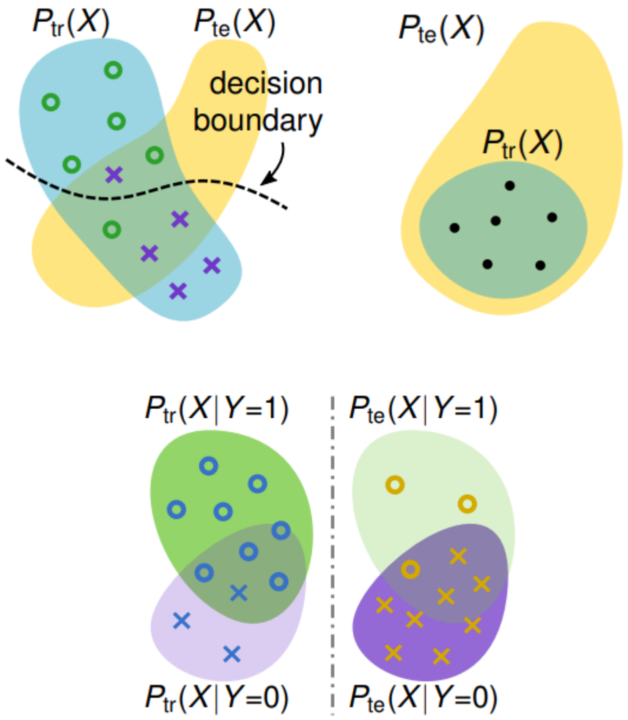
# Deep Model Generalization for Medical Image Computing at Scale

DOU Qi

Department of Computer Science and Engineering
co-affiliated with T Stone Robotics Institute
The Chinese University of Hong Kong

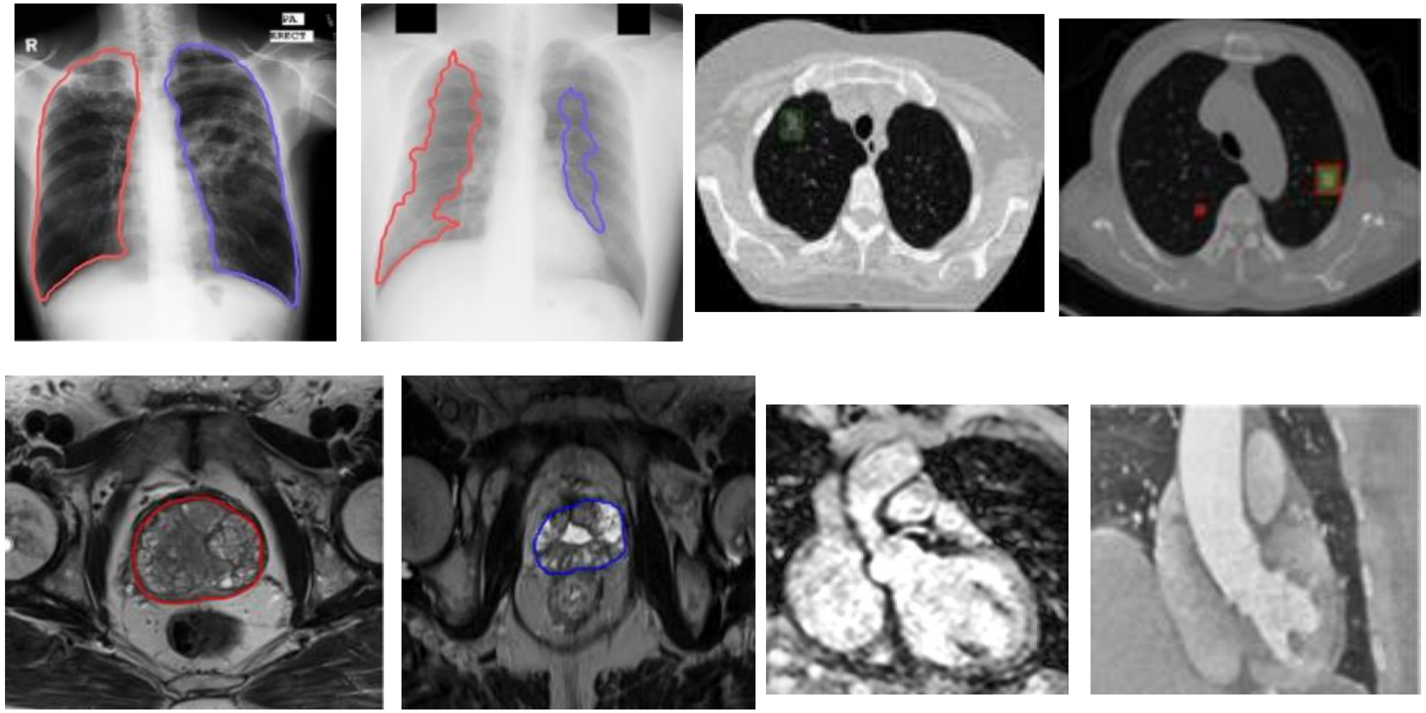
# Model Generalization in Real World Conditions

**Data-driven method is sensitive to data mismatch**




[D. Castro, I. Walker, and B. Glocker. 2019]

- **Large-scale data always encounter data heterogeneity**
- **Medical imaging:** different vendors, imaging protocols, patient population, etc.



# Tackling Data Heterogeneity: does Normalization Help?

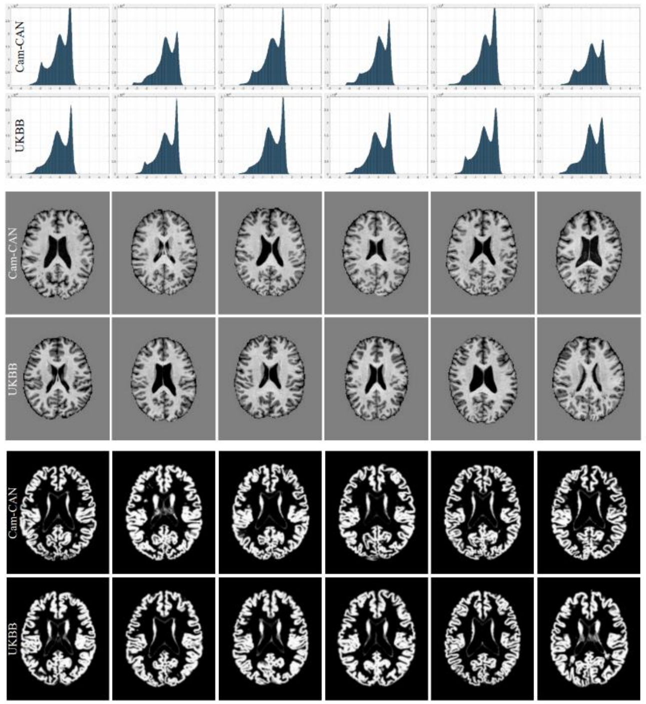
## An empirical study on the impact of scanner effects with brain imaging

Construct an **age- and sex-matched** dataset with T1-weighted brain MRI from n = 592 individuals, where 296 subjects (146 F) are taken each from the Cam-CAN and UKBB, to simulate a somewhat 'best case scenario' to **remove population bias**.

Very **careful pre-processing** is conducted, including: 1) reorientation, 2) skull stripping, 3) bias field correction, 4) intensity-based linear registration (rigid and affine) to MNI space, 5) whitening for intensity normalization

### Site classification with random forest binary classifier

| Stripped | Bias Field | Aligned | Intensities | Accuracy | Avg. Entropy | Avg. Prob. |
|---|---|---|---|---|---|---|
| ✓ | ✓ | rigid | whitening | 96.96% | 0.4039 | 0.8296 |
| ✓ | ✓ | affine | whitening | 98.82% | 0.3876 | 0.8397 |
| **SPM12 – Segment** | | | | **Accuracy** | **Avg. Entropy** | **Avg. Prob.** |
| ✗ | ✓ | rigid | graymatter | 80.24% | 0.6363 | 0.6399 |
| ✗ | ✓ | non-linear | graymatter | 96.62% | 0.5675 | 0.7234 |
| **FSL – FAST** | | | | **Accuracy** | **Avg. Entropy** | **Avg. Prob.** |
| ✓ | ✓ | rigid | graymatter | 93.24% | 0.4542 | 0.7968 |



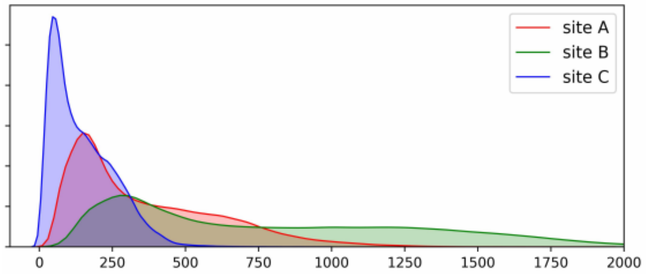

*B. Glocker et al. "Machine Learning with Multi-site Imaging Data: An Empirical Study on the Impact of Scanner Effects." Medical Imaging meets NeurIPS Workshop, 2019.*

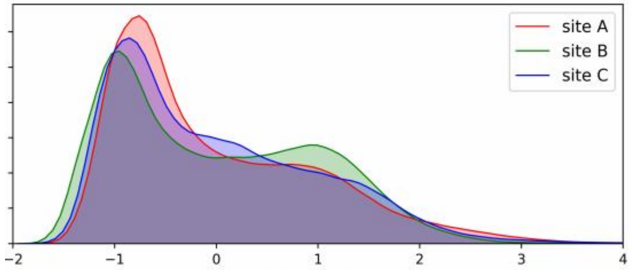
*Related work: [Shafto et al., 2014; Taylor et al., 2017; Sudlow et al., 2015; Miller et al., 2016; Alfaro-Almagro et al., 2018]*

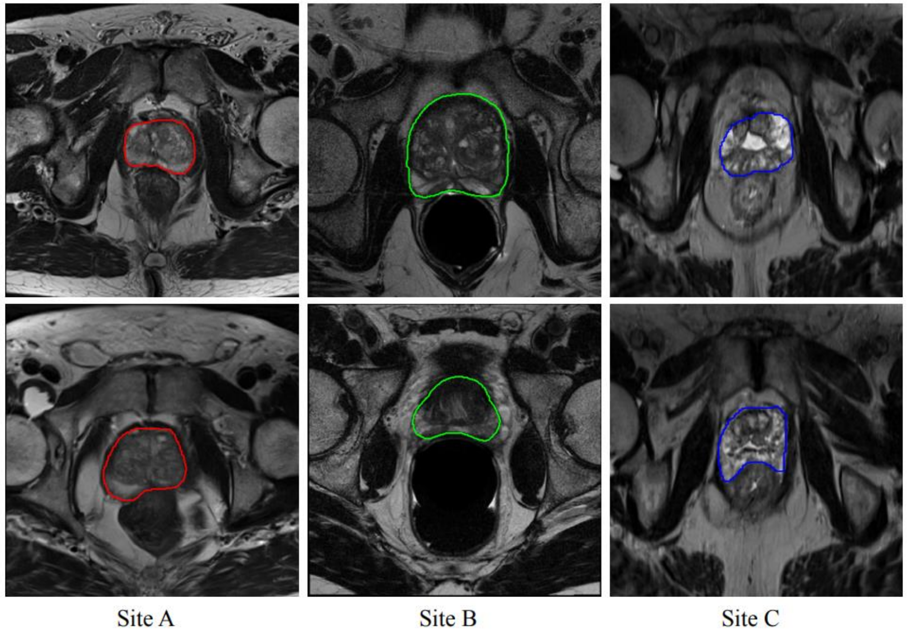
# Tackling Data Heterogeneity with Supervised Learning

## A case study with prostateT2-weighted MRI image segmentation

| Dataset | Case num | Field strength (T) | Resolution(in-plane/through-plane)(mm) | Coil | Manufactor |
|---|---|---|---|---|---|
| Site A | 30 | 3 | 0.6-0.625/3.6-4 | Surface | Siemens |
| Site B | 30 | 1.5 | 0.4/3 | Endorectal | Philips |
| Site C | 19 | 3 | 0.67-0.79/1.25 | No | Siemens |



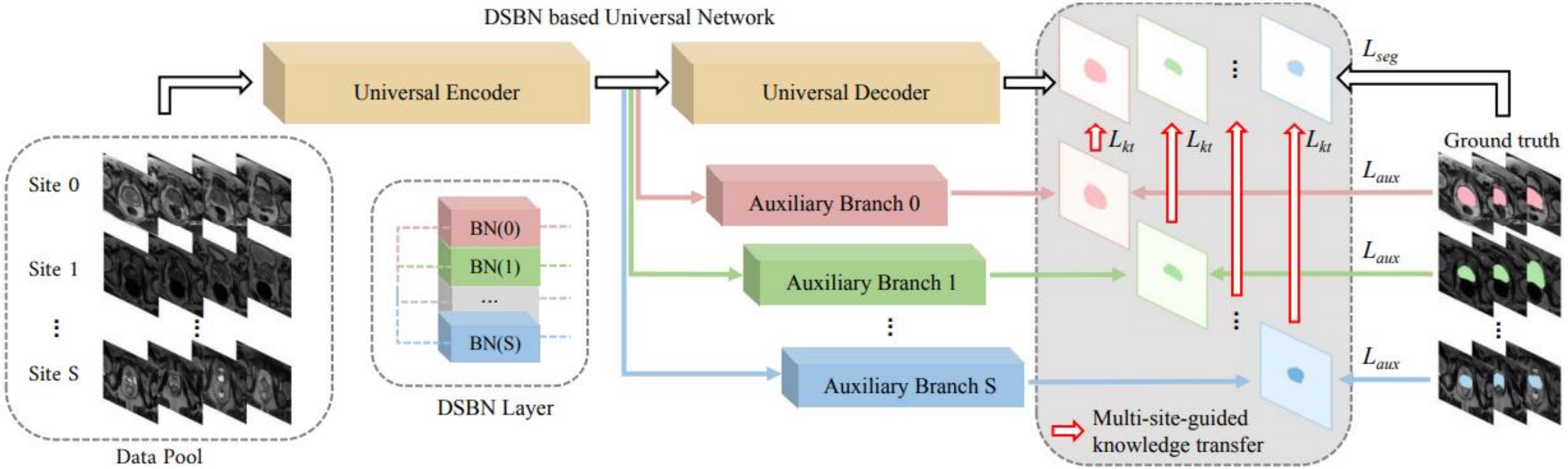



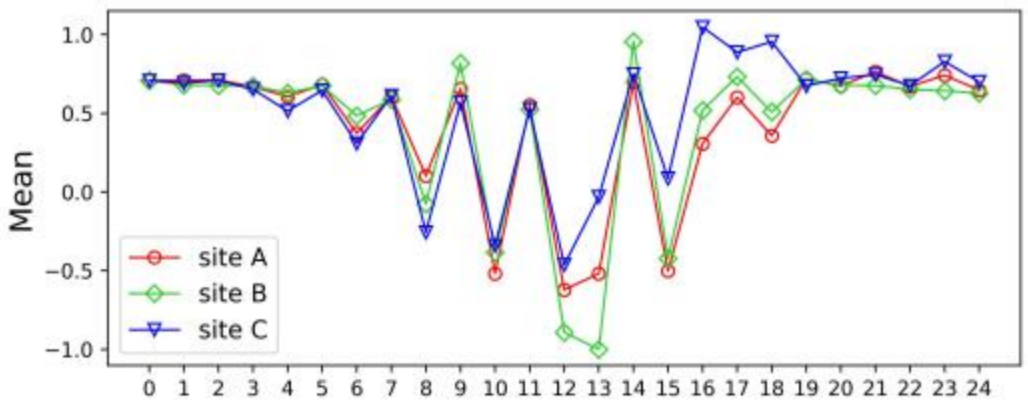

| Methods | BFC | NF | Intensities | Site A | Site B | Site C | Overall |
|---|---|---|---|---|---|---|---|
| Separate (A) | ✗ | ✗ | whitening | 90.47 | 76.44 | 56.81 | |
| Separate (B) | ✗ | ✗ | whitening | 70.11 | 90.52 | 50.25 | 90.56 |
| Separate (C) | ✗ | ✗ | whitening | 57.93 | 55.25 | 90.70 | |
| Joint | ✗ | ✗ | ✗ | 86.51 | 88.00 | 86.78 | 87.10 |
| Joint | ✗ | ✗ | histogram | 87.68 | 88.02 | 89.46 | 88.39 |
| Joint | ✗ | ✗ | scaled | 90.43 | 88.06 | 88.26 | 88.92 |
| Joint | ✗ | ✗ | whitening | 90.69 | 89.53 | 90.55 | 90.25 |
| Joint | ✗ | ✓ | whitening | 90.76 | 89.46 | 90.91 | 90.37 |
| Joint | ✓ | ✗ | whitening | 90.84 | 89.81 | 90.81 | 90.49 |
| Joint | ✓ | ✓ | whitening | 91.14 | 89.75 | 90.83 | 90.58 |

*Q. Liu, Q. Dou, et al. "MS-Net: Multi-Site Network for Improving Prostate Segmentation with Heterogeneous MRI Data", IEEE Trans. on Medical Imaging, 2020.*

*Related work: [Karani et al. MICCAI 2018; Gibson et al. MICCAI 2018; John et al. ISBI 2019]*

# Tackling Data Heterogeneity with Supervised Learning



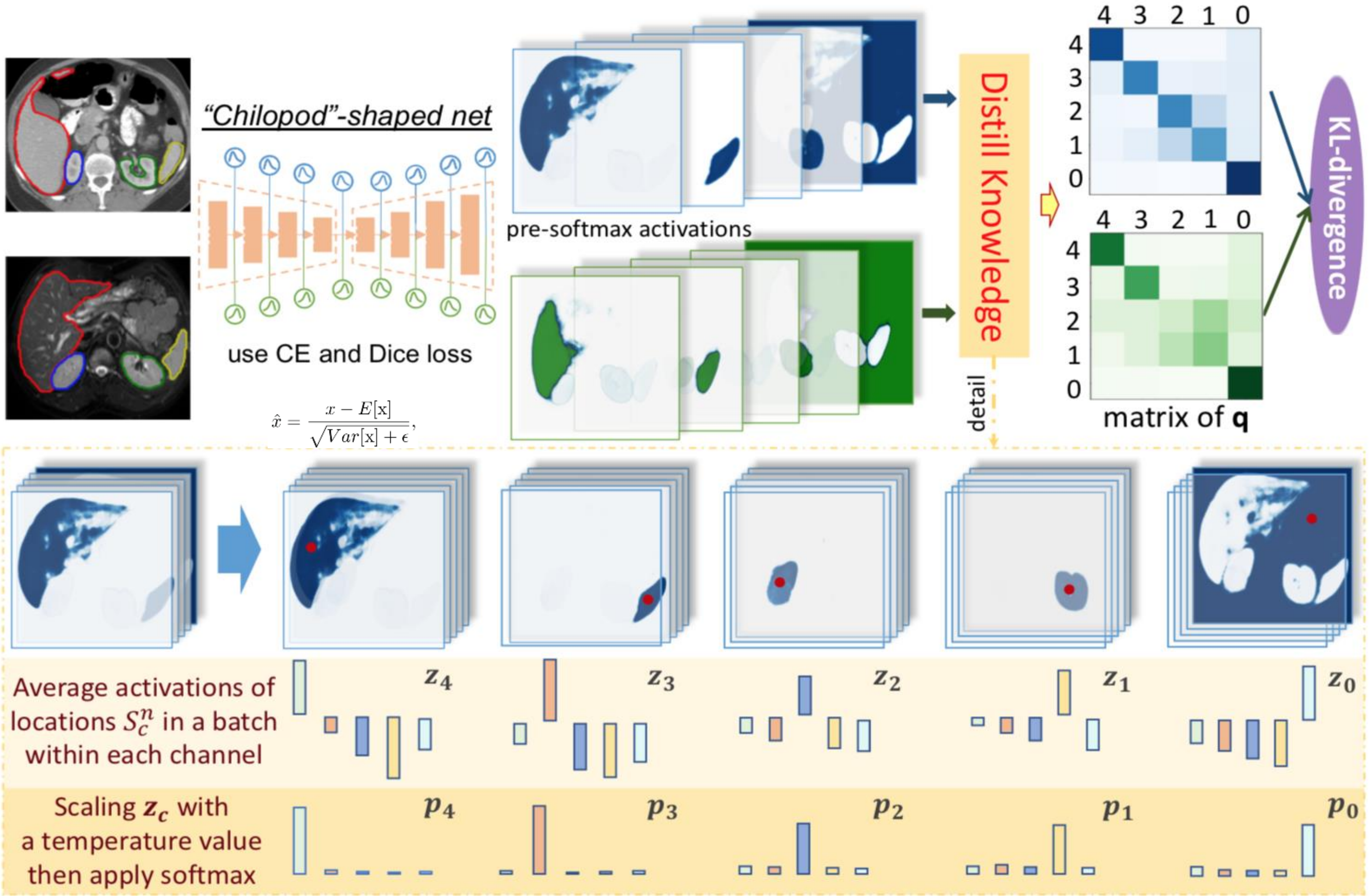




| Methods | Dice Coefficient (mean±std, %) Site A | Site B | Site C | Overall |
|---|---|---|---|---|
| Tian *et al.* [26] | 88.23 | 88.23 | — | |
| Rundo *et al.* [9] | — | — | 88.66 | |
| Separate | 90.47±3.00 | 90.52±2.45 | 90.70±3.34 | 90.56±2.88 |
| Joint | 90.69±3.05 | 89.53±2.97 | 90.55±3.18 | 90.25±3.08 |
| USE-Net [19] | 90.90±2.41 | 90.17±2.61 | 90.73±2.36 | 90.60±2.50 |
| Dual-Stream [47] | 90.87±2.85 | 90.57±2.12 | 90.10±3.28 | 90.51±2.72 |
| Series-Adapter [48] | 90.80±2.72 | 89.92±2.80 | 91.24±2.21 | 90.65±2.71 |
| Parallel-Adapter [23] | 90.61±3.54 | 90.71±2.17 | 91.30±2.06 | 90.88±2.79 |
| DSBN (ours) | 90.98±2.69 | 90.67±2.22 | 91.07±1.86 | 90.91±2.36 |
| **MS-Net (ours)** | **91.54±2.01** | **91.24±1.97** | **92.18±1.62** | **91.66±1.95** |

*Q. Liu, Q. Dou, et al. "MS-Net: Multi-Site Network for Improving Prostate Segmentation with Heterogeneous MRI Data", IEEE Trans. on Medical Imaging, 2020.*

# Tackling Data Heterogeneity with Supervised Learning

## Unpaired Multi-modal Learning with Knowledge Distillation




**Distill activations per-class:**

$$\mathbf{z}_c^i = \frac{1}{\sum_n |S_c^n|} \sum_n \sum_{(w,h)\in S_c^n} z_{nwhi},$$

$$\mathbf{p}_c^i = \frac{\exp(\mathbf{z}_c^i/T)}{\sum_j \exp(\mathbf{z}_c^j/T)},$$

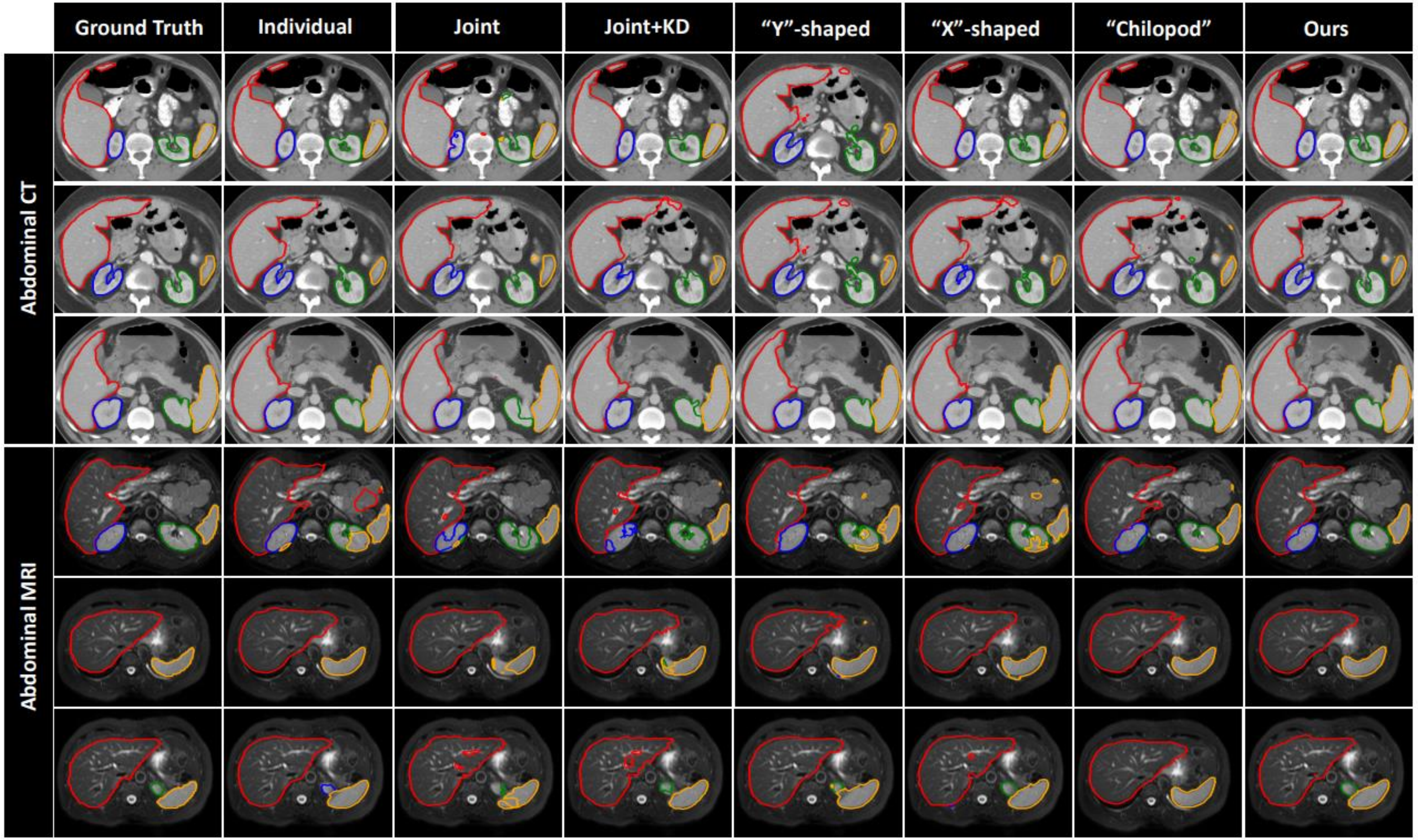
**Minimize probability divergence:**

$$\mathcal{L}_{\text{kd}} = \frac{1}{C}\sum_c \Big(\mathcal{D}_{\text{KL}}(\mathbf{q}_c^a||\mathbf{q}_c^b) + \mathcal{D}_{\text{KL}}(\mathbf{q}_c^b||\mathbf{q}_c^a)\Big),$$

$$\text{where } \mathcal{D}_{\text{KL}}(\mathbf{q}_c^a||\mathbf{q}_c^b) = \sum \mathbf{q}_c^a \log \frac{\mathbf{q}_c^a}{\mathbf{q}_c^b}.$$

*Q. Dou, Q. Liu et al. "Unpaired Multi-modal Segmentation via Knowledge Distillation", IEEE Trans. on Medical Imaging, 2020.*

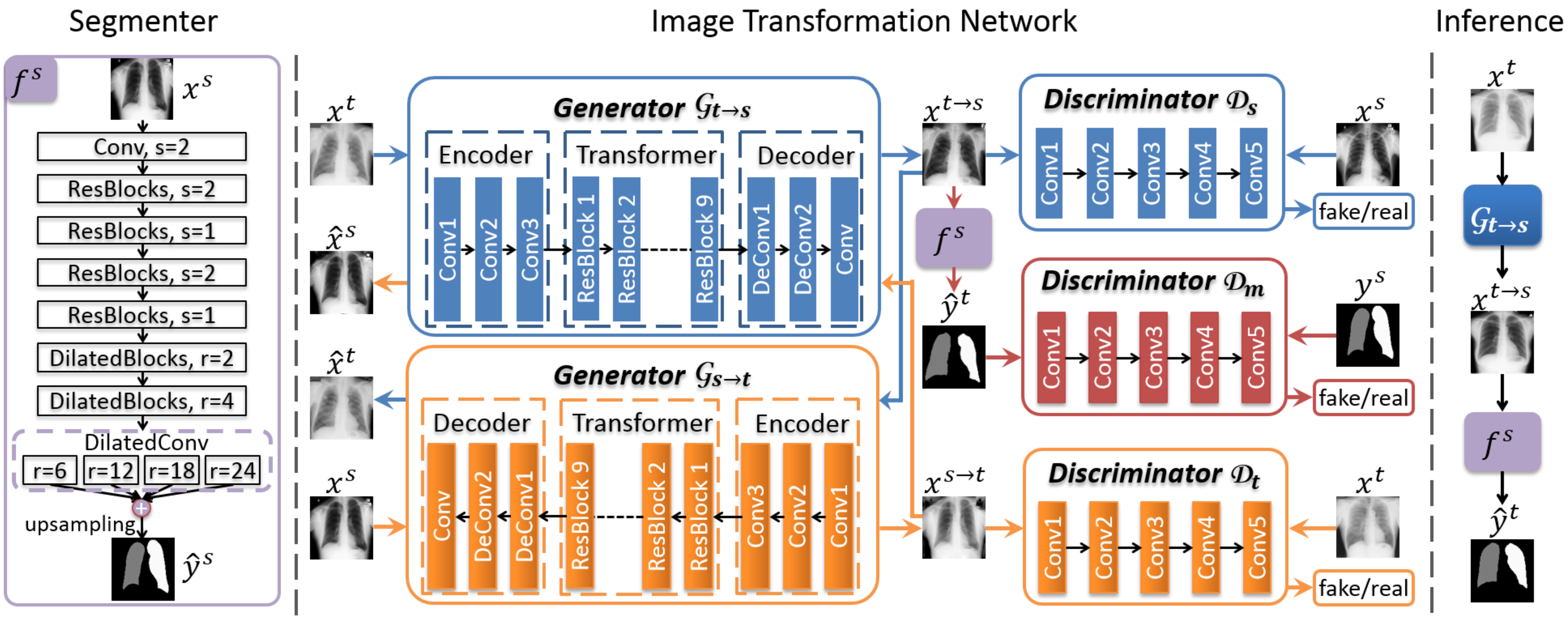
# Tackling Data Heterogeneity with Supervised Learning




*Q. Dou, Q. Liu et al. "Unpaired Multi-modal Segmentation via Knowledge Distillation", IEEE Trans. on Medical Imaging, 2020.*

# Tackling Data Heterogeneity with UDA

## Unsupervised domain adaptation through pixel-level alignment

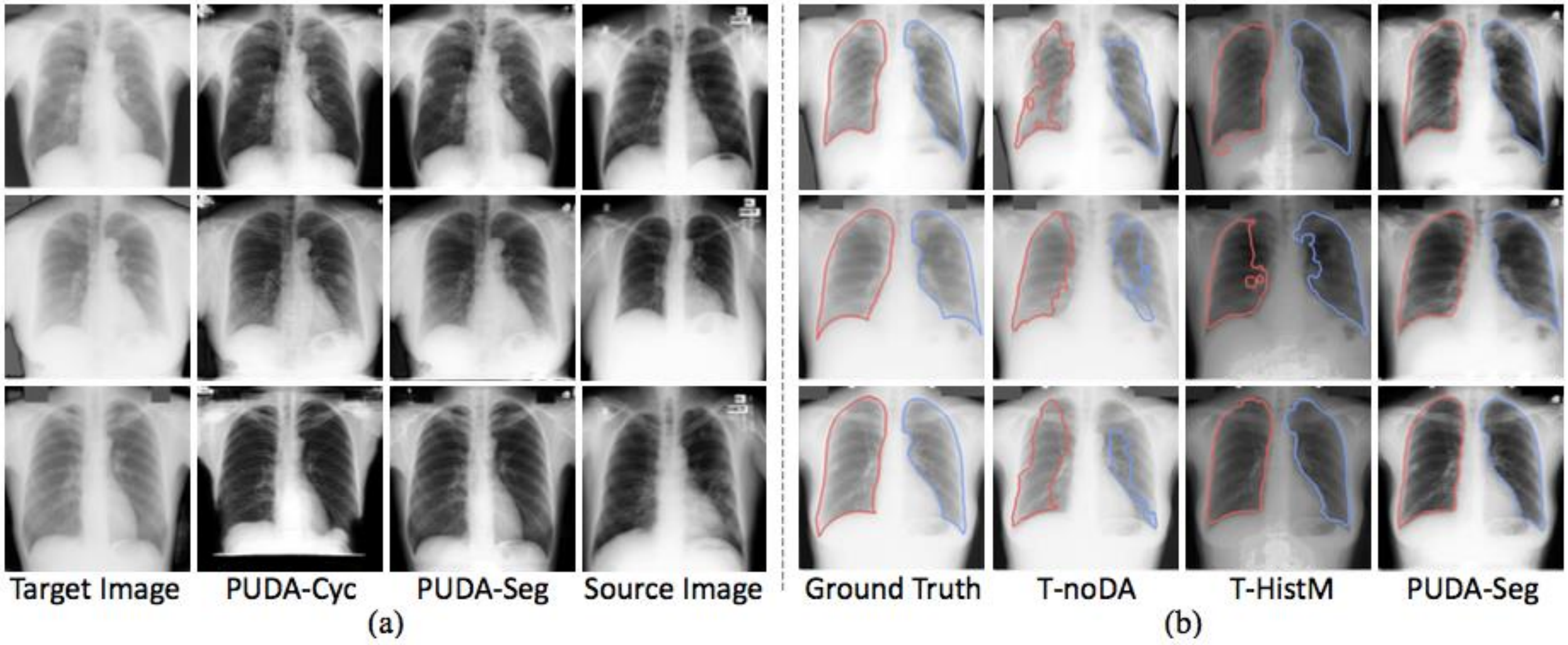



*C. Chen, Q. Dou, et al. "Semantic-Aware Generative Adversarial Nets for Unsupervised Domain Adaptation in Chest X-ray Segmentation." MICCAI-MLMI'18 (Oral)*

*Related work: [Y. Huo et al., ISBI 2018; Z. Zhang et al. CVPR 2018; Y. Zhang et al. MICCAI 2018]*

# Tackling Data Heterogeneity with UDA

## Image-to-image transformation with generative adversarial nets




*C. Chen, Q. Dou, et al. "Semantic-Aware Generative Adversarial Nets for Unsupervised Domain Adaptation in Chest X-ray Segmentation." MICCAI-MLMI'18 (Oral)*

*Related work: [Y. Huo et al., ISBI 2018; Z. Zhang et al. CVPR 2018; Y. Zhang et al. MICCAI 2018]*

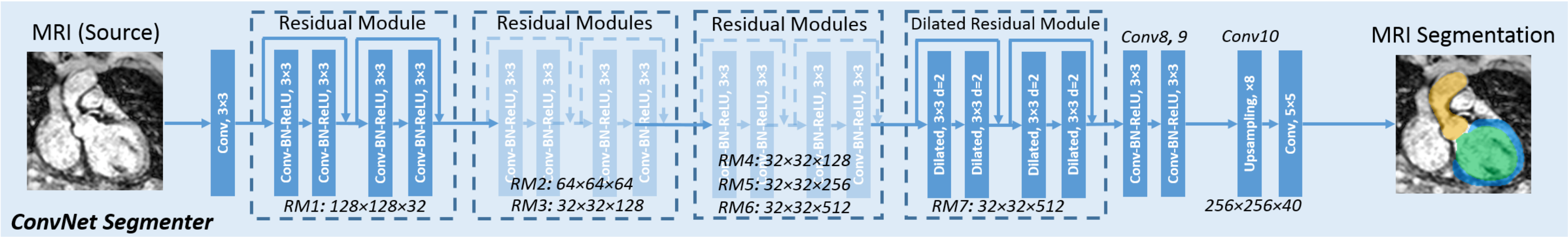
# Tackling Data Heterogeneity with UDA

## Unsupervised Domain Adaptation: Feature-level Alignment

**Train a source domain segmentation model**

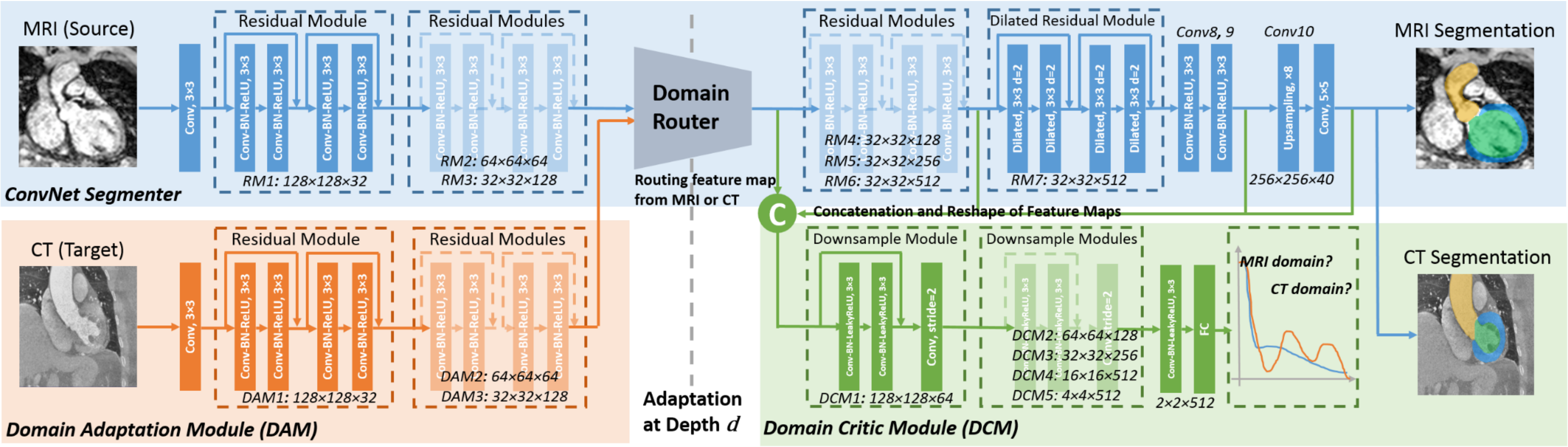
- joint cross-entropy loss and dice loss

$$\mathcal{L}_{\text{seg}} = -\sum_{i=1}^{N^s}\sum_{c\in C} w_c^s \cdot y_{i,c}^s \log(\hat{p}_{i,c}^s) - \lambda \sum_{c\in C} \frac{\sum_{i=1}^{N^s} 2y_{i,c}^s \hat{y}_{i,c}^s}{\sum_{i=1}^{N^s} y_{i,c}^s y_{i,c}^s + \sum_{i=1}^{N^s} \hat{y}_{i,c}^s \hat{y}_{i,c}^s}$$




*Q. Dou*, C. Ouyang*, et al. "Unsupervised Cross-Modality Domain Adaptation of ConvNets for Biomedical Image Segmentations with Adversarial Loss.." IJCAI 2018.*

# Unsupervised Domain Adaptation: Feature-level Alignment



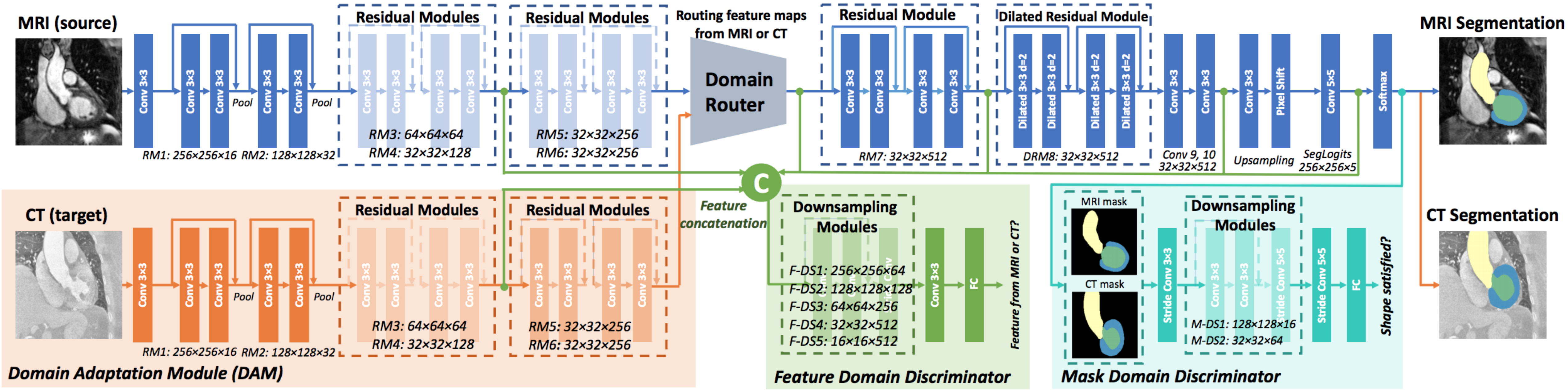

## Unsupervised learning with adversarial loss

domain adaptation module (generator): $\min\limits_{\mathcal{M}} \mathcal{L}_{\mathcal{M}}(X^t, \mathcal{D}) = -\mathbb{E}_{(\mathcal{M}_A(x^t), F_H(x^t)) \sim \mathbb{P}_g}[\mathcal{D}(\mathcal{M}_A(x^t), F_H(x^t))]$

domain critic module (discriminator): $\min\limits_{\mathcal{D}} \mathcal{L}_{\mathcal{D}}(X^s, X^t, \mathcal{M}) = \mathbb{E}_{(\mathcal{M}_A(x^t), F_H(x^t)) \sim \mathbb{P}_g}[\mathcal{D}(\mathcal{M}_A(x^t), F_H(x^t))] - \mathbb{E}_{(M_A^s(x^s), F_H(x^s)) \sim \mathbb{P}_s}[\mathcal{D}(M_A^s(x^s), F_H(x^s))], s.t. \|\mathcal{D}\|_{L \leq K}$

*Q. Dou*, C. Ouyang*, et al. "Unsupervised Cross-Modality Domain Adaptation of ConvNets for Biomedical Image Segmentations with Adversarial Loss.." IJCAI 2018.*

# Unsupervised Domain Adaptation: Feature-level Alignment




## Unsupervised learning with adversarial loss

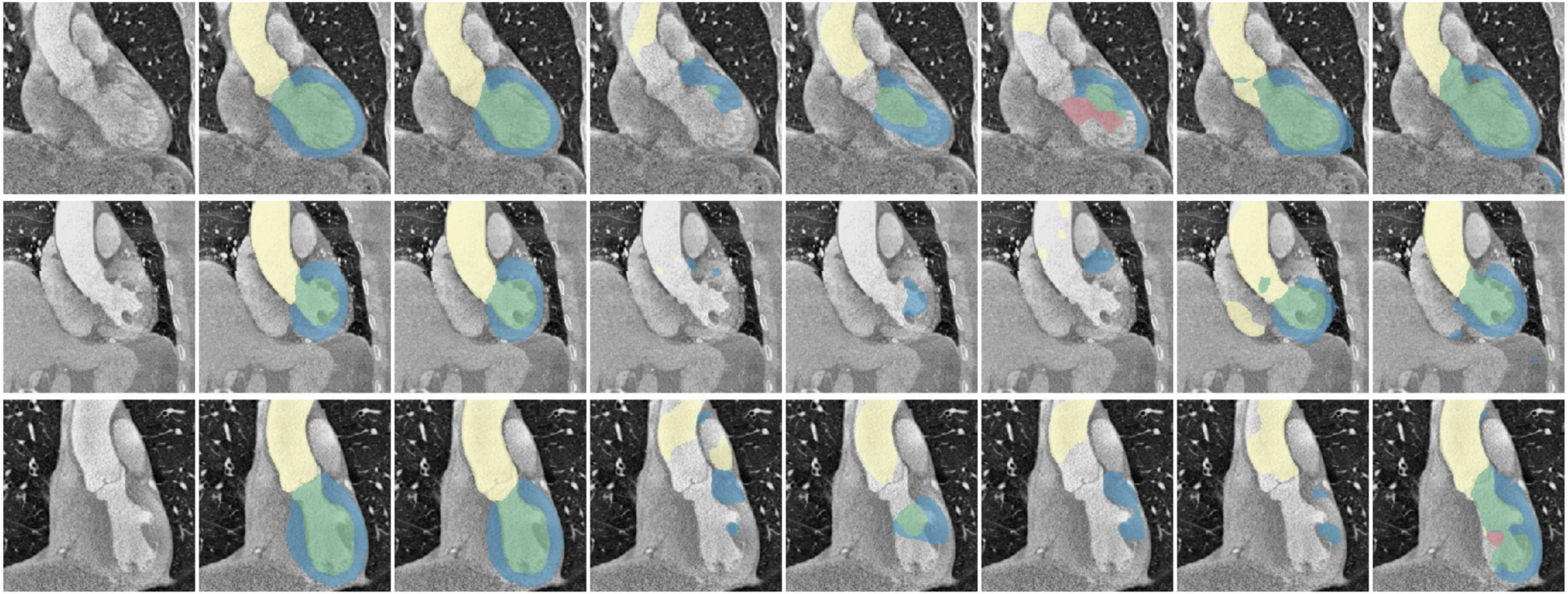
domain adaptation module (generator): $$\min_{\mathcal{M}} \mathcal{L}_{\mathcal{M}}(X^t, \mathcal{D}) = -\mathbb{E}_{(\mathcal{M}_A(x^t), F_H(x^t))\sim \mathbb{P}_g}[\mathcal{D}(\mathcal{M}_A(x^t), F_H(x^t))]$$

domain critic module (discriminator): $$\min_{\mathcal{D}} \mathcal{L}_{\mathcal{D}}(X^s, X^t, \mathcal{M}) = \mathbb{E}_{(\mathcal{M}_A(x^t), F_H(x^t))\sim \mathbb{P}_g}[\mathcal{D}(\mathcal{M}_A(x^t), F_H(x^t))] - \mathbb{E}_{(M_A^s(x^s), F_H(x^s))\sim \mathbb{P}_s}[\mathcal{D}(M_A^s(x^s), F_H(x^s))], s.t. \|\mathcal{D}\|_{L\leq K}$$

*Q. Dou*, C. Ouyang*, et al. "Unsupervised Cross-Modality Domain Adaptation of ConvNets for Biomedical Image Segmentations with Adversarial Loss.." IJCAI 2018.*

*Related work: [K Kamnitsas et al. IPMI 2017]*

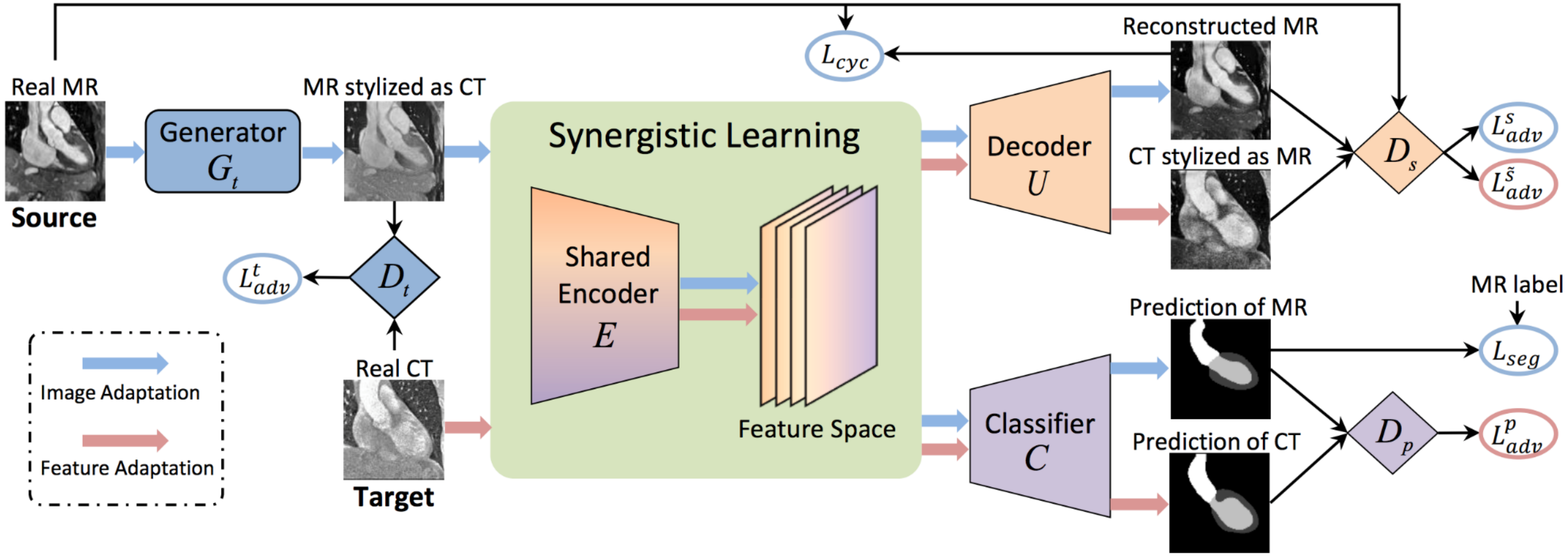
# Unsupervised Domain Adaptation: Feature-level Alignment

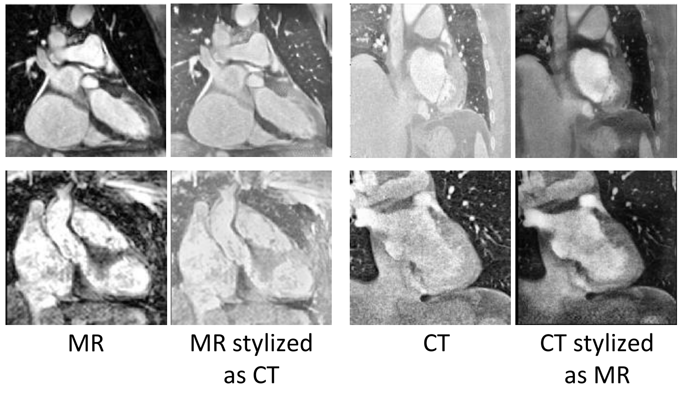


(a) CT image (b) CT label (c) Seg-CT (d) Seg-CT-noDA (e) DANN (f) ADDA (g) CycleGAN (h) Ours

*Related work: [DANN, Ganin et al. JMLR 2016; ADDA, Tzeng et al. CVPR 2017; CycleGAN, Zhu et al. ICCV 2017]*

# Unsupervised Domain Adaptation: Synergistic Alignment




| Methods | Adaptation | | Dice | | | | | ASD | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Image | Feature | AA | LAC | LVC | MYO | Average | AA | LAC | LVC | MYO | Average |
| W/o adaptation | | | 28.4 | 27.7 | 4.0 | 8.7 | 17.2 | 20.6 | 16.2 | N/A | 48.4 | N/A |
| DANN (Ganin et al. 2016) | | ✓ | 39.0 | 45.1 | 28.3 | 25.7 | 34.5 | 16.2 | 9.2 | 12.1 | 10.1 | 11.9 |
| ADDA (Tzeng et al. 2017) | | ✓ | 47.6 | 60.9 | 11.2 | 29.2 | 37.2 | 13.8 | 10.2 | N/A | 13.4 | N/A |
| CycleGAN (Zhu et al. 2017) | ✓ | | 73.8 | 75.7 | 52.3 | 28.7 | 57.6 | 11.5 | 13.6 | 9.2 | 8.8 | 10.8 |
| CyCADA (Hoffman et al. 2018) | ✓ | ✓ | 72.9 | **77.0** | 62.4 | 45.3 | 64.4 | **9.6** | 8.0 | 9.6 | 10.5 | 9.4 |
| Dou et al. (Dou et al. 2018) | | ✓ | 74.8 | 51.1 | 57.2 | 47.8 | 57.7 | 27.5 | 20.1 | 29.5 | 31.2 | 27.1 |
| Joyce et al. (Joyce et al. 2018) | | ✓ | - | - | 66 | 44 | - | - | - | - | - | - |
| SIFA (Ours) | ✓ | ✓ | **81.1** | 76.4 | **75.7** | **58.7** | **73.0** | 10.6 | **7.4** | **6.7** | **7.8** | **8.1** |

*C. Chen, Q. Dou et al. "Synergistic Image and Feature Adaptation: Towards Cross-Modality Domain Adaptation for Medical Image Segmentation", AAAI, 2019. (Oral)*

*Related work: [DANN, Ganin et al. JMLR 2016; ADDA, Tzeng et al. CVPR 2017; CycleGAN, Zhu et al. ICCV 2017]*

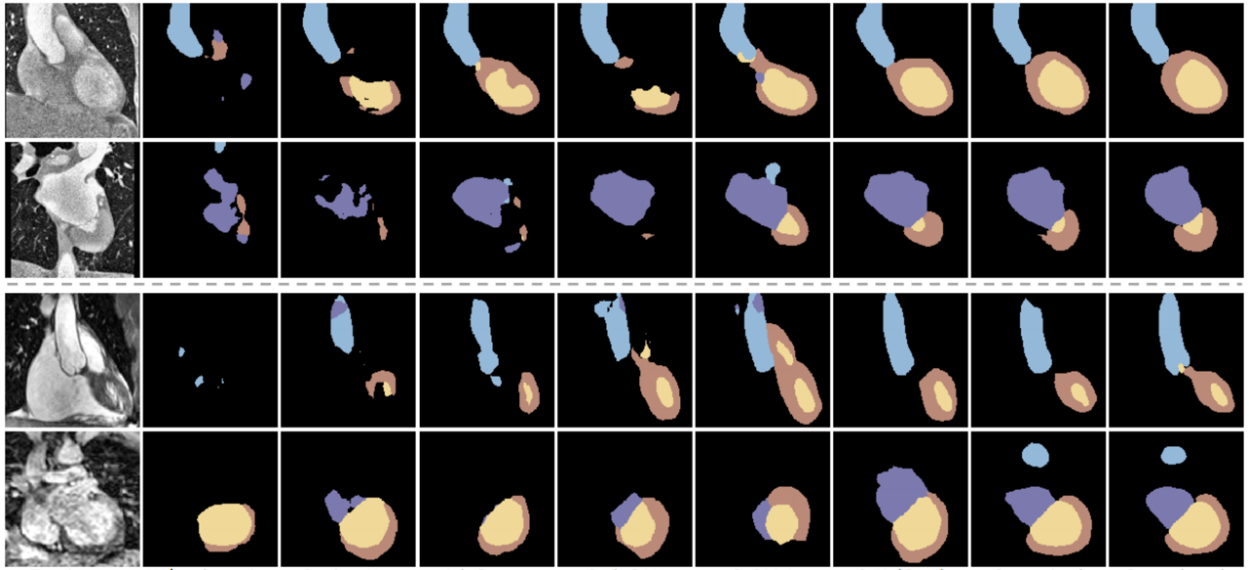
# Bidirectional Adaptation via Deeply Supervised SIFA

Cardiac MRI → Cardiac CT

| Method | Dice | | | | | ASD | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | AA | LAC | LVC | MYO | Average | AA | LAC | LVC | MYO | Average |
| Supervised training | 92.7 | 91.1 | 91.9 | 87.7 | 90.9 | 1.5 | 3.5 | 1.7 | 2.1 | 2.2 |
| W/o adaptation | 28.4 | 27.7 | 4.0 | 8.7 | 17.2 | 20.6 | 16.2 | N/A | 48.4 | N/A |
| PnP-AdaNet [44] | 74.0 | 68.9 | 61.9 | 50.8 | 63.9 | 12.8 | 6.3 | 17.4 | 14.7 | 12.8 |
| SynSeg-Net [37] | 71.6 | 69.0 | 51.6 | 40.8 | 58.2 | 11.7 | 7.8 | 7.0 | 9.2 | 8.9 |
| AdaOutput [15] | 65.2 | 76.6 | 54.4 | 43.6 | 59.9 | 17.9 | **5.5** | 5.9 | 8.9 | 9.6 |
| CycleGAN [8] | 73.8 | 75.7 | 52.3 | 28.7 | 57.6 | 11.5 | 13.6 | 9.2 | 8.8 | 10.8 |
| CyCADA [18] | 72.9 | 77.0 | 62.4 | 45.3 | 64.4 | 9.6 | 8.0 | 9.6 | 10.5 | 9.4 |
| Prior SIFA [20] | 81.1 | 76.4 | **75.7** | 58.7 | 73.0 | 10.6 | 7.4 | 6.7 | **7.8** | 8.1 |
| SIFA (Ours) | **81.3** | **79.5** | 73.8 | **61.6** | **74.1** | **7.9** | 6.2 | **5.5** | 8.5 | **7.0** |

Cardiac CT → Cardiac MRI

| Method | Dice | | | | | ASD | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | AA | LAC | LVC | MYO | Average | AA | LAC | LVC | MYO | Average |
| Supervised training | 82.8 | 80.5 | 92.4 | 78.8 | 83.6 | 3.6 | 3.9 | 2.1 | 1.9 | 2.9 |
| W/o adaptation | 5.4 | 30.2 | 24.6 | 2.7 | 15.7 | 15.4 | 16.8 | 13.0 | 10.8 | 14.0 |
| PnP-AdaNet [44] | 43.7 | 47.0 | 77.7 | **48.6** | 54.3 | 11.4 | 14.5 | 4.5 | 5.3 | 8.9 |
| SynSeg-Net [37] | 41.3 | 57.5 | 63.6 | 36.5 | 49.7 | 8.6 | 10.7 | 5.4 | 5.9 | 7.6 |
| AdaOutput [15] | 60.8 | 39.8 | 71.5 | 35.5 | 51.9 | **5.7** | 8.0 | 4.6 | 4.6 | **5.7** |
| CycleGAN [8] | 64.3 | 30.7 | 65.0 | 43.0 | 50.7 | 5.8 | 9.8 | 6.0 | 5.0 | 6.6 |
| CyCADA [18] | 60.5 | 44.0 | 77.6 | 47.9 | 57.5 | 7.7 | 13.9 | 4.8 | 5.2 | 7.9 |
| Prior SIFA [20] | **67.0** | 60.7 | 75.1 | 45.8 | 62.1 | 6.2 | 9.8 | 4.4 | **4.4** | 6.2 |
| SIFA (Ours) | 65.3 | **62.3** | **78.9** | 47.3 | **63.4** | 7.3 | **7.4** | **3.8** | **4.4** | **5.7** |

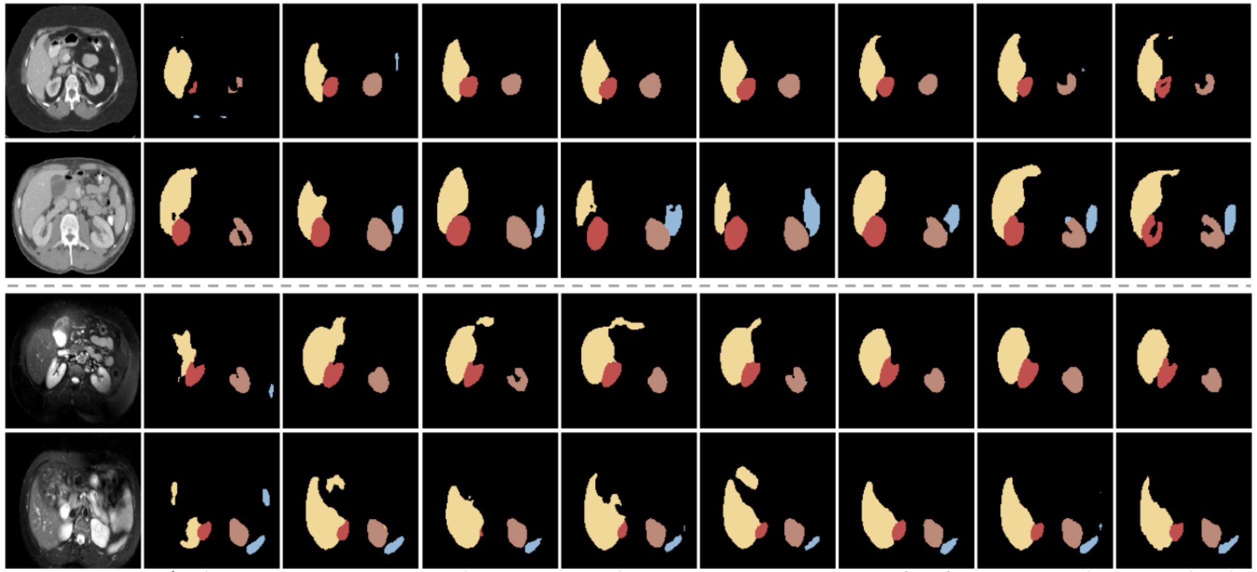



Abdominal MRI → Abdominal CT

| Method | Dice | | | | | ASD | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Liver | R. kidney | L. kidney | Spleen | Average | Liver | R. kidney | L. kidney | Spleen | Average |
| Supervised training | 92.8 | 86.4 | 87.4 | 88.2 | 88.7 | 1.0 | 1.8 | 0.9 | 1.2 | 1.2 |
| W/o adaptation | 73.1 | 47.3 | 57.3 | 55.1 | 58.2 | 2.9 | 5.6 | 7.7 | 7.4 | 5.9 |
| SynSeg-Net [37] | 85.0 | 82.1 | 72.7 | 81.0 | 80.2 | 2.2 | 1.3 | 2.1 | 2.0 | 1.9 |
| AdaOutput [15] | 85.4 | 79.7 | 79.7 | 81.7 | 81.6 | 1.7 | 1.2 | 1.8 | **1.6** | 1.6 |
| CycleGAN [8] | 83.4 | 79.3 | 79.4 | 77.3 | 79.9 | 1.8 | 1.3 | **1.2** | 1.9 | 1.6 |
| CyCADA [18] | 84.5 | 78.6 | 80.3 | 76.9 | 80.1 | 2.6 | 1.4 | 1.3 | 1.9 | 1.8 |
| Prior SIFA [20] | 87.9 | **83.7** | 80.1 | 80.5 | 83.1 | 2.1 | 1.1 | 1.6 | 1.8 | 1.6 |
| SIFA (Ours) | **88.0** | 83.3 | **80.9** | **82.6** | **83.7** | **1.2** | **1.0** | 1.5 | **1.6** | **1.3** |

Abdominal CT → Abdominal MRI

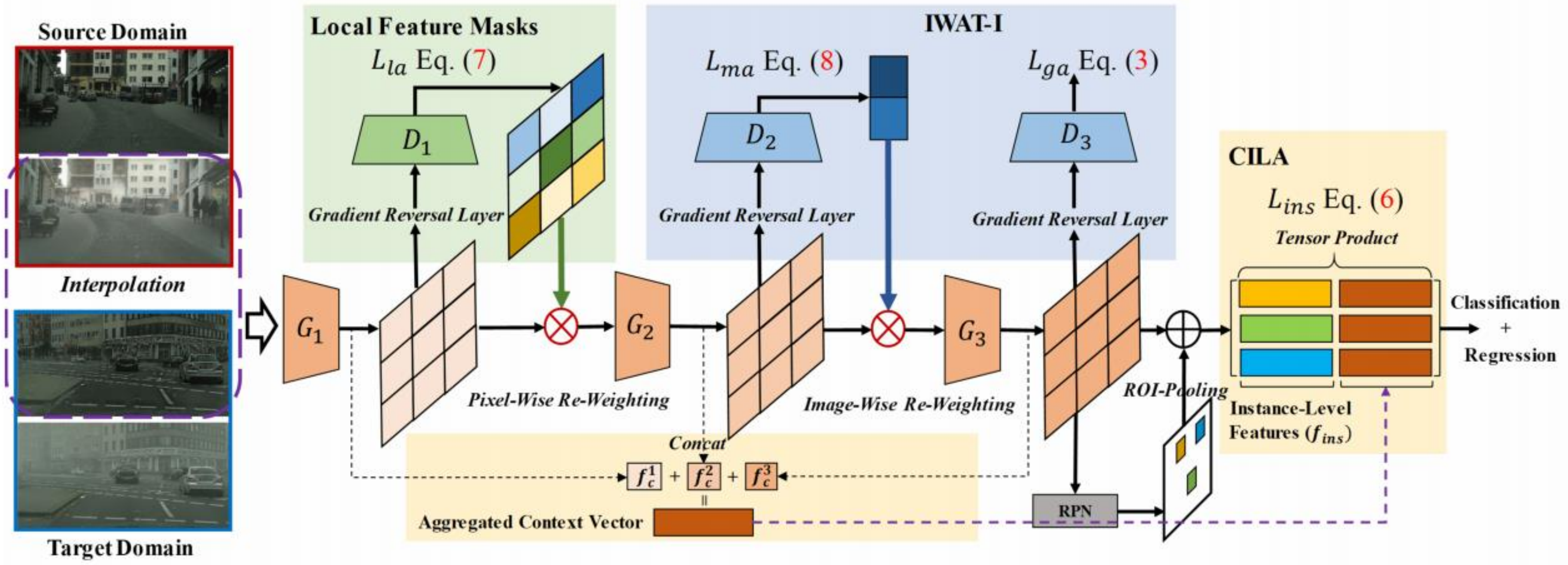
| Method | Dice | | | | | ASD | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Liver | R. kidney | L. kidney | Spleen | Average | Liver | R. kidney | L. kidney | Spleen | Average |
| Supervised training | 92.0 | 91.1 | 80.6 | 85.7 | 87.3 | 1.3 | 2.0 | 1.5 | 1.3 | 1.5 |
| W/o adaptation | 48.9 | 50.9 | 65.3 | 65.7 | 57.7 | 4.5 | 12.3 | 6.8 | 4.5 | 7.0 |
| SynSeg-Net [37] | 87.2 | **90.2** | 76.6 | 79.6 | 83.4 | 2.8 | 0.7 | 4.8 | 2.5 | 2.7 |
| AdaOutput [15] | 85.8 | 89.7 | 76.3 | 82.2 | 83.5 | 1.9 | 1.4 | 3.0 | 1.8 | 2.1 |
| CycleGAN [8] | 88.8 | 87.3 | 76.8 | 79.4 | 83.1 | 2.0 | 3.2 | 1.9 | 2.6 | 2.4 |
| CyCADA [18] | 88.7 | 89.3 | 78.1 | 80.2 | 84.1 | **1.5** | 1.7 | **1.3** | **1.6** | **1.5** |
| Prior SIFA [20] | 88.5 | 90.0 | 79.7 | 81.3 | 84.9 | 2.3 | 0.9 | 1.4 | 2.4 | 1.7 |
| SIFA (Ours) | **90.0** | 89.1 | **80.2** | **82.3** | **85.4** | **1.5** | **0.6** | 1.5 | 2.4 | **1.5** |



# Harmonizing Transferability and Discriminability for Adapting

## HTCN: Hierarchical Transferability Calibration Network

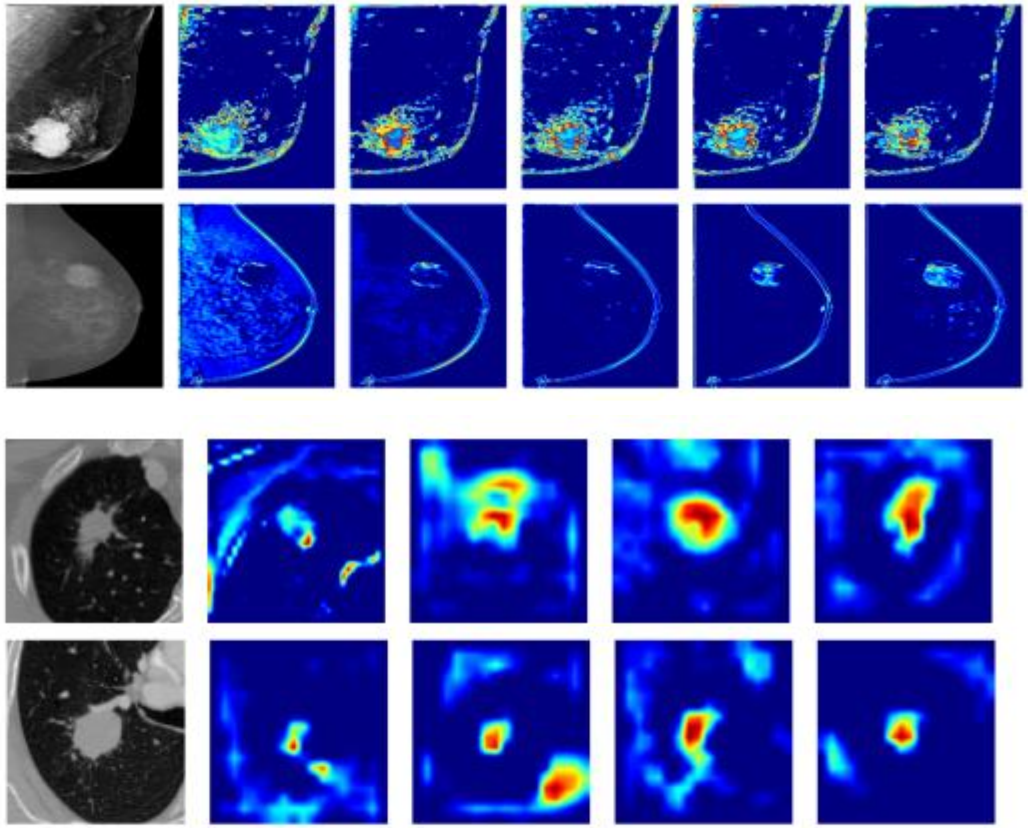
- transferability and discriminability may come at a contradiction given the complex combinations of objects
- hierarchically (local-region/image/instance) calibrates the transferability of feature representations




*C. Chen et al. "Harmonizing Transferability and Discriminability for Adapting Object Detectors.." CVPR 2020.*

# Harmonizing Transferability and Discriminability for Adapting

## HTCN: Hierarchical Transferability Calibration Network

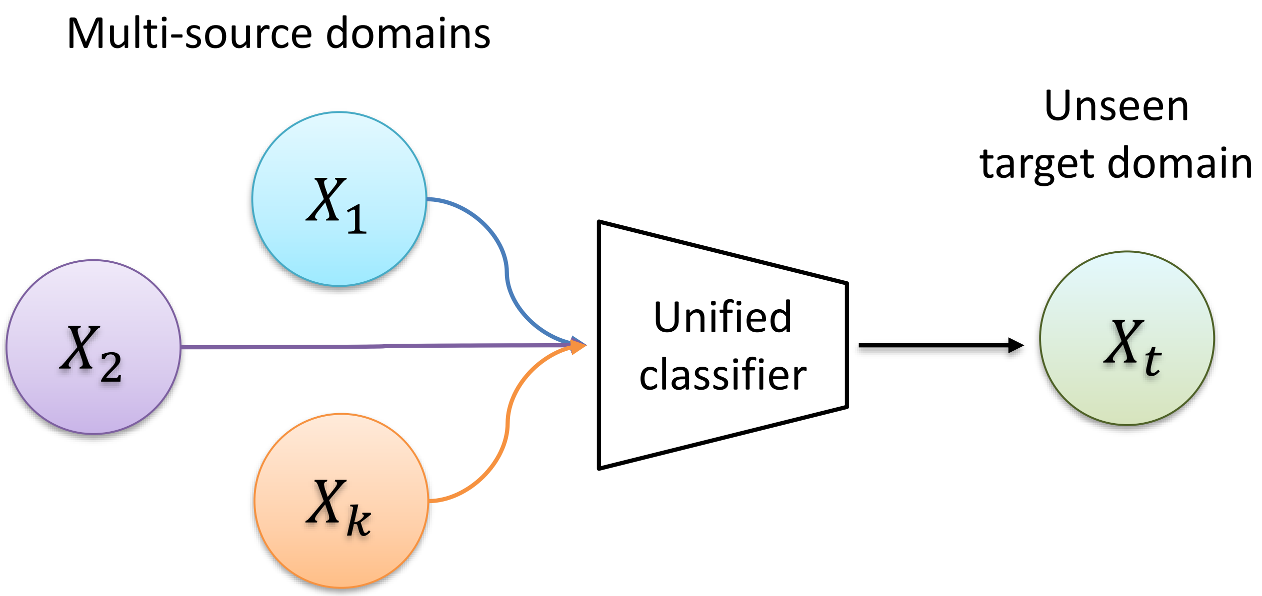



## Domain Generalization

***Problem setting:*** train on multiple source domains and **directly** generalize to unseen domains

### Regularization for generic semantic features

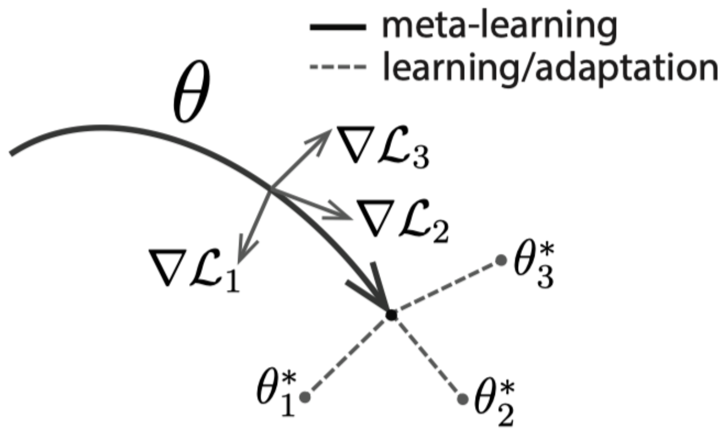
- adversarial feature alignment for domain invariance [Li et al. ECCV 2018]
- decompose networks parameters to domain-specific/invariant [Khosla ECCV 2012]
- data augmentation based methods [Shankar et al. ICLR 2018; Volpi et al. NeurIPS 2018]
- multi-task or self-supervised signals [Ghifary et al. ICCV 2015; Carlucci et al. CVPR 2019]



# Tackling Data Heterogeneity for Domain Generalization

## Domain Generalization with Gradient-based Meta-learning

**Model-agnostic learning:** MAML (model-agnostic meta-learning) [Finn et al. ICML 2017]




***<u>Applying to domain generalization:</u>***

- MLDG: directly applying episodic training paradigm [Li et al. AAAI 2018]
- MetaReg: meta-learning of weights regularization term [Balaji et al. NeurIPS 2018]
- Episodic training with alternative model updates [Li et al. ICCV 2019]

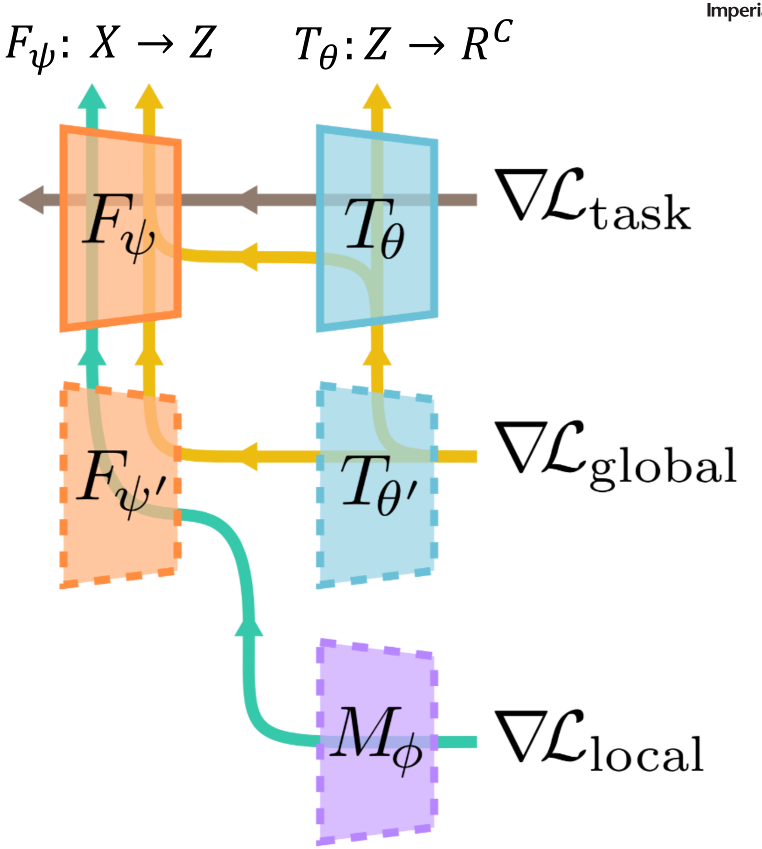
*Q. Dou, D. Castro, et al. "Domain Generalization via Model-Agnostic Learning of Semantic Features", NeurIPS, 2019.*

# MASF: Model-Agnostic Learning of Semantic Features

## Episodic training paradigm

Available domains: $D = \{D_1, D_2, \ldots, D_K\}$

Neural network is composed of: $F_\psi \circ T_\theta$




***<u>Learning with explicit simulation of domain shift:</u>***

At each iteration, split into meta-train $D_{tr}$ and meta-test $D_{te}$

Update the parameters one or more steps with gradient descent:

$$(\psi', \theta') = (\psi, \theta) - \alpha \nabla_{\psi,\theta} \mathcal{L}_{\text{task}}(\mathcal{D}_{\text{tr}}; \psi, \theta)$$

Then, apply meta-learning step, to enforce certain properties to be exhibited on held-out domain $D_{te}$, to regularize semantic features

*Q. Dou, D. Castro, et al. "Domain Generalization via Model-Agnostic Learning of Semantic Features", NeurIPS, 2019.*

## MASF: Model-Agnostic Learning of Semantic Features

### Global Class Alignment

Inter-class relationships concept is domain-invariant and transferable

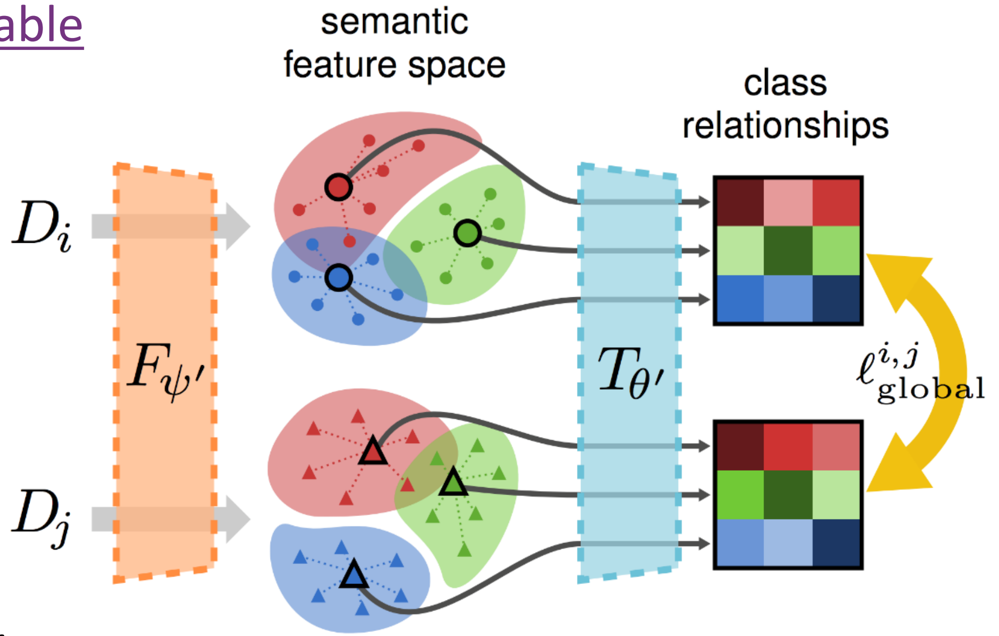
- In each domain, compute class-specific mean feature vector:

$$\bar{\mathbf{z}}_c^{(k)} = \frac{1}{N_k^{(c)}} \sum_{n:y_n^{(k)}=c} F_{\psi'}(\mathbf{x}_n^{(k)}) \approx \mathbb{E}_{D_k}[F_{\psi'}(\mathbf{x}) \,|\, y=c]$$

- Compute soft label distribution: $\mathbf{s}_c^{(k)} = \mathrm{softmax}(T_{\theta'}(\bar{\mathbf{z}}_c^{(k)})/\tau)$

- With $(D_i, D_j) \in D_{tr} \times D_{te}$, regularize consistency of inter-class alignment:

$$\ell_{\mathrm{global}}(D_i, D_j; \psi', \theta') = \frac{1}{C} \sum_{c=1}^{C} \frac{1}{2}[D_{\mathrm{KL}}(\mathbf{s}_c^{(i)} \,\|\, \mathbf{s}_c^{(j)}) + D_{\mathrm{KL}}(\mathbf{s}_c^{(j)} \,\|\, \mathbf{s}_c^{(i)})]$$




*(Note: complexity of pairs is controllable via mini-batch sampling in large-scale scenarios.)*

## Local Sample Clustering

feature clusters with domain-independent class-specific cohesion and separation

Use a *metric-learning* approach, with an embedding network and operates in semantic feature space:

- obtain a learnable distance function:

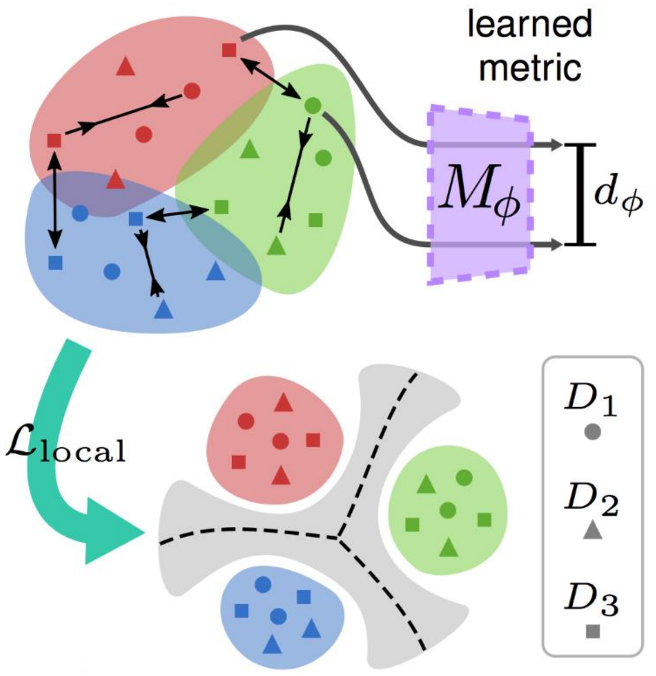
$$d_\phi(\mathbf{z}_n, \mathbf{z}_m) = \|\mathbf{e}_n - \mathbf{e}_m\|_2 = \|M_\phi(\mathbf{z}_n) - M_\phi(\mathbf{z}_m)\|_2$$

- metric-learning can rely on contrastive loss [Hadsell et al. CVPR 2006]:

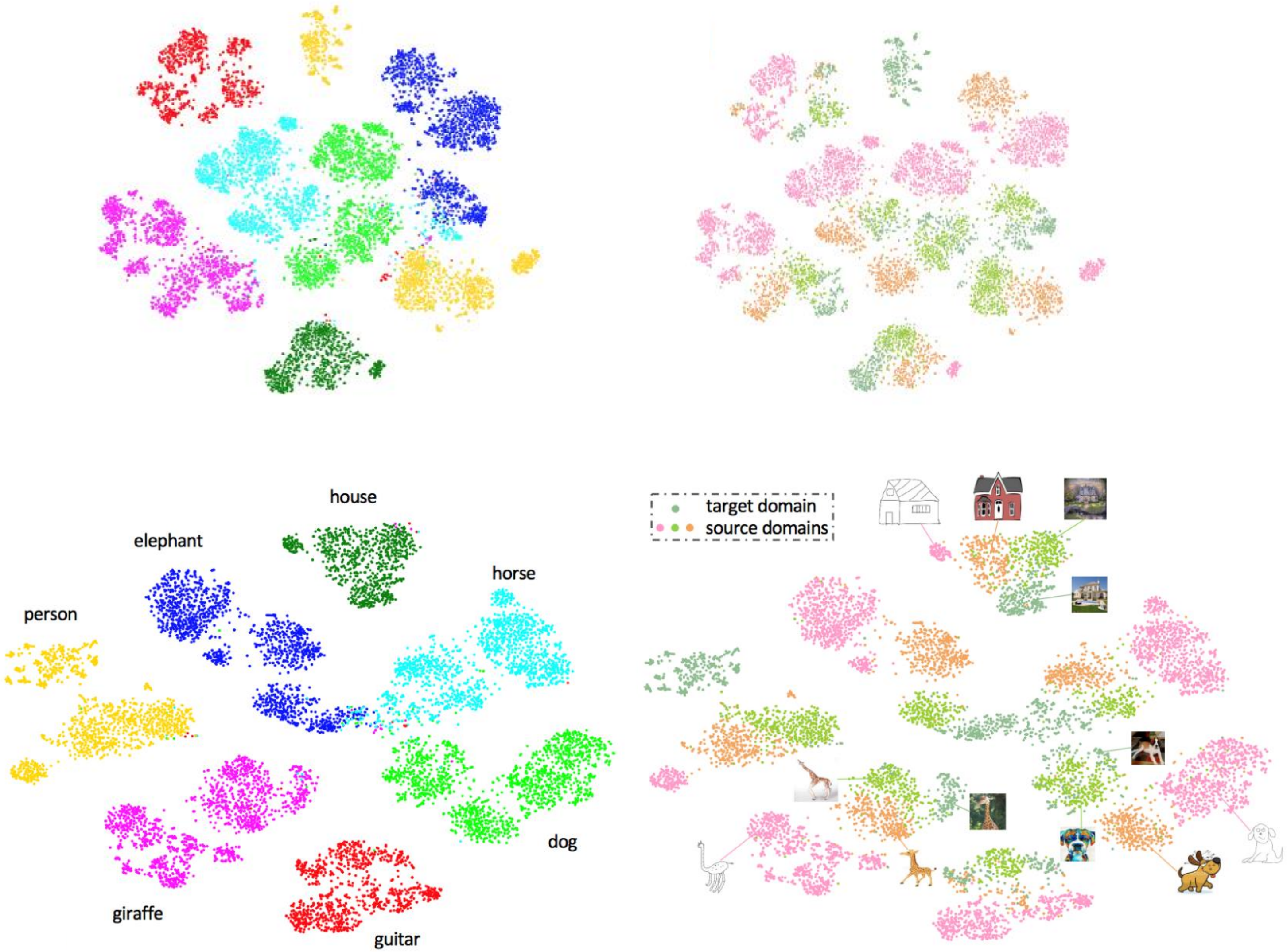
$$\ell_{\text{con}}^{n,m} = \begin{cases} d_\phi(\mathbf{z}_n, \mathbf{z}_m)^2\,, & \text{if } y_n = y_m \\ (\max\{0,\ \xi - d_\phi(\mathbf{z}_n, \mathbf{z}_m)\})^2\,, & \text{if } y_n \neq y_m \end{cases}$$

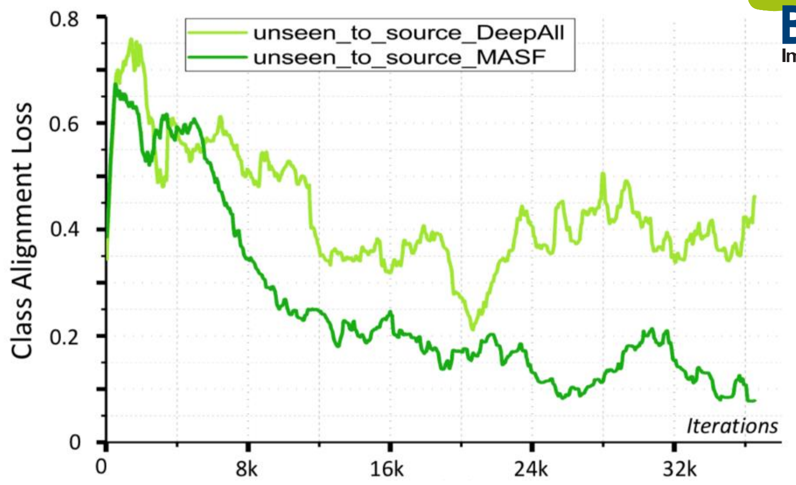
- or triplet loss [Schroff et al. CVPR 2015]:

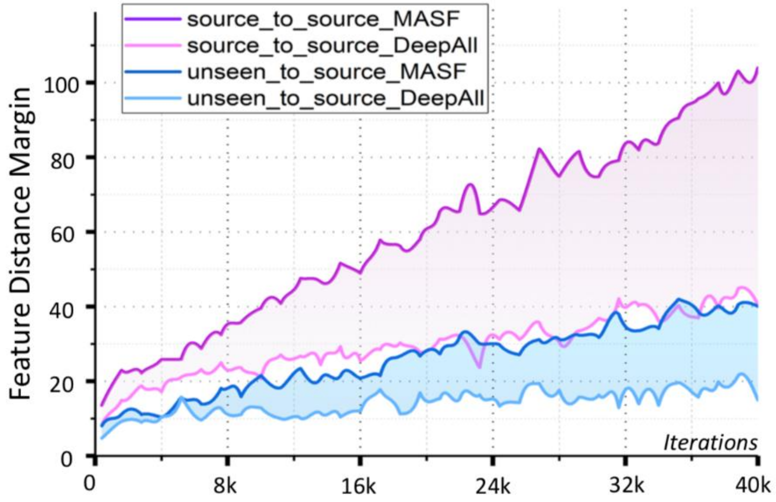
$$\ell_{\text{tri}}^{a,p,n} = \max\{0,\ d_\phi(\mathbf{z}_a, \mathbf{z}_p)^2 - d_\phi(\mathbf{z}_a, \mathbf{z}_n)^2 + \xi\}$$



# MASF: Model-Agnostic Learning of Semantic Features



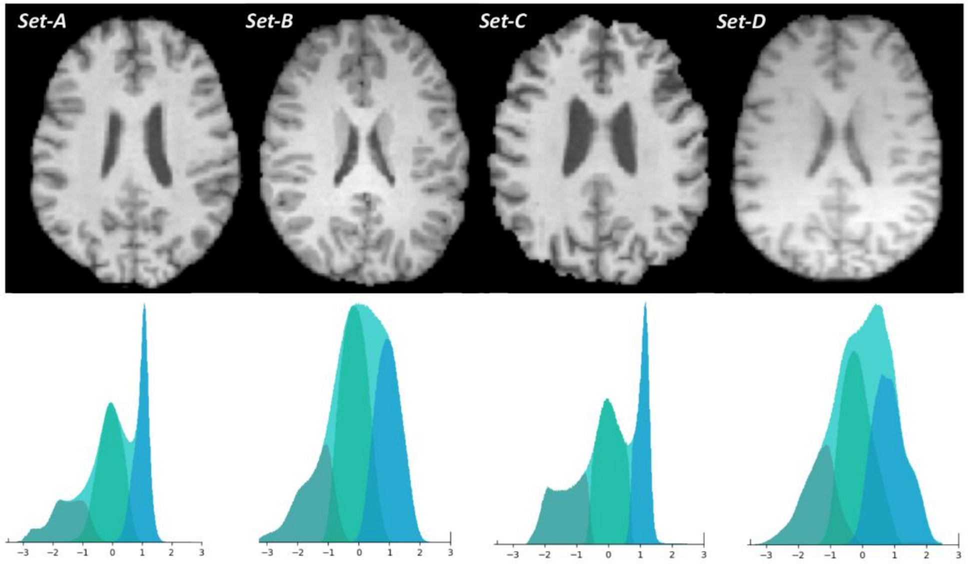



# MASF: Model-Agnostic Learning of Semantic Features

## Medical application of brain tissue segmentation

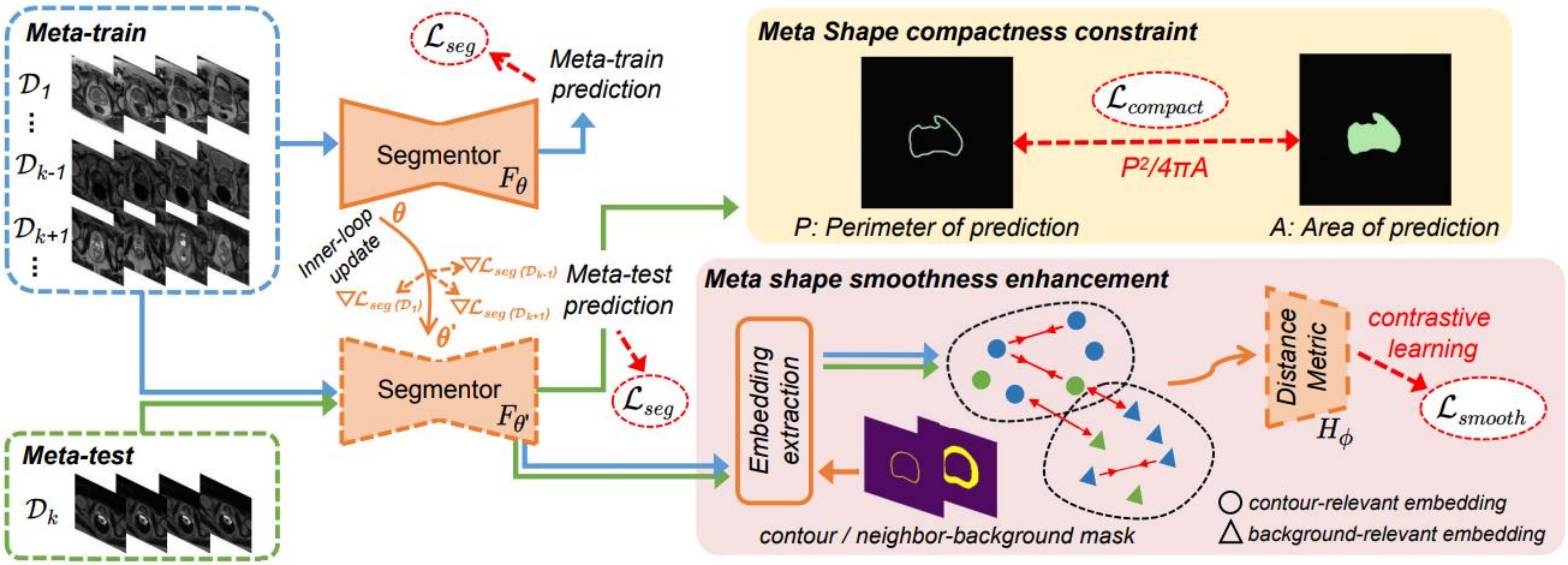
- data acquisition differences in scanners, imaging protocols, and many other factors
- posing severe limitations for translating learning-based methods in clinical practice
- segmentation of 3 brain tissues: white matter, gray matter and cerebrospinal fluid
- 4 domains corresponding to 4 hospitals




| Test \ Train | Set-A | Set-B | Set-C | Set-D | DeepAll | MASF |
|---|---|---|---|---|---|---|
| Set-A | 90.62 | 88.91 | 88.81 | 85.03 | 89.09 | 89.82 |
| Set-B | 85.03 | 94.22 | 81.38 | 88.31 | 90.41 | 91.71 |
| Set-C | 93.14 | 92.80 | 95.40 | 88.68 | 94.30 | 94.50 |
| Set-D | 76.32 | 88.39 | 73.50 | 94.29 | 88.62 | 89.51 |

## Shape-awareness in MASF scheme for segmentation tasks

- Encourage complete segmentation shape at domain shift
- Learn domain-invariant contour-relevant and background-relevant embedding

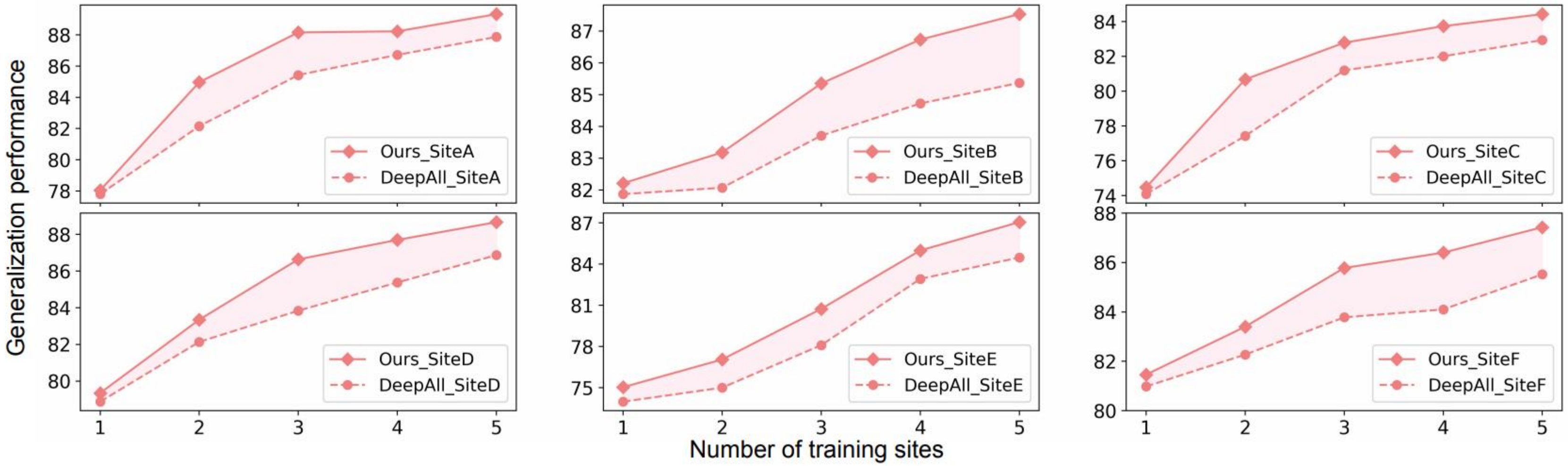



Q. Liu, Q. Dou, P. A. Heng. "Shape-aware Meta-learning for Generalizing Prostate MRI Segmentation to Unseen Domains", MICCAI, 2020.

## Experimental results with prostate MRI segmentation

| Method | Site A | | Site B | | Site C | | Site D | | Site E | | Site F | | Average | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intra-site | 89.27 | 1.41 | <u>88.17</u> | 1.35 | <u>88.29</u> | 1.56 | 83.23 | 3.21 | 83.67 | 2.93 | 85.43 | 1.91 | 86.34 | 2.06 |
| DeepAll (baseline) | 87.87 | 2.05 | 85.37 | 1.82 | 82.94 | 2.97 | 86.87 | 2.25 | 84.48 | 2.18 | 85.58 | 1.82 | 85.52 | 2.18 |
| Epi-FCR [19] | 88.35 | 1.97 | 85.83 | 1.73 | 82.56 | 2.99 | 86.97 | 2.05 | 85.03 | 1.89 | 85.66 | 1.76 | 85.74 | 2.07 |
| LatReg [2] | 88.17 | 1.95 | 86.65 | 1.53 | 83.37 | 2.91 | 87.27 | 2.12 | 84.68 | 1.93 | 86.28 | 1.65 | 86.07 | 2.01 |
| BigAug [35] | 88.62 | 1.70 | 86.22 | 1.56 | 83.76 | 2.72 | 87.35 | 1.98 | 85.53 | 1.90 | 85.83 | 1.75 | 86.21 | 1.93 |
| MASF [6] | 88.70 | 1.69 | 86.20 | 1.54 | 84.16 | 2.39 | 87.43 | 1.91 | 86.18 | 1.85 | 86.57 | 1.47 | 86.55 | 1.81 |
| Plain meta-learning | 88.55 | 1.87 | 85.92 | 1.61 | 83.60 | 2.52 | 87.52 | 1.86 | 85.39 | 1.89 | 86.49 | 1.63 | 86.24 | 1.90 |
| + $\mathcal{L}_{compact}$ | 89.08 | 1.61 | 87.11 | 1.49 | 84.02 | 2.47 | 87.96 | 1.64 | 86.23 | 1.80 | 87.19 | 1.32 | 86.93 | 1.72 |
| + $\mathcal{L}_{smooth}$ **(Ours)** | **89.66** | **1.38** | **87.53** | **1.46** | **84.43** | **2.07** | **88.67** | **1.56** | **87.37** | **1.77** | **88.34** | **1.22** | **87.67** | **1.58** |

***Influence of training domain numbers on generalization***




*Q. Liu, Q. Dou, P. A. Heng. "Shape-aware Meta-learning for Generalizing Prostate MRI Segmentation to Unseen Domains", MICCAI, 2020.*

# Datasets

1. Multi-Modality Whole Heart Segmentation (MMWHS) Challenge
   http://www.sdspeople.fudan.edu.cn/zhuangxiahai/0/mmwhs/
   https://github.com/carrenD/Medical-Cross-Modality-Domain-Adaptation

2. MICCAI 2019 MS-CMRSeg Multi-sequence Cardiac MR Segmentation Challenge
   https://zmiclab.github.io/mscmrseg19/

3. MICCAI iSeg 2019 Challenge 6-month Infant Brain MRI segmentation from Multiple Sites
   http://iseg2019.web.unc.edu

4. ISBI 2019 CHAOS Challenge CT-MRI Abdominal Multi-Organ Segmentation
   https://chaos.grand-challenge.org

5. Prostate Segmentation, with several public datasets,
   i.e., NCI-ISBI 2013 dataset, I2CVB dataset (include multiple sites), PROMISE12 dataset (include multiple sites)

6. Chest X-Ray, with several public datasets,
   i.e., ChestX-ray14 NIH, CheXpert, PadChest, Mimic-CXR
   MIDL 2019: https://openreview.net/forum?id=S1gvm2E-t4

**Paper & Code Available at:**

# Acknowledgement

# Thanks for your attention!
# Q & A