled	Dice[%]	Jaccard[%]	ASD[voxel]	95HD[voxel]
Vanilla V-Net	16	0	84.13	73.26	4.75	17.93
Bayesian V-Net	16	0	86.03	76.06	3.51	14.26
Vanilla V-Net	80	0	90.25	82.40	1.91	8.29
Bayesian V-Net	80	0	91.14	83.82	1.52	5.75
Self-training [1]	16	64	86.92	77.28	2.21	9.19
DAN [18]	16	64	87.52	78.29	2.42	9.01
ASDNet [12]	16	64	87.90	78.85	**2.08**	9.24
TCSE [10]	16	64	88.15	79.20	2.44	9.57
UA-MT-UN (ours)	16	64	88.83	80.13	3.12	10.04
UA-MT (ours)	16	64	**88.88**	**80.21**	2.26	**7.32**

Evaluation

- Metrics
  - Dice
  - Jaccard
  - Average Surface Distance (ASD)
  - 95% Hausdorff Distance (95HD)

Table 2: Quantitative analysis of our method.

Method	# scans used		Metrics			
	Labeled	Unlabeled	Dice[%]	Jaccard[%]	ASD[voxel]	95HD[voxel]
MT	16	64	88.23	79.29	2.73	10.64
MT-Dice [5]	16	64	88.32	79.37	2.76	10.50
Our UA-MT	16	64	88.88	80.21	2.26	7.32
Bayesian V-Net	8	0	79.99	68.12	5.48	21.11
Our UA-MT	8	72	84.25	73.48	3.36	13.84
Bayesian V-Net	24	0	88.52	79.70	2.60	10.45
Our UA-MT	24	56	90.16	82.18	2.73	8.90

- **Outperforms** the state-of-the-art semi-supervised methods

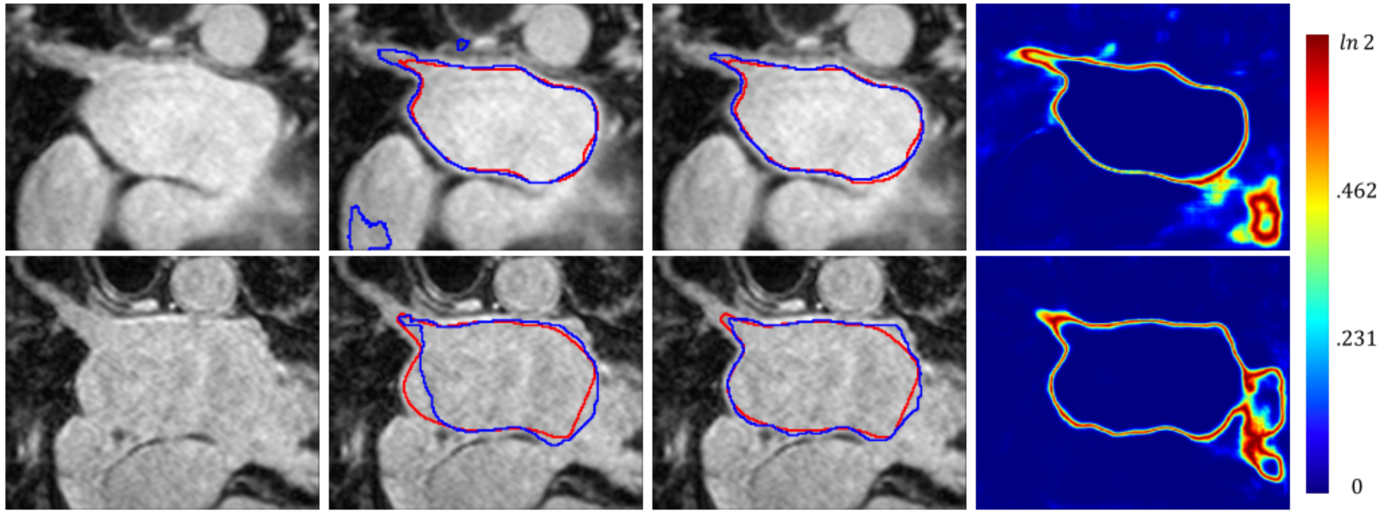
Zhanke Zhou
zhankezhou@hust.edu.cn

Case Study

Segmentation examples

- Compared with the **supervised** method, UA-MT
  - have higher overlap ratio with the ground truth
  - produce less false positives

- The network estimates high uncertainty
  - near the **boundary**
  - **ambiguous regions** of great vessels



(a) Image *(b) Supervised* *(c) Our method* *(d) Uncertainty*

Zhanke Zhou
zhankezhou@hust.edu.cn

Outline

- Background
- Introduction to WSL
- Introduction to UA-MT
- UA-MT
- Experiments
- Case Study
- **Improvement**
- Conclusion
- Discussion
- Q&A

Zhanke Zhou
zhankezhou@hust.edu.cn

Improvement

Better Uncertainty Estimation Module (UEM)

- **Uncertainty Estimation Matters**
  - The key of success lies in **the quality of targets**
  - **Effect**
    - Identify difficult cases
    - Detect out-of-distribution samples
- Weakness of UEM in UA-MT
  - Only concerning single type of noise
  - Only one metric

Uncertainty	Agg. Metric	Bal. Acc.	AUC. UNK
MC Drop.	Entropy	0.476	0.613
	Variance	0.508	0.645
	BC	0.525	0.579
Test Aug.	Entropy	0.411	0.660
	Variance	0.390	0.684
	BC	0.377	0.622
Both	Entropy	0.437	0.670
	Variance	0.349	**0.692**
	BC	0.379	0.622
Control	-	**0.550**	0.500

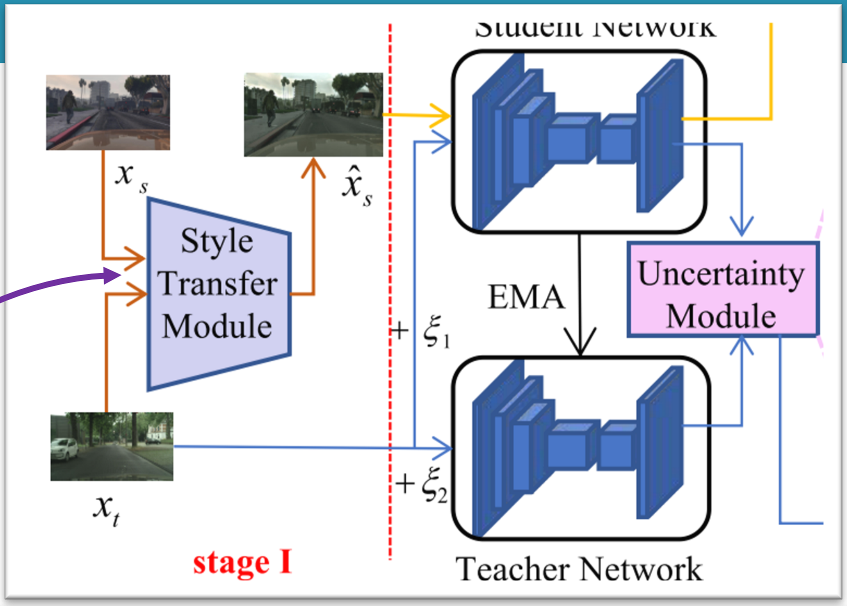
- **Improvement (of Presentation Level)**
  - **More metrics** to quantify the uncertainty [1]
    - Variance
    - Bhattacharyya Coefficient (BC)
  - Design an uncertainty-Aware **attention module** [2]

$$\boxed{\sigma^2(\mathbf{p}_T(x))} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{p}_t(x) - \mathbf{p}_T(X))^2$$

$$\boxed{BC(h_{c1}, h_{c2})(x)} = \sum_{n=1}^{N}\sqrt{h_{c1}[n] * h_{c2}[n]}$$

Zhanke Zhou
zhankezhou@hust.edu.cn

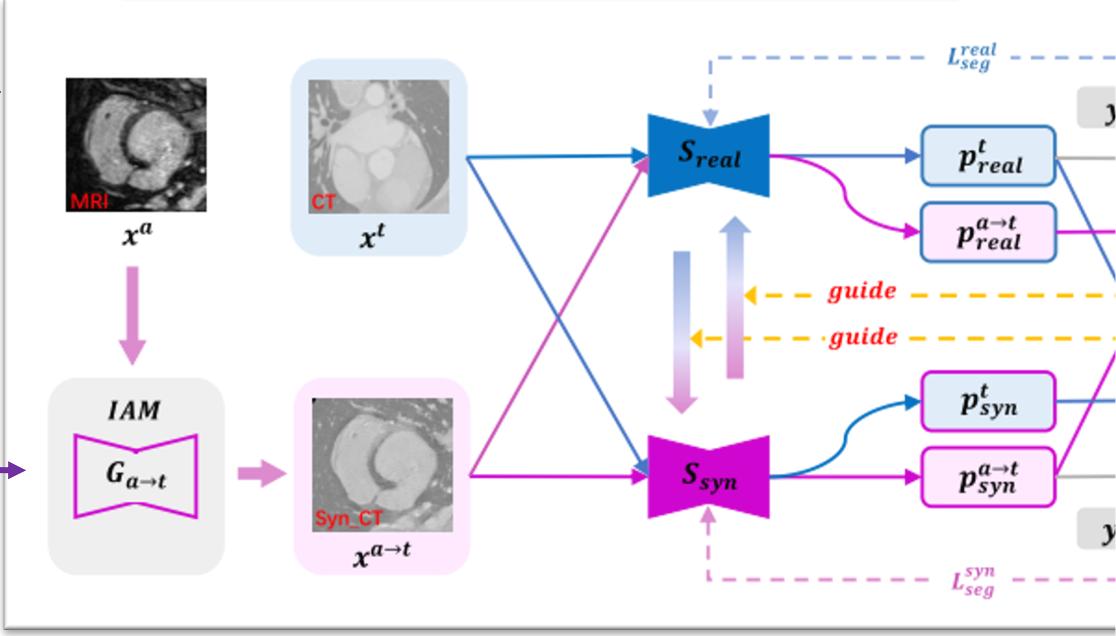
Improvement



Transferable Prior Knowledge

- To make up annotation scarcity
- Weakness of UA-MT
    - Only using single small-scale dataset
    - No enough knowledge

- **Improvement**
    - Exploit the prior knowledge (e.g. shape priors)
        - learned from assistant modality
        - to improve the performance on target modality
    - **Style Transfer Module** [3]
        - Produce translated images to address the domain gap
    - **Image Alignment Module** [4]
        - To narrow the appearance gap between assistant and target modality data



Zhanke Zhou
zhankezhou@hust.edu.cn

Outline

- Background
- Introduction to WSL
- Introduction to UA-MT
- UA-MT
- Experiments
- Case Study
- Improvement
- **Conclusion**
- Discussion
- Q&A

Zhanke Zhou
zhankezhou@hust.edu.cn

Conclusion

UA-MT

- A novel **uncertainty-aware** **semi-supervised learning** method for left atrium segmentation from 3D MR images
- Encourages the segmentation to be **consistent**
  - For the same input under different **perturbations**
- Explore the model **uncertainty** to improve the quality of the target
- **Outperforms** the state-of-the-art semi-supervised methods

Zhanke Zhou
zhankezhou@hust.edu.cn

Discussion

Use UA-MT for reference (in our project)

- **Better Performance**
  - Utilize abundant unlabeled data

- **Higher Stability**
  - Uncertainty estimation
  - Improving training process

- **Explore new ideas**
  - Weakly supervised Learning
    - Semi-supervised learning
    - Label-noise representation learning
    - Transfer learning

Zhanke Zhou
zhankezhou@hust.edu.cn

Reference

1. Bai, W., Oktay, O., Sinclair, M.e.a.: Semi-supervised learning for network-based cardiac mr image segmentation. In: MICCAI. pp. 253–260 (2017)
2. Baur, C., Albarqouni, S., Navab, N.: Semi-supervised deep learning for fully con- volutional networks. In: MICCAI. pp. 311–319 (2017)
3. Chartsias, A., Joyce, T., Papanastasiou, G., Semple, S., Williams, M., Newby, D., Dharmakumar, R., Tsaftaris, S.A.: Factorised spatial representation learning: ap- plication in semi-supervised myocardial segmentation. MICCAI pp. 490–498 (2018)
4. Chen, C., Bai, W., Rueckert, D.: Multi-task learning for left atrial segmentation on ge-mri. arXiv preprint arXiv:1810.13205 (2018)
5. Cui, W., Liu, Y., Li, Y., Guo, M., Li, Y., Li, X., Wang, T., Zeng, X., Ye, C.: Semi-supervised brain lesion segmentation with an adapted mean teacher model. In: IPMI. pp. 554–565 (2019)
6. Dong, N., Kampffmeyer, M., Liang, X., Wang, Z., Dai, W., Xing, E.: Unsupervised domain adaptation for automatic estimation of cardiothoracic ratio. In: MICCAI. pp. 544–552 (2018)
7. Ganaye, P.A., Sdika, M., Benoit-Cattin, H.: Semi-supervised learning for segmen- tation under semantic constraint. In: MICCAI. pp. 595–602 (2018)
8. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? In: NIPS. pp. 5574–5584 (2017)
9. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. arXiv preprint (2016)
10. Li, X., Yu, L., Chen, H., Fu, C.W., Heng, P.A.: Semi-supervised skin lesion seg- mentation via transformation consistent self-ensembling model. BMVC (2018)
11. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 3DV. pp. 565–571 (2016)
12. Nie, D., Gao, Y., Wang, L., Shen, D.: Asdnet: Attention based semi-supervised deep networks for medical image segmentation. In: MICCAI. pp. 370–378 (2018)
13. Perone, C.S., Cohen-Adad, J.: Deep semi-supervised segmentation with weight- averaged consistency targets. In: DLMIA workshop (2018)
14. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: NIPS (2017)
15. Xiong, Z., Fedorov, V.V., Fu, X., Cheng, E., Macleod, R., Zhao, J.: Fully automatic left atrium segmentation from late gadolinium enhanced magnetic resonance imag- ing using a dual fully convolutional neural network. TMI 38(2), 515–524 (2019)
16. Yang, X., Bian, C., Yu, L., Ni, D., Heng, P.A.: Hybrid loss guided convolutional networks for whole heart parsing. In: International Workshop on STACOM (2017)
17. Yu, L., Cheng, J.Z., Dou, Q., Yang, X., Chen, H., Qin, J., Heng, P.A.: Automatic 3d cardiovascular mr segmentation with densely-connected volumetric convnets. In: MICCAI. pp. 287–295. Springer (2017)
18. Zhang, Y., Yang, L., Chen, J., Fredericksen, M., Hughes, D.P., Chen, D.Z.: Deep adversarial networks for biomedical image segmentation utilizing unannotated im- ages. In: MICCAI. pp. 408–416 (2017)
19. Zhou, Y., Wang, Y., Tang, P., Bai, S., Shen, W., Fishman, E.K., Yuille, A.L.: Semi- supervised multi-organ segmentation via multi-planar co-training. arXiv preprint arXiv:1804.02586 (2018)

Zhanke Zhou
zhankezhou@hust.edu.cn

Q&A

Thanks for your attention!

Outline

- Background
- Introduction
- UA-MT
- Experiments
- Case Study
- Improvement
- Conclusion