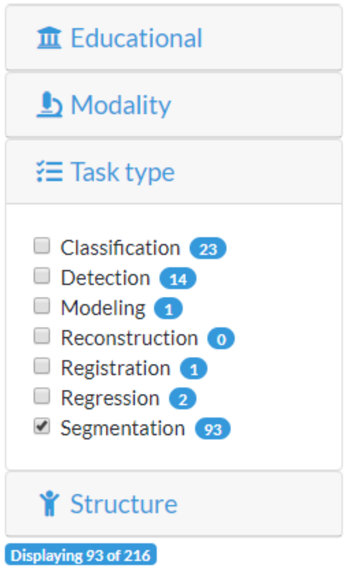
Deep Model Generalization in Medical Image Segmentation

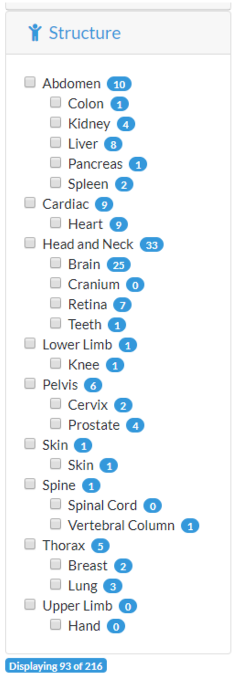
DOU Qi

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

Medical Image Segmentation at a Glance

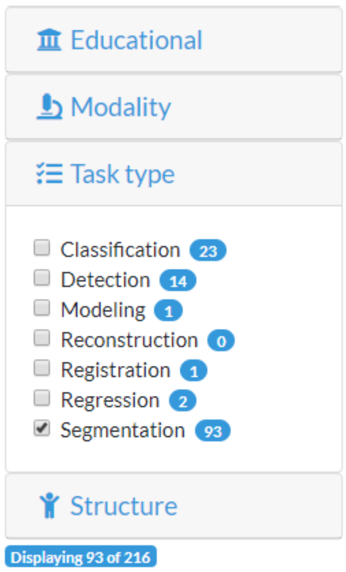


https://grand-challenge.org/

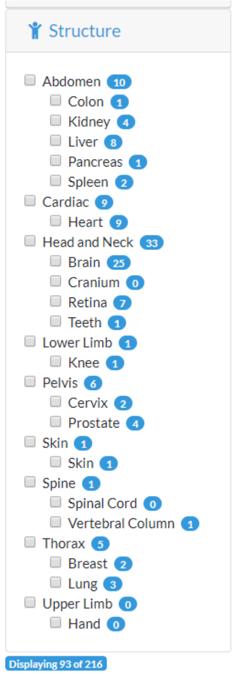


	Normal Tissues	Abnormal Tissues
Scenarios	organs (whole/substructure), vessels, cells	tumors, lesions, cancerous cells
Clinical Relevance	quantification (volume), visualization, intra-operative navigation, radiotherapy (organs at risk), clinical-oriented analysis	tumor quantification, diagnosis, prognosis, biopsy, surgical planning, radiotherapy (GTV/CTV), monitoring, radiomics
Challenges	limited contrast, multi-class problems, variances in shape for some large organs, cell overlapping, endoscopy artefacts	small scale, sample imbalance, ambiguous boundaries, various context, irregular shape, limited visibility with specific applications
Common Issues	Data perspective: high-dimensional, multi-modal, labor-intensive annotations Value perspective: speed and accuracy, a prerequisite for following-up tasks	

Medical Image Segmentation at a Glance



https://grand-challenge.org/



	Normal Tissues	Abnormal Tissues
Scenarios	organs (whole/substructure), vessels, cells	tumors, lesions, cancerous cells
Clinical Relevance	quantification (volume), visualization, intra-operative navigation, radiotherapy (organs at risk), clinical-oriented analysis	tumor quantification, diagnosis, prognosis, biopsy, surgical planning, radiotherapy (GTV/CTV), monitoring, radiomics
Challenges	limited contrast, multi-class problems, variances in shape for some large organs, cell overlapping, endoscopy artefacts	small scale, sample imbalance, ambiguous boundaries, various context, irregular shape, limited visibility with specific applications
Common Issues	Data perspective: high-dimensional, multi-modal, labor-intensive annotations Value perspective: speed and accuracy, a prerequisite for following-up tasks	

Spectrum of medical image segmentation methods

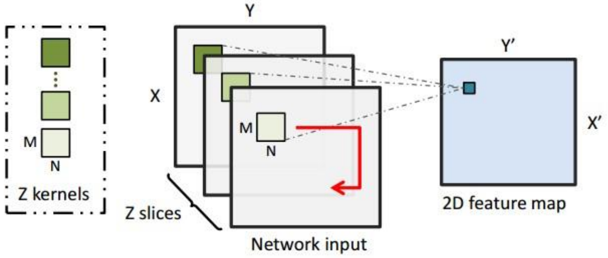
- Deformable models, such as snakes, active contours, level-set [T. McInerney and D. Terzopoulos D. MedIA 1996]
- Statistical inference via parametric or nonparametric probability models [D. Pham, C. Xu, and J. Prince Annu Rev Biomed Eng 2000]
- Multi-atlas based segmentation via registration and label fusion [J. Iglesias and M. Sabuncu MedIA 2015]
- Discriminative classifier based methods

$F_\theta(X) \to Z \in R^z;\ \ C_\phi(Z) \to Y \in R^c$ $\qquad X \to F_\theta(\cdot) \circ C_\phi(\cdot) \to Y$ *Deep Learning*

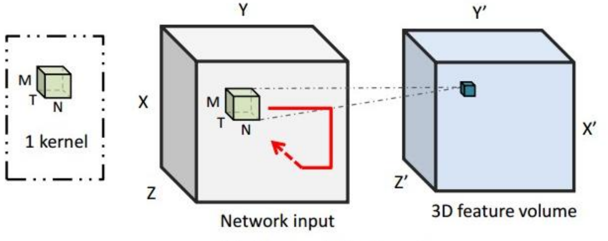
End-to-end 3D Networks for Semantic Segmentation

3D models

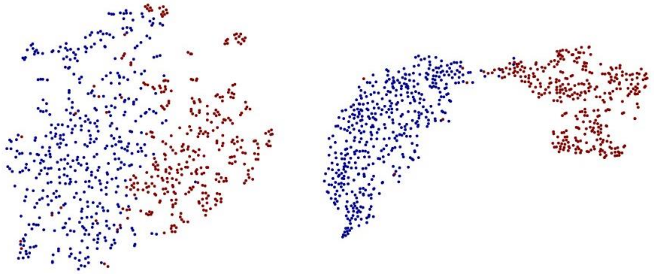
- Encode richer spatial information
- Fit to domain knowledge



(a) 2D convolution kernel



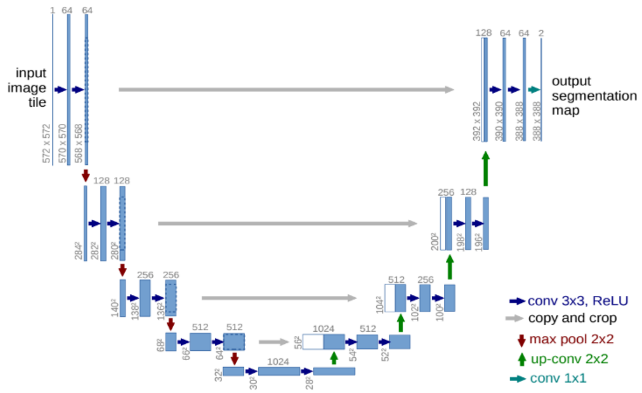
(b) 3D convolution kernel



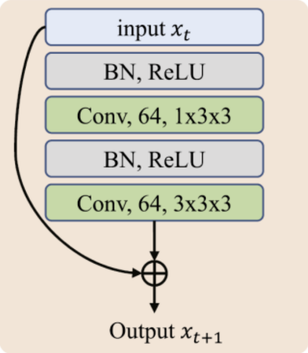
[Dou et al. TMI 2016]

Skip connection

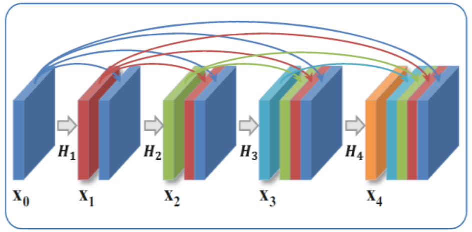
- Short-cut connection facilitate information flow
- Mutli-scale feature fusion for pixel-wise prediction



[Ronneberger et al. MICCAI 2015]



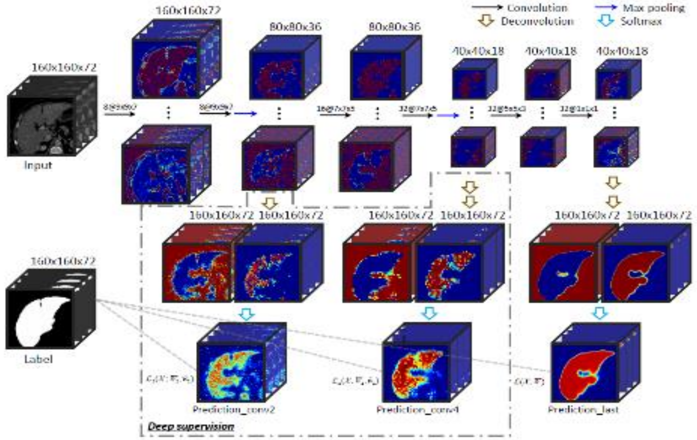
[Chen et al. NeuroImage 2018]



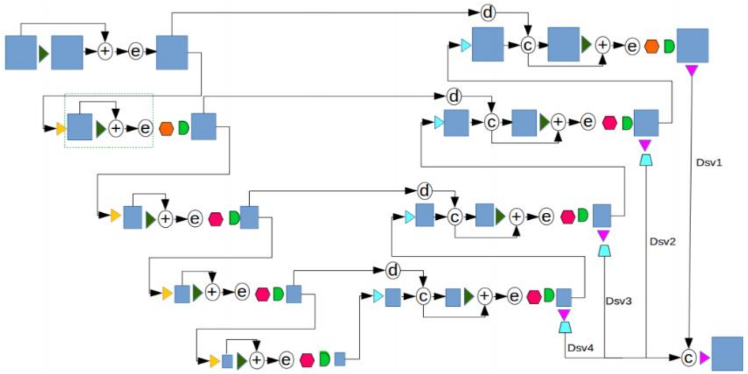
[Yu et al. MICCAI 2017]

Deep supervision

- Enhance gradients with explicit loss propagation
- Speed up training process



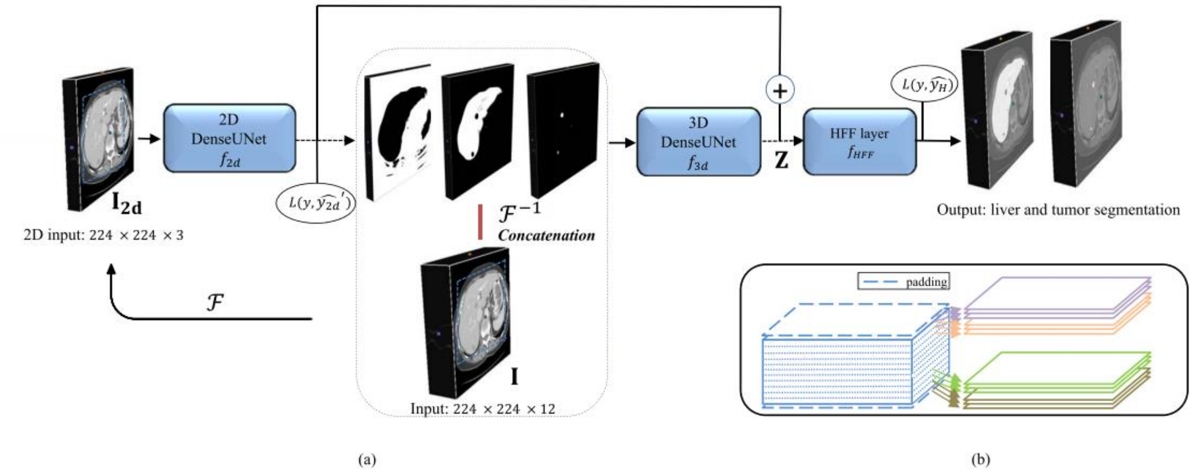
[Dou et al. MICCAI 2016; MedIA 2017]

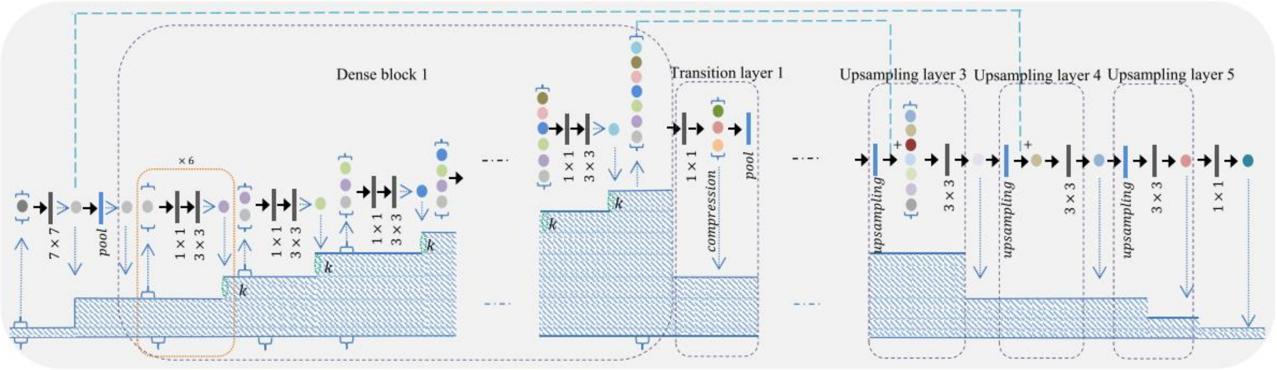


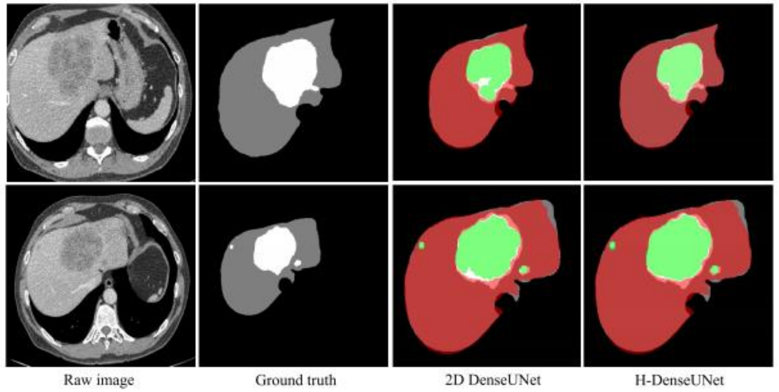
[Liu et al. IEEE Access 2020]

Improve Computational Efficiency by Network Design

H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation





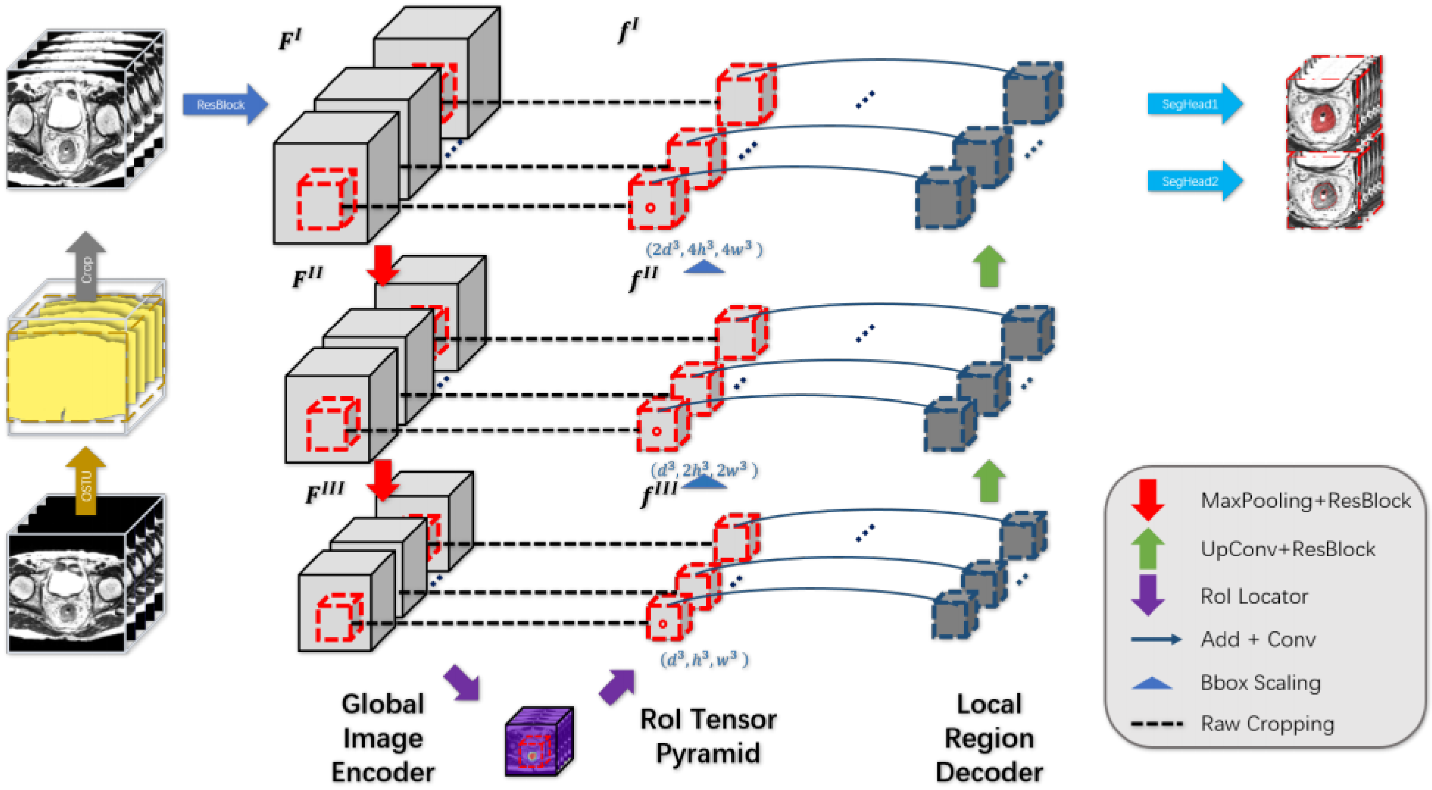


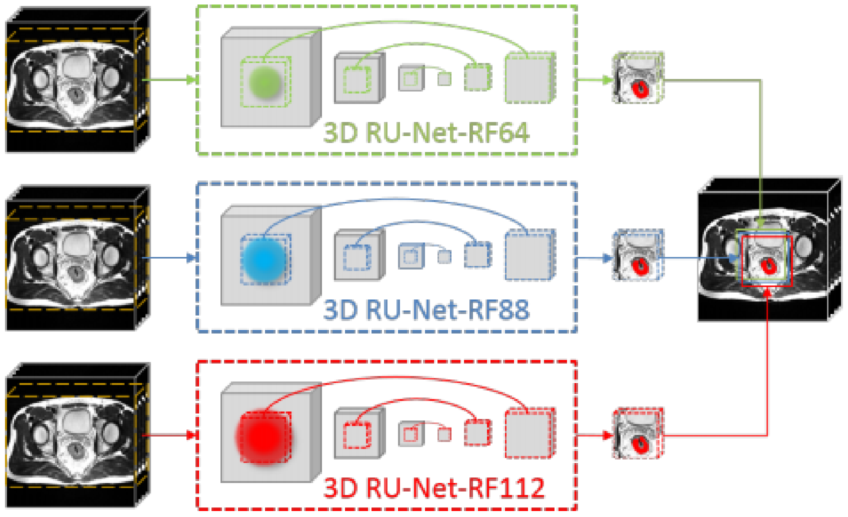
Team	Lesion		Liver		Tumor Burden
	Dice per case	Dice global	Dice per case	Dice global	RMSE
our	**72.2**	**82.4**	**96.1**	**96.5**	**0.015**
IeHealth	70.2	79.4	**96.1**	96.4	0.017
hans.meine	67.6	79.6	96.0	**96.5**	0.020
superAI	67.4	81.4	0.0	0.0	1251.447
Elehanx [40]	67.0	-	-	-	-
medical	66.1	78.3	95.1	95.1	0.023
deepX [48]	65.7	82.0	**96.3**	**96.7**	0.017
Njust768	65.5	76.8	4.10	13.5	0.920
Medical [41]	65.0	-	-	-	-
Gchlebus [42]	65.0	-	-	-	-
predible	64.0	77.0	95.0	95.0	0.020
Lei [49]	64.0	-	-	-	-
ed10b047	63.0	77.0	94.0	94.0	0.020
chunliang	62.5	78.8	95.8	96.2	0.016
yaya	62.4	79.2	95.9	96.3	0.016

Li et al. "H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes." IEEE Trans. on Medical Imaging, 2018.

Improve Computational Efficiency by Network Design

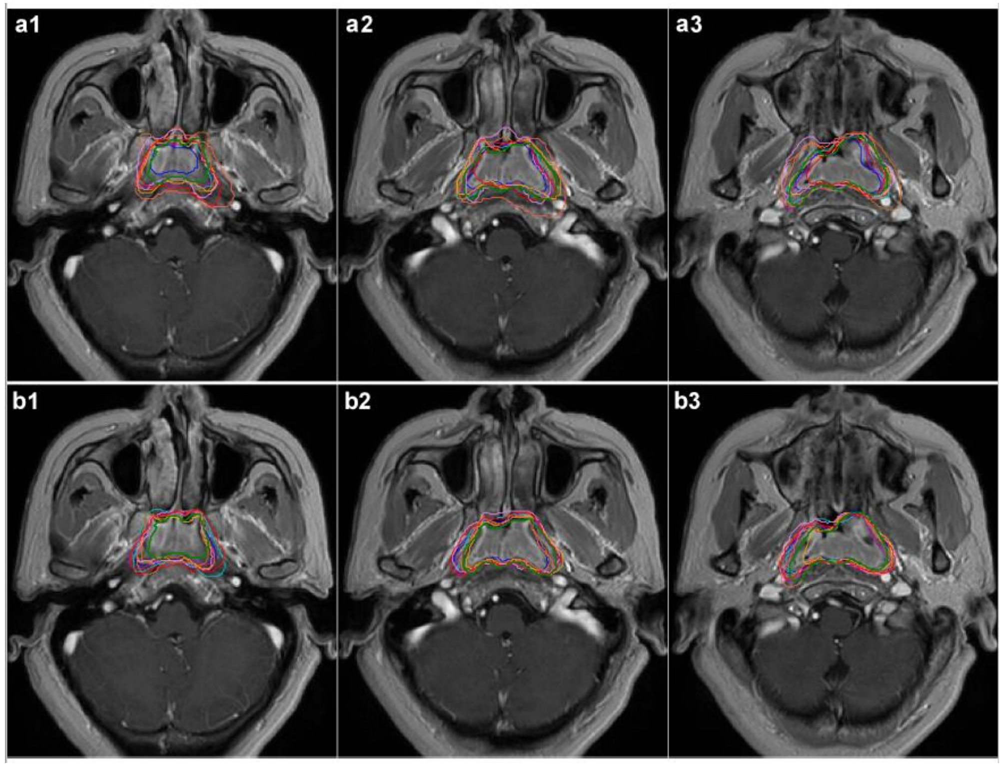
3D RoI-aware U-Net for Accurate and Efficient Colorectal Tumor Segmentation



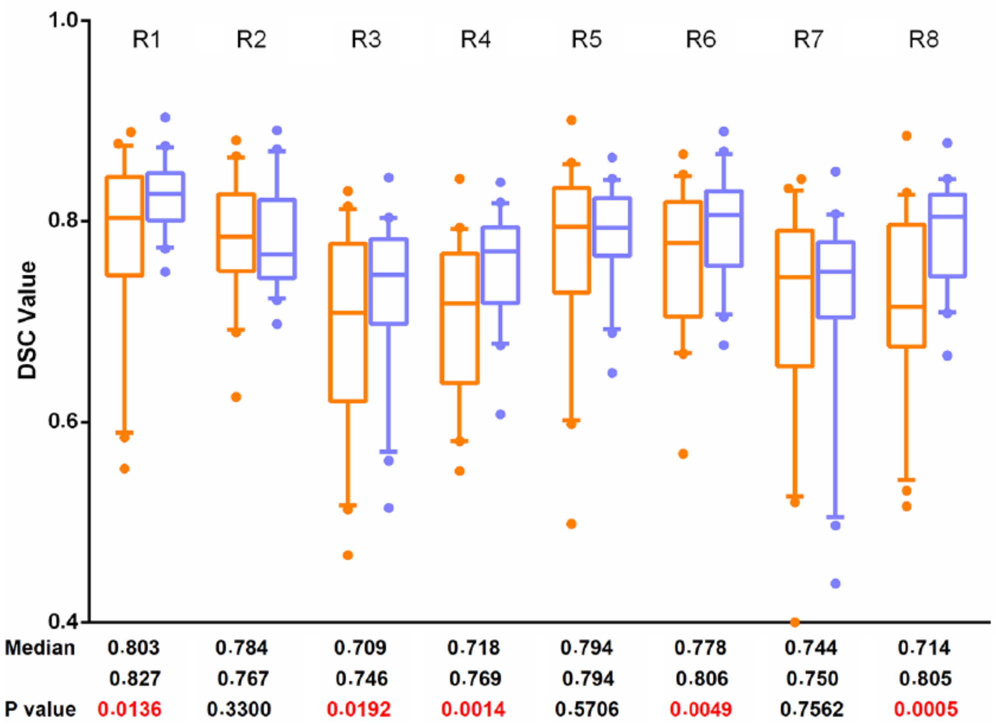


Huang et al. "3D RoI-aware U-Net for Accurate and Efficient Colorectal Tumor Segmentation." IEEE Trans. on Cybernetics 2020.

How can AI Segmentations Influence Clinicians?



Editing AI-based Contour	Traditional Manual Draw
18 mins	30 mins

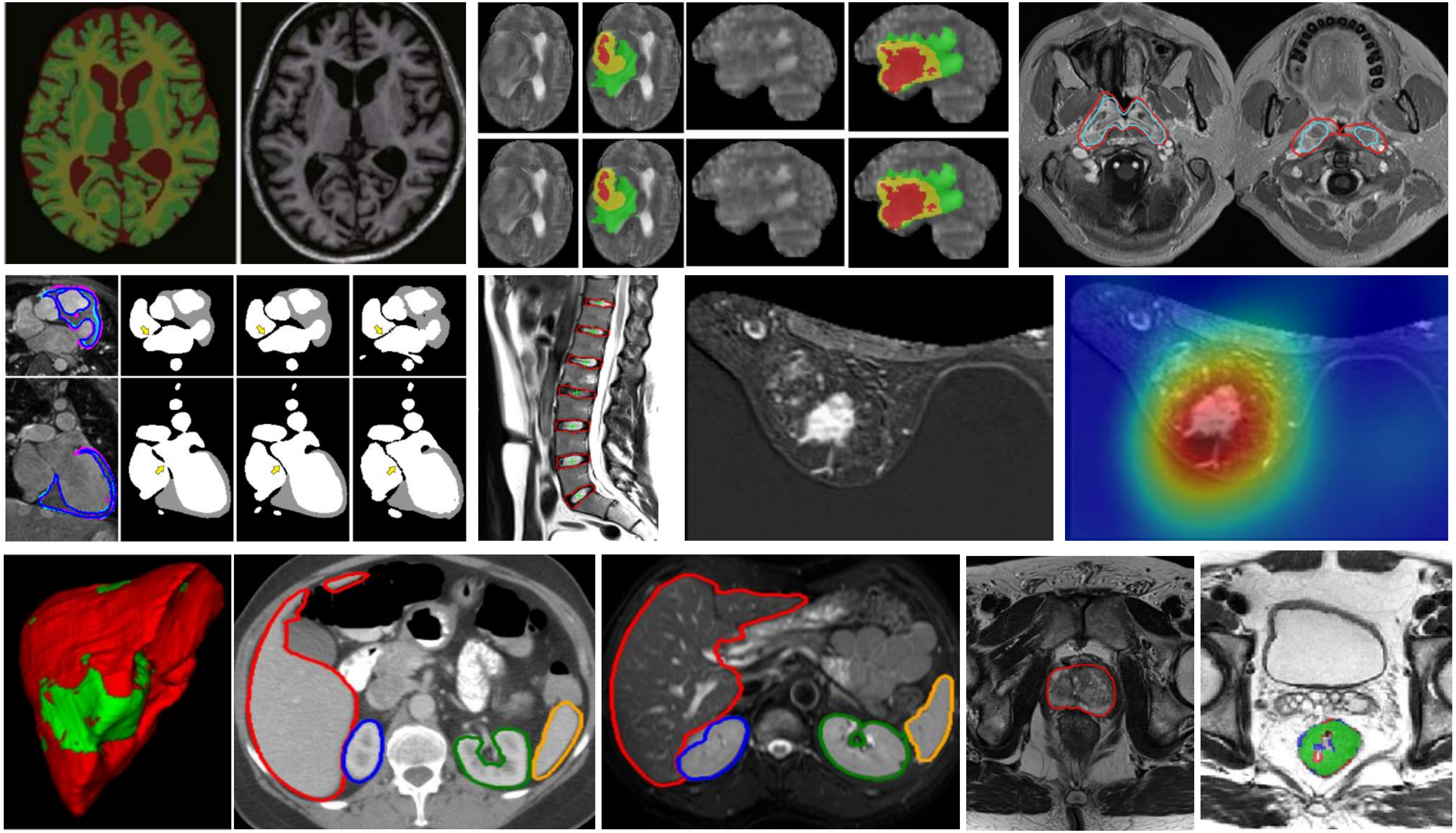


- saving 40% time, per case
- process 1 more cancer patient in 1 hour

Do more calculations …

L. Lin, Q. Dou* et al. “Deep Learning for Automated Contouring of Primary Tumor in Radiotherapy of Nasopharynx Cancer on Multiparametric Magnetic Resonance Imaging”, Radiology, 2019.*

A Gallery of Image Segmentation Scenarios



Open Problems

***Weak label usage**, e.g., semi-supervised learning; noisy-label training; unsupervised learning, active learning.*

***Data scarcity issue**, e.g., transfer learning; data augmentation; limited data training.*

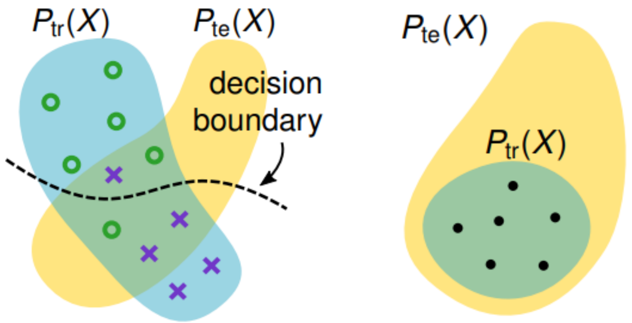
***Interpretability issue**, e.g., uncertainty estimation; explainable deep learning; relationship and causality.*

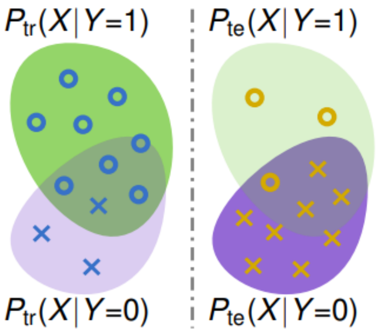
***Clinical oriented**, e.g., large-scale validation; beyond imaging data;*

Model generalization.

Model Generalization in Real World Conditions

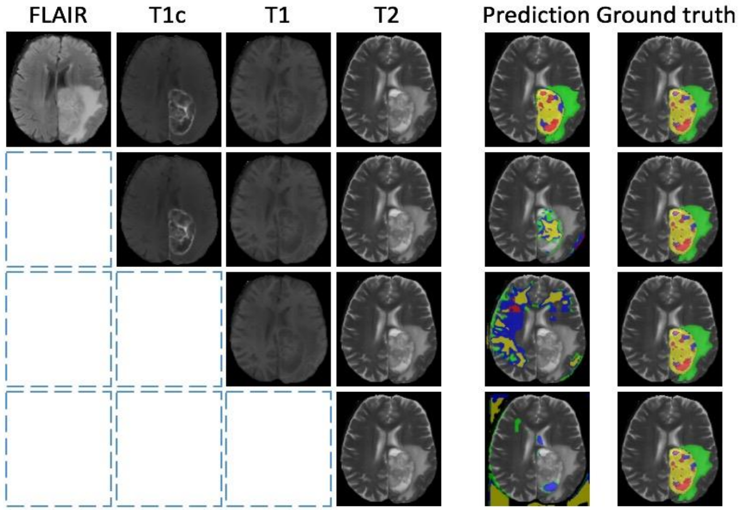
Data-driven method is sensitive to data mismatch



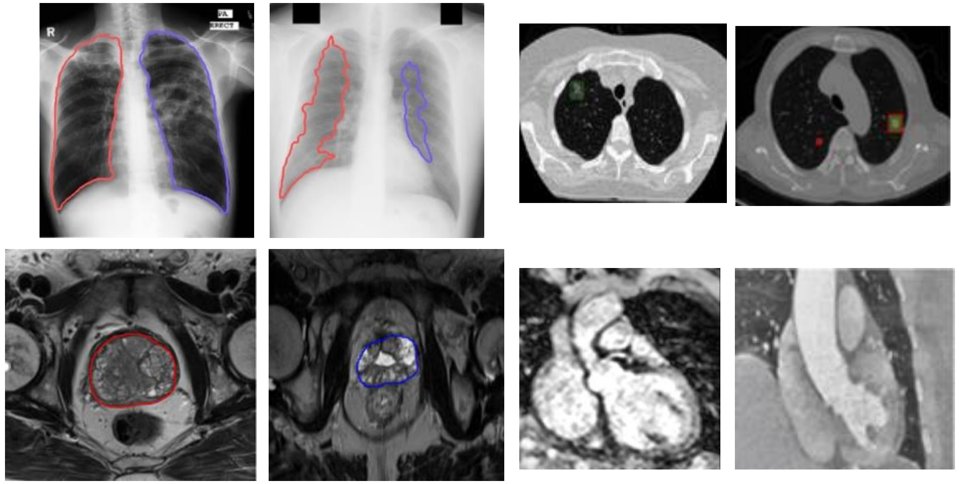


[D. Castro, I. Walker, and B. Glocker. 2019]

- **Low data quality at inference**
  - artefacts,
  - missing modality,
  - unseen severe cases, etc.



- **Data heterogeneity**
  - different vendors,
  - imaging protocols,
  - patient population



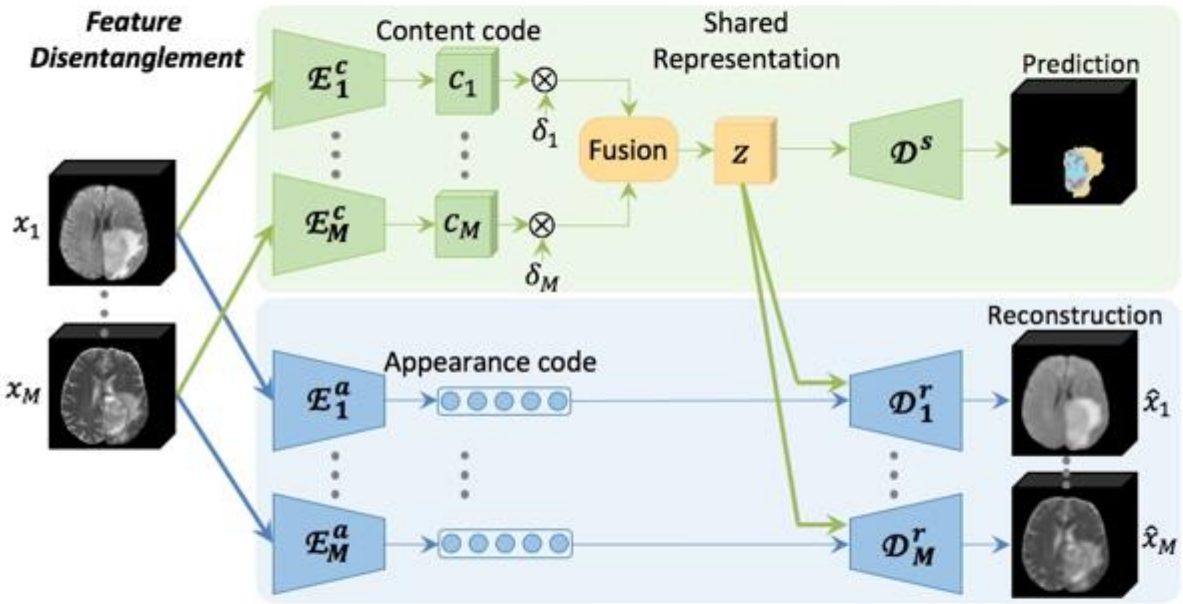
Tackling Missing Modality via Feature Disentanglement

Loss for disentangled appearance code:

$$\mathcal{L}_{\text{KL}} = \sum_{i=1}^{M} \mathbb{E}[D_{\text{KL}}(p(a_i)||\mathcal{N}(\mathbf{0}, \boldsymbol{I}))]$$

Loss for reconstruction with content code:

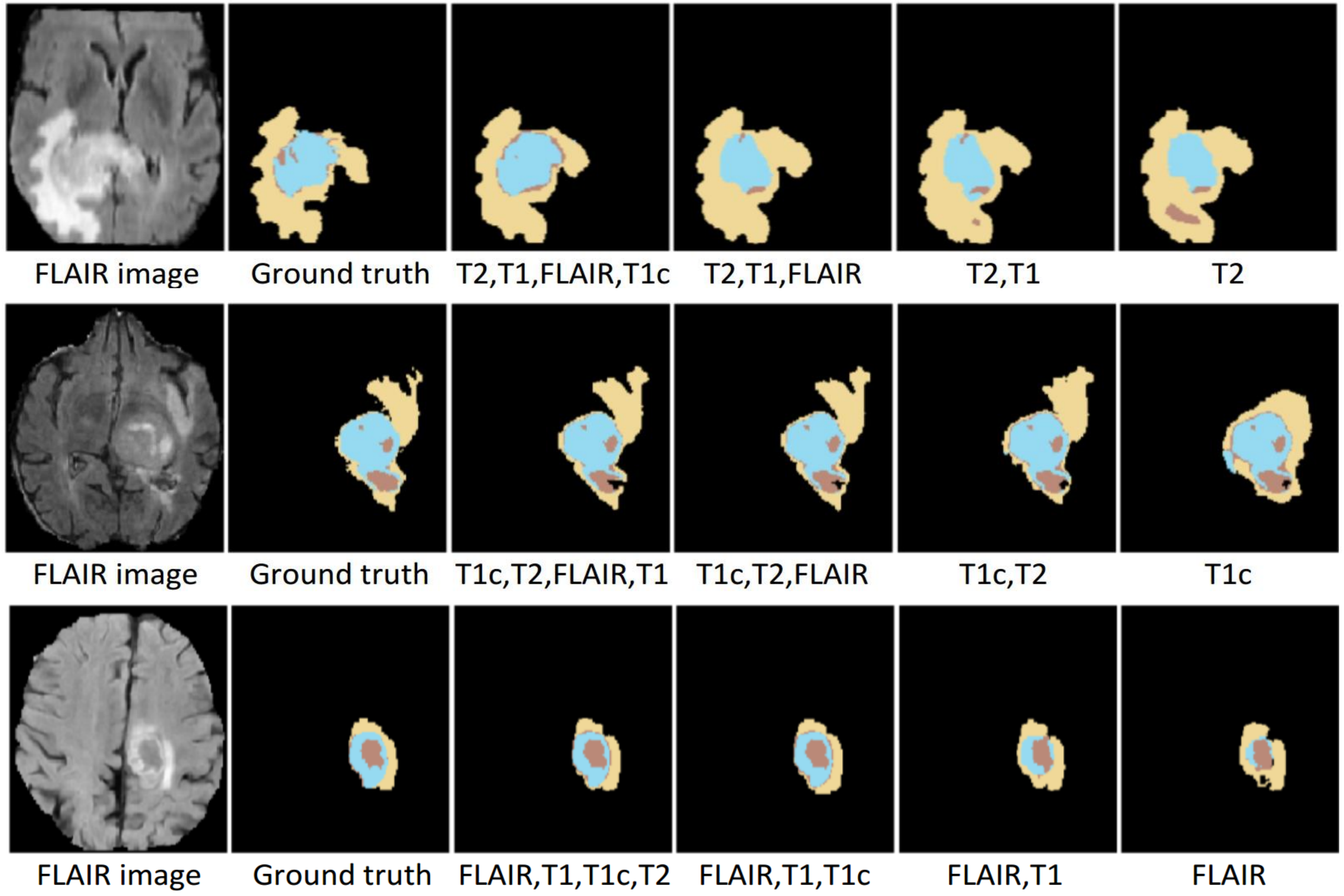
$$\mathcal{L}_{\text{rec}} = \sum_{i=1}^{M} ||D_i^r(z, a_i) - x_i||_1, \text{where } z = \mathcal{F}(\delta_i c_1, ..., \delta_M c_M)$$

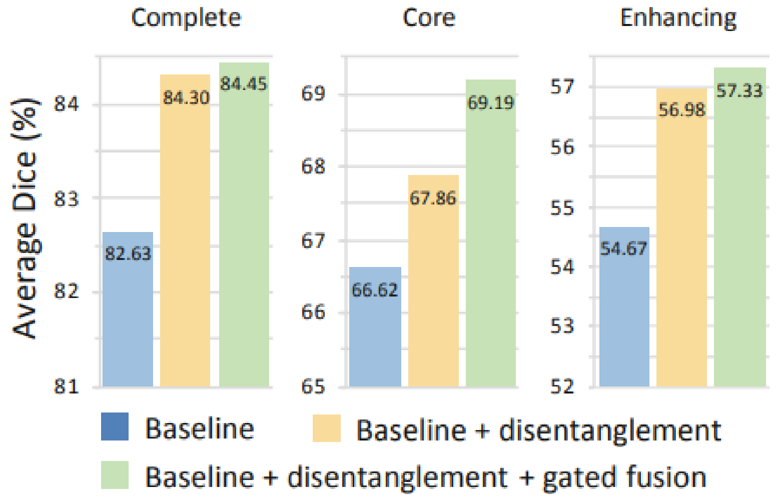


Modalities				Dice (%)								
				Complete			Core			Enhancing		
F	T1	T1c	T2	Ours	HeMIS	MLP	Ours	HeMIS	MLP	Ours	HeMIS	MLP
–	–	–	✓	**85.49**	58.48	61.50	**58.66**	40.18	37.32	**37.66**	20.31	18.62
–	–	✓	–	**71.86**	33.46	2.04	**72.87**	44.55	17.70	**70.22**	49.93	32.92
–	✓	–	–	**68.40**	33.22	2.07	**50.00**	17.42	10.52	**22.67**	4.67	10.78
✓	–	–	–	**83.02**	71.26	63.81	**46.67**	37.45	34.26	**28.30**	5.57	15.90
–	–	✓	✓	**87.53**	67.59	64.97	**78.46**	63.39	49.38	**76.82**	65.38	60.30
–	✓	✓	–	**74.59**	45.93	1.99	**76.40**	55.06	26.55	**73.95**	62.40	40.93
✓	✓	–	–	**87.66**	80.28	78.13	**60.17**	49.52	48.97	**35.28**	22.26	25.18
–	✓	–	✓	**87.87**	69.56	66.88	**64.88**	47.26	43.66	**41.05**	23.56	26.37
✓	–	–	✓	**89.08**	82.10	81.35	**63.51**	53.42	52.41	**39.72**	23.19	25.01
✓	–	✓	–	**88.01**	79.80	81.13	**78.09**	66.12	65.51	**76.62**	67.12	66.19
✓	✓	✓	–	**87.73**	80.88	82.19	**80.68**	69.26	69.34	**78.81**	71.30	70.93
✓	✓	–	✓	**89.07**	83.87	80.40	**65.99**	57.76	53.46	**43.04**	28.46	28.34
✓	–	✓	✓	**89.06**	82.78	83.37	**79.47**	70.62	70.45	**78.07**	70.52	70.56
–	✓	✓	✓	**88.26**	70.98	67.85	**80.84**	66.60	55.40	**78.56**	67.84	64.81
✓	✓	✓	✓	**89.07**	83.15	82.43	**81.19**	72.50	71.46	**79.13**	75.37	72.08
Average				**84.45**	68.22	60.01	**69.19**	54.07	47.09	**57.33**	43.86	41.93

C. Chen, Q. Dou, et al. "Robust Multimodal Brain Tumor Segmentation via Feature Disentanglement and Gated Fusion", MICCAI, 2019.

Tackling Missing Modality via Feature Disentanglement

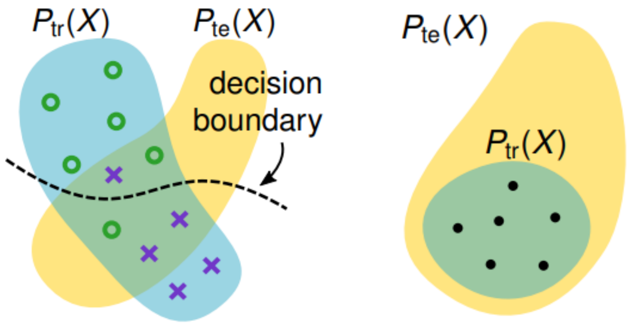


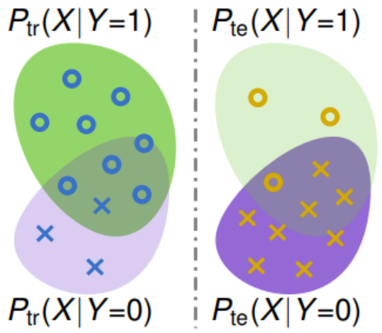


C. Chen, Q. Dou, et al. "Robust Multimodal Brain Tumor Segmentation via Feature Disentanglement and Gated Fusion", MICCAI, 2019.

Model Generalization in Real World Conditions

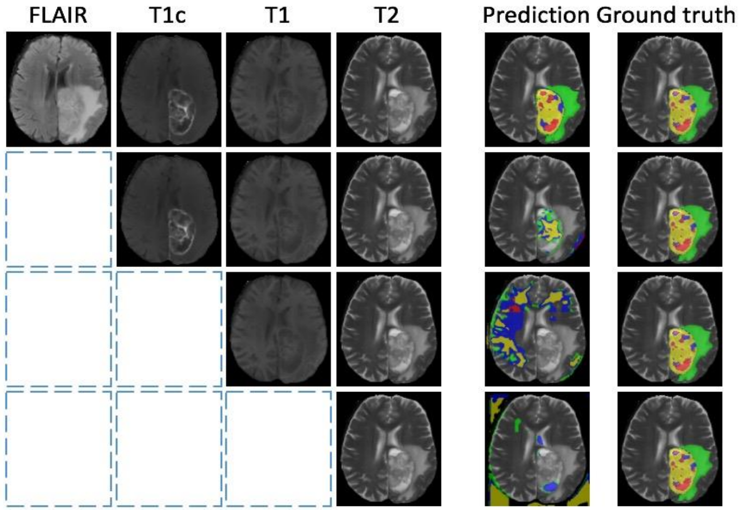
Data-driven method is sensitive to data mismatch



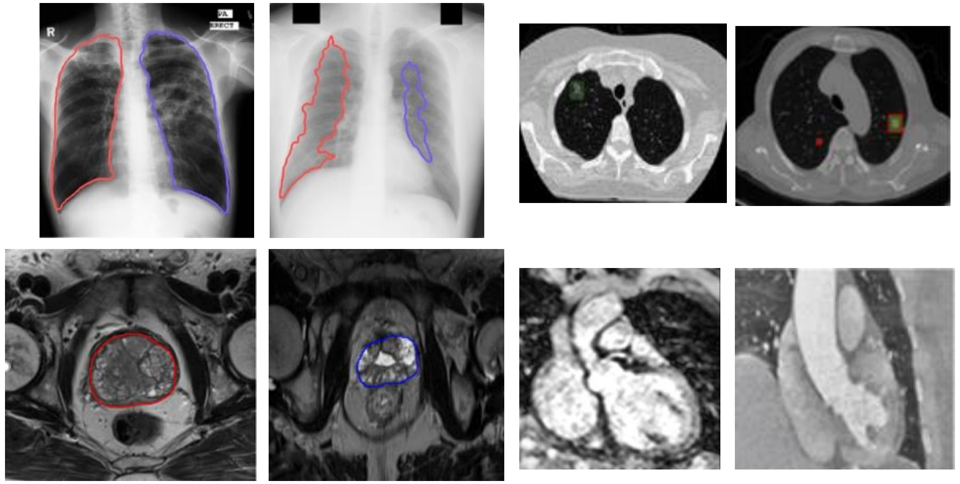


[D. Castro, I. Walker, and B. Glocker. 2019]

- **Low data quality at inference**
  - artefacts,
  - missing modality,
  - unseen severe cases, etc.



- **Data heterogeneity**
  - different vendors,
  - imaging protocols,
  - patient population



Tackling Data Heterogeneity: does Normalization Help?

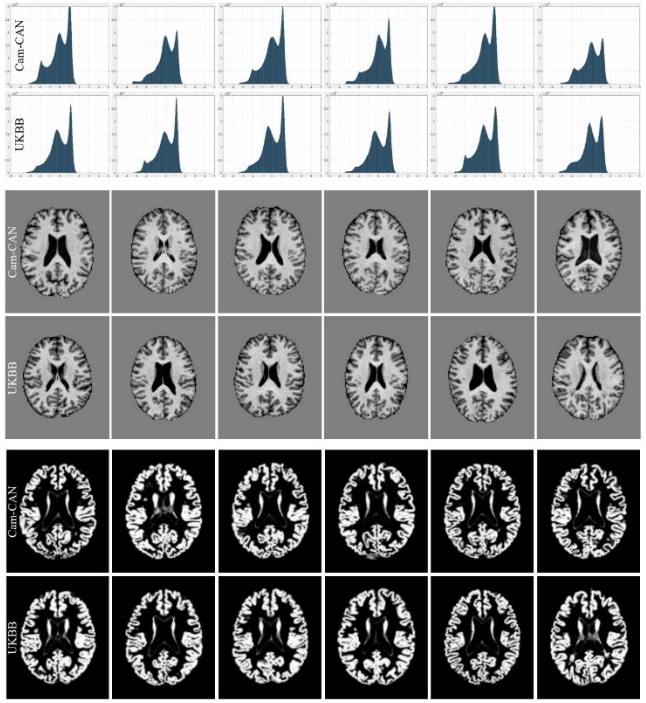
An empirical study on the impact of scanner effects with brain imaging

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



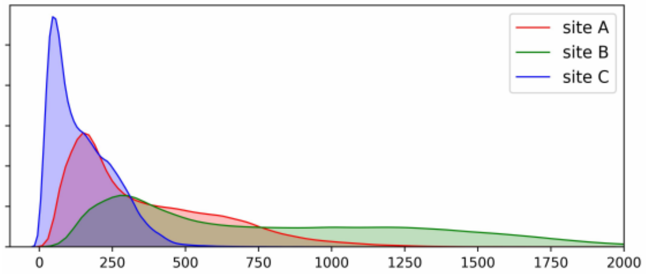
B. Glocker et al. “Machine Learning with Multi-site Imaging Data: An Empirical Study on the Impact of Scanner Effects.” Medical Imaging meets NeurIPS Workshop, 2019.

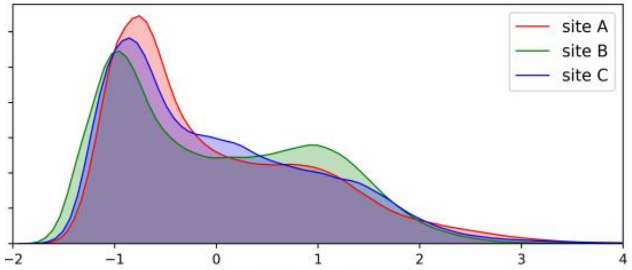
Related work: [Shafto et al., 2014; Taylor et al., 2017; Sudlow et al., 2015; Miller et al., 2016; Alfaro-Almagro et al., 2018]

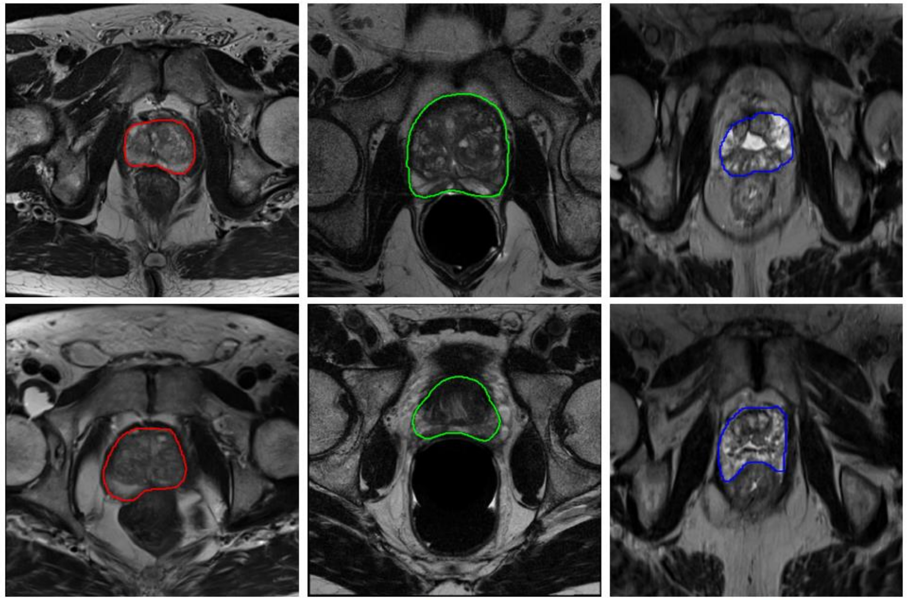
Tackling Data Heterogeneity with Supervised Learning

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







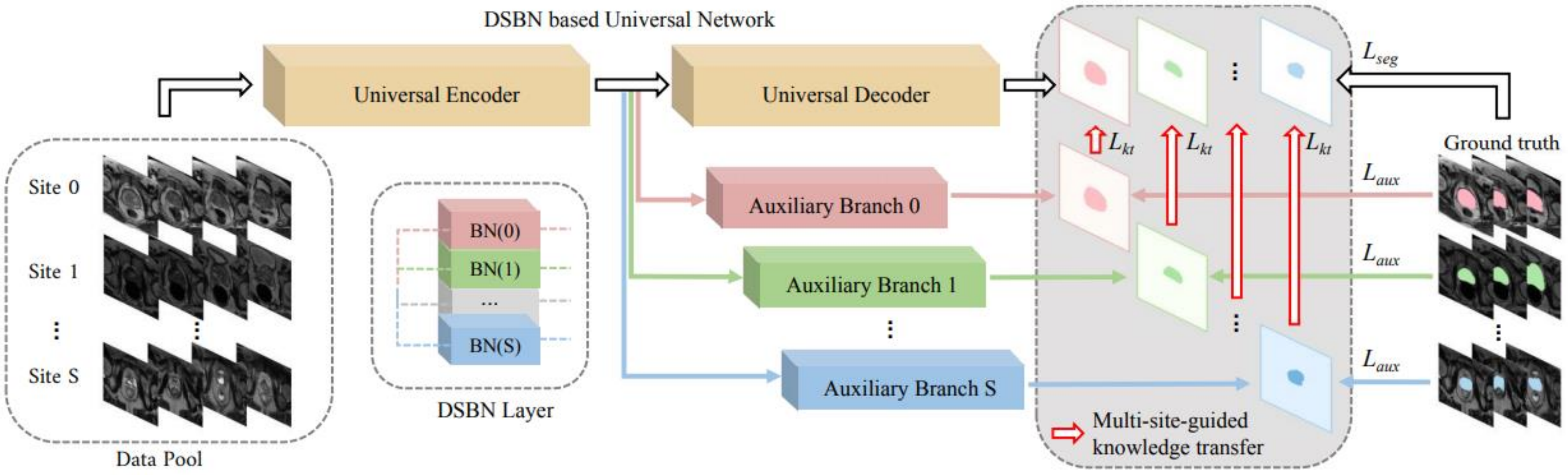
Site A Site B Site C

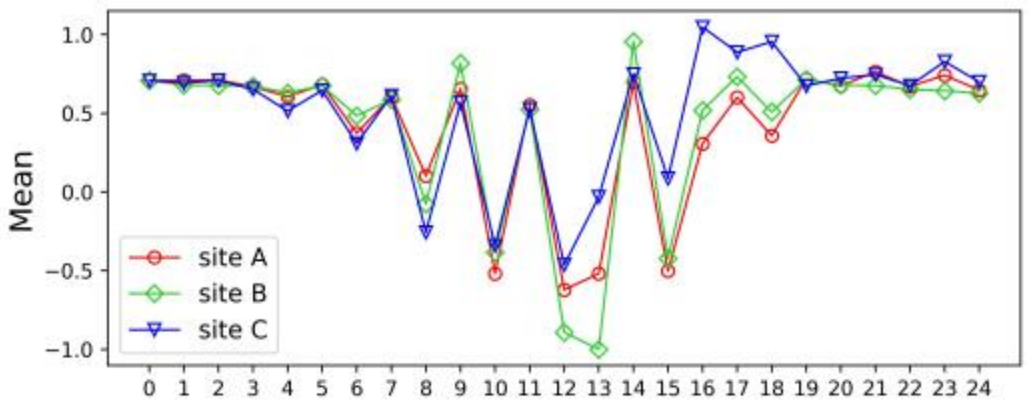
Methods	BFC	NF	Intensities	Site A	Site B	Site C	Overall
Separate (A)	✗	✗	whitening	90.47	76.44	56.81	
Separate (B)	✗	✗	whitening	70.11	90.52	50.25	90.56
Separate (C)	✗	✗	whitening	57.93	55.25	90.70	
Joint	✗	✗	✗	86.51	88.00	86.78	87.10
Joint	✗	✗	histogram	87.68	88.02	89.46	88.39
Joint	✗	✗	scaled	90.43	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



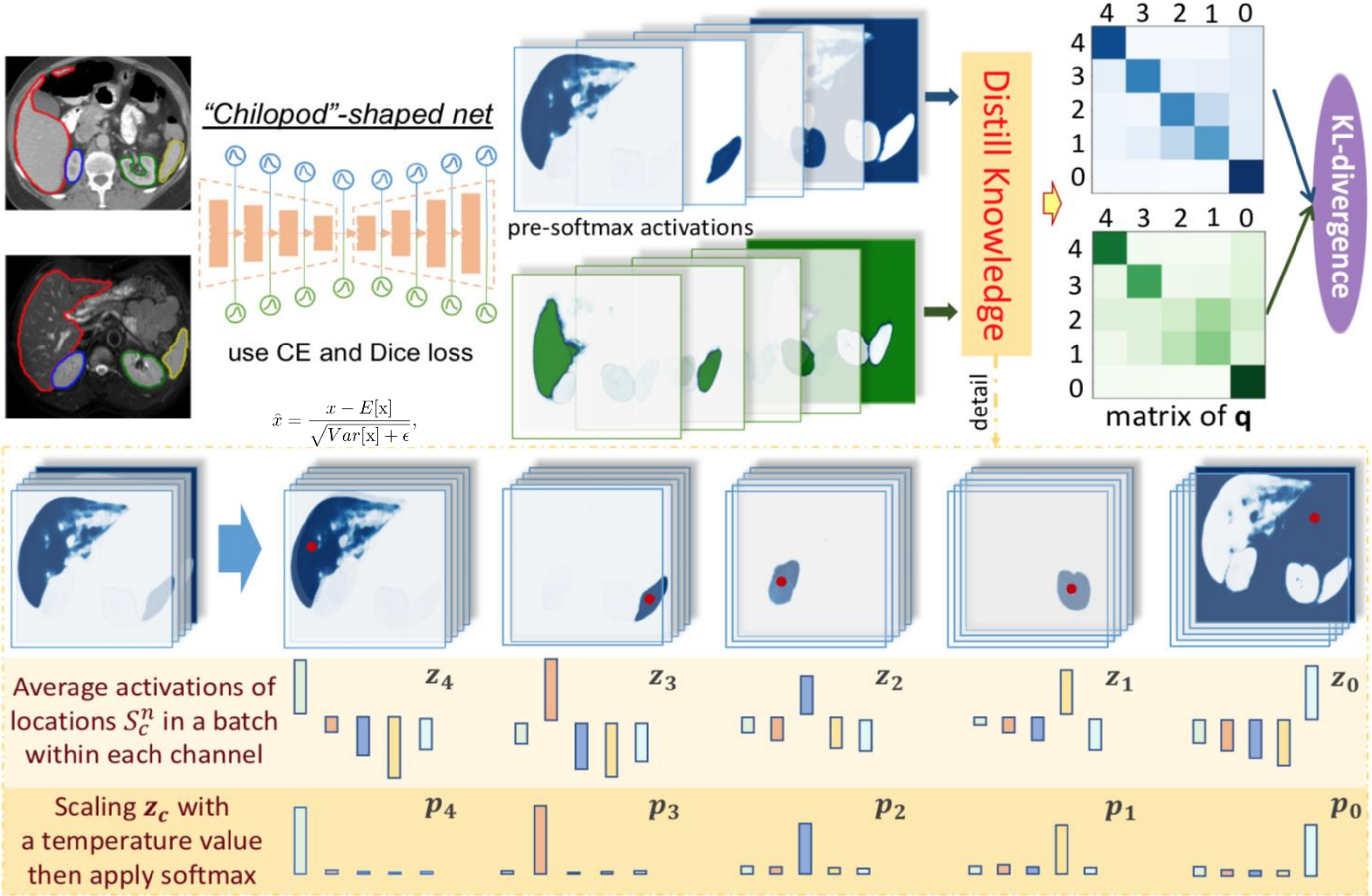


	Dice Coefficient (mean±std, %)			
Methods	Site A	Site B	Site C	Overall
Tian *et al.* [26]	88.23	88.23	—	
Rundo *et al.* [9]	—	—	88.66	
Separate	90.47±3.00	90.52±2.45	90.70±3.34	90.56±2.88
Joint	90.69±3.05	89.53±2.97	90.55±3.18	90.25±3.08
USE-Net [19]	90.90±2.41	90.17±2.61	90.73±2.36	90.60±2.50
Dual-Stream [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:

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

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

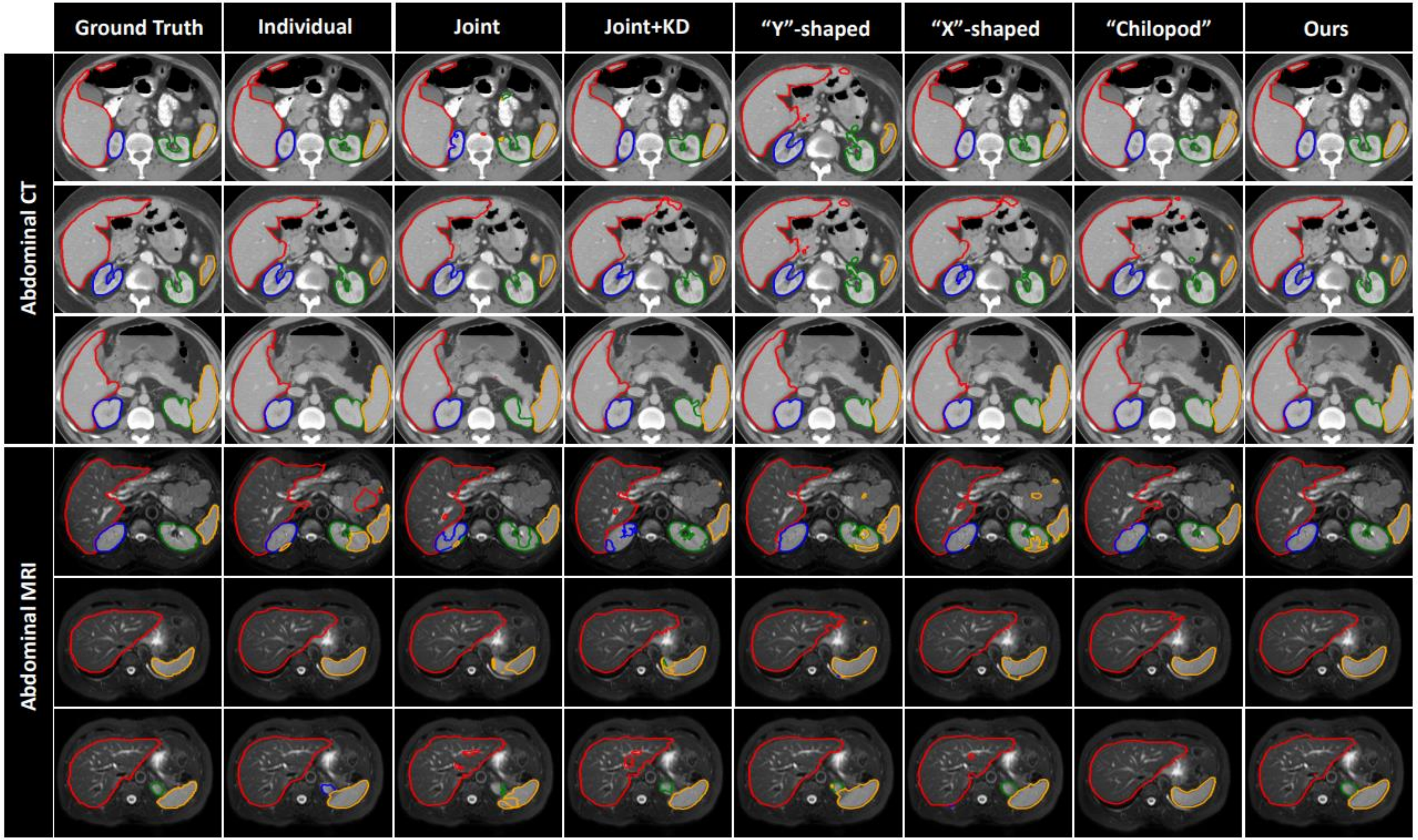
Minimize probability divergence:

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

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

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

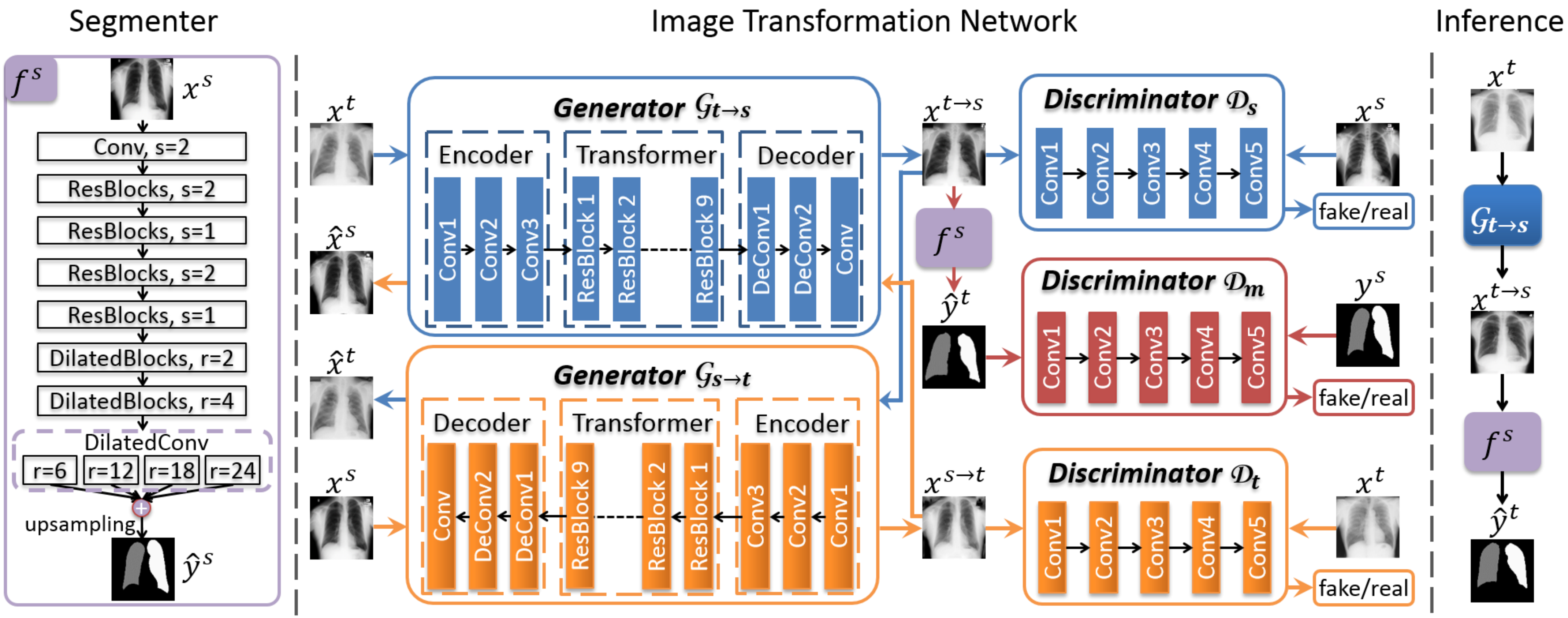
Tackling Data Heterogeneity with Supervised Learning



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

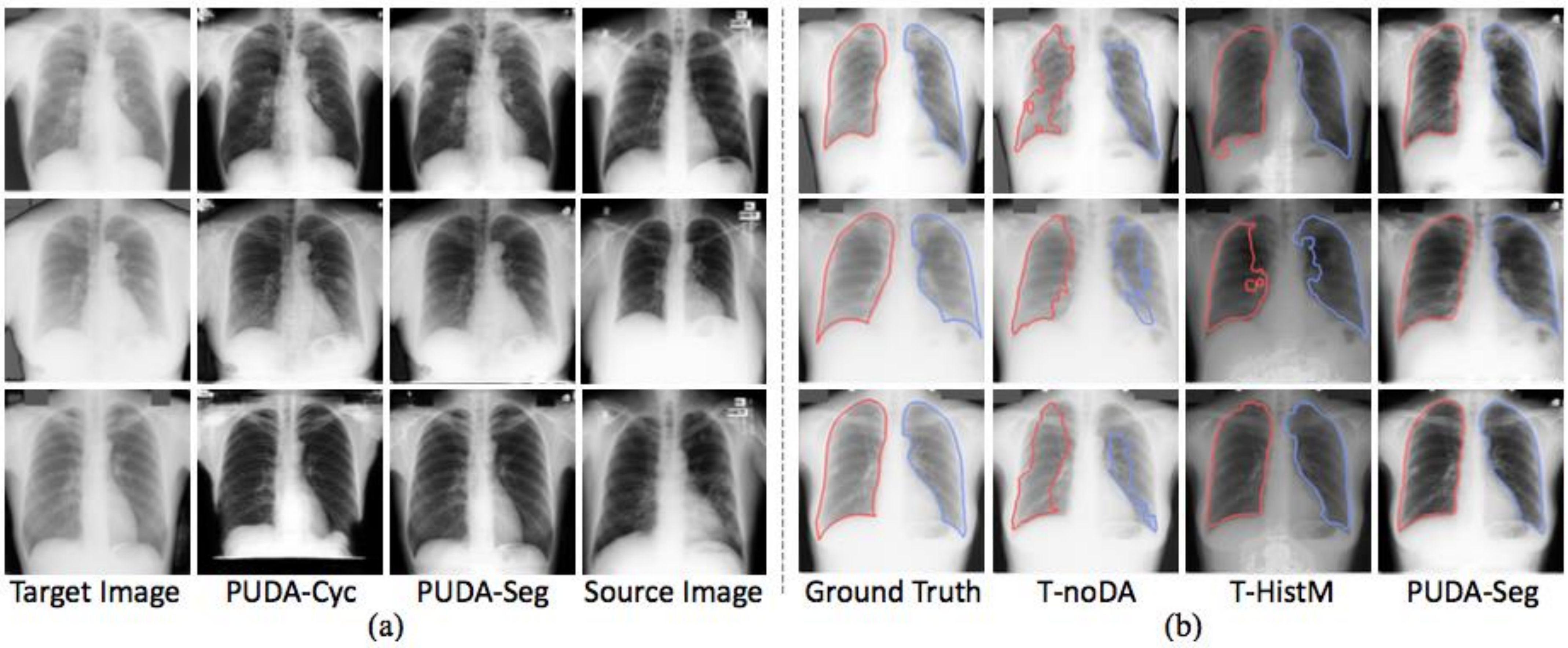


C. Chen, Q. Dou, et al. "Semantic-Aware Generative Adversarial Nets for Unsupervised Domain Adaptation in Chest X-ray Segmentation." MICCAI-MLMI'18 (Oral)

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)

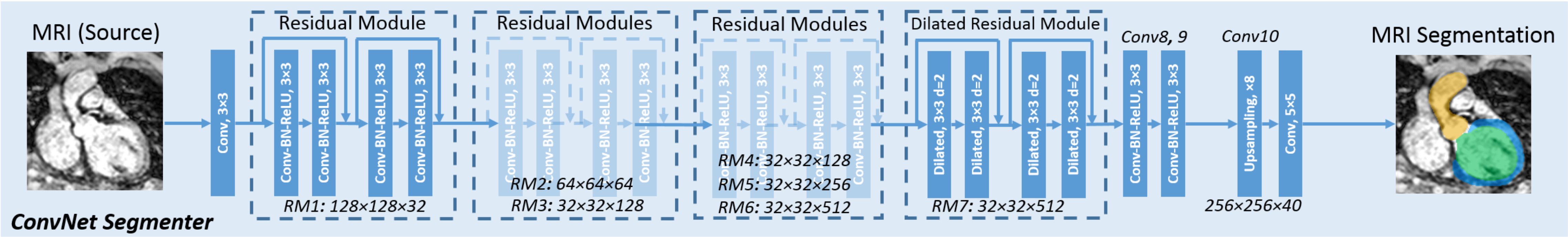
Related work: [Y. Huo et al., ISBI 2018; Z. Zhang et al. CVPR 2018; Y. Zhang et al. MICCAI 2018]

Unsupervised Domain Adaptation: Feature-level Alignment

Train a source domain segmentation model

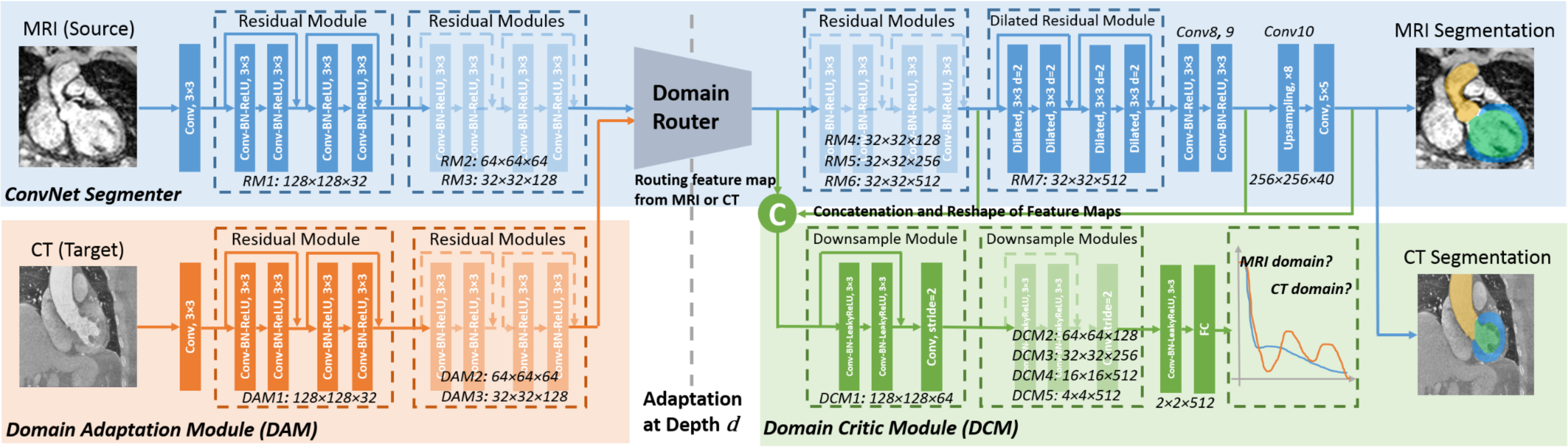
- joint cross-entropy loss and dice loss

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



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

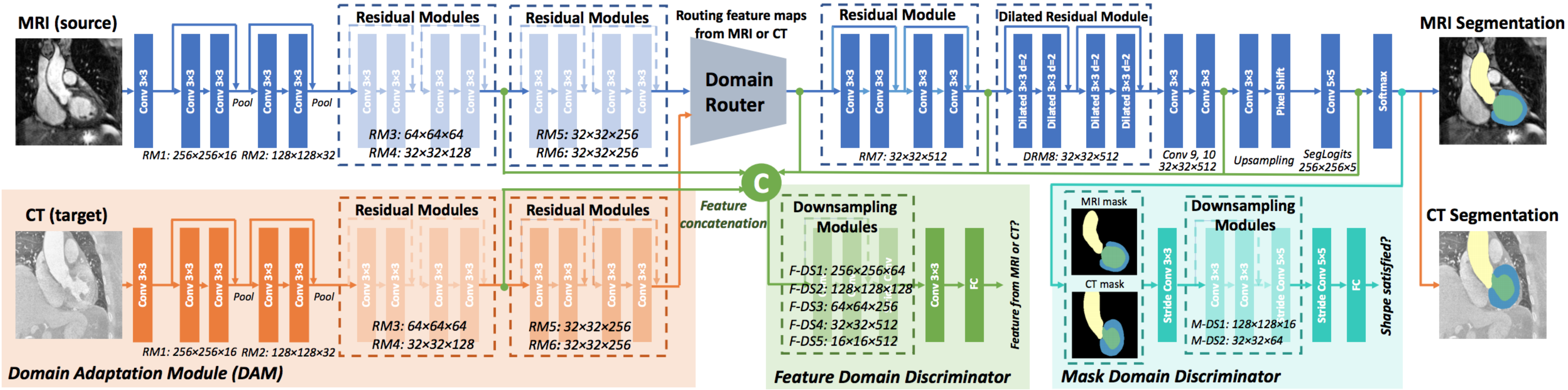
domain adaptation module (generator): $\min_{\mathcal{M}} \mathcal{L}_{\mathcal{M}}(X^t, \mathcal{D}) = -\mathbb{E}_{(\mathcal{M}_A(x^t), F_H(x^t)) \sim \mathbb{P}_g}[\mathcal{D}(\mathcal{M}_A(x^t), F_H(x^t))]$

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



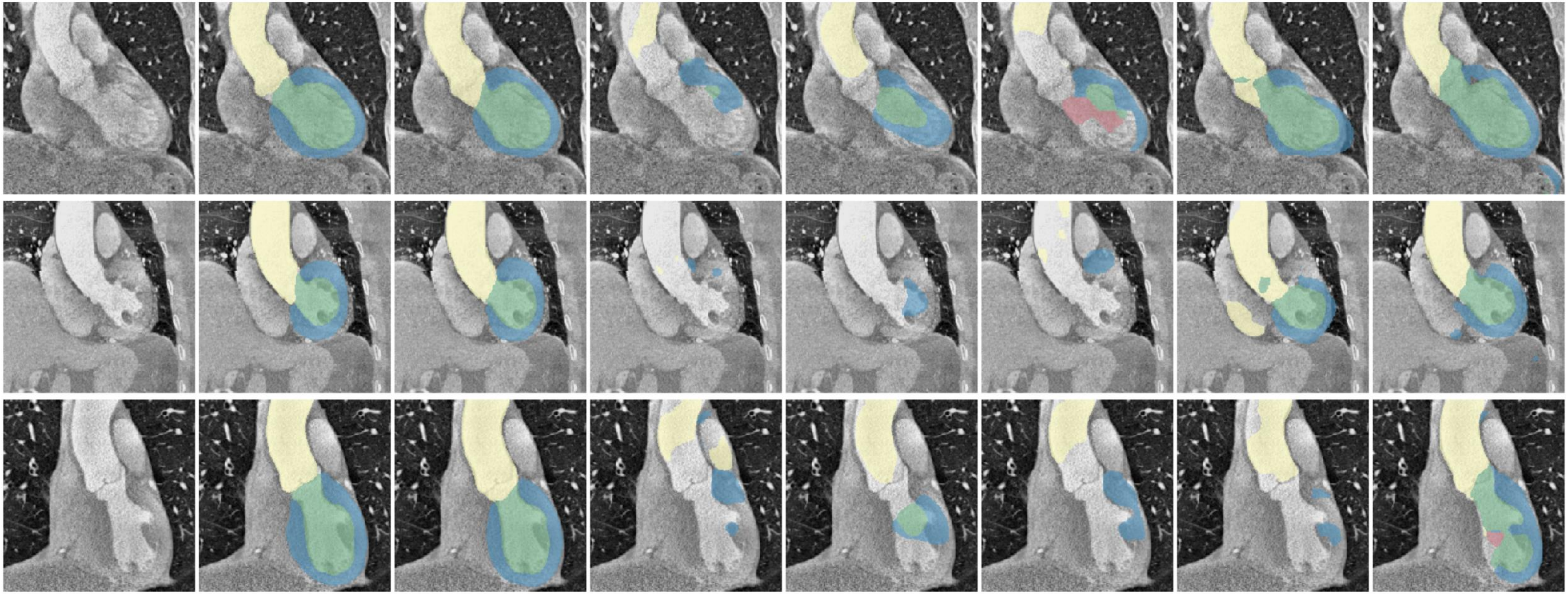
Unsupervised learning with adversarial loss

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

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

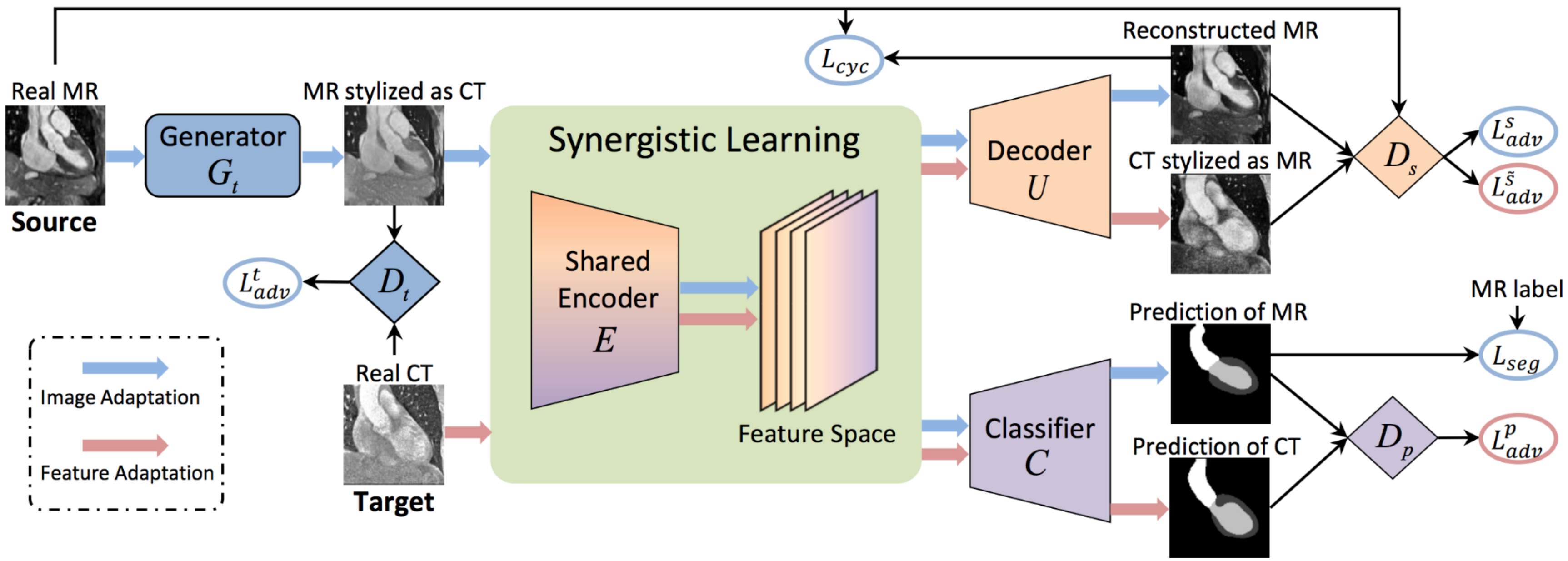
Unsupervised Domain Adaptation: Feature-level Alignment

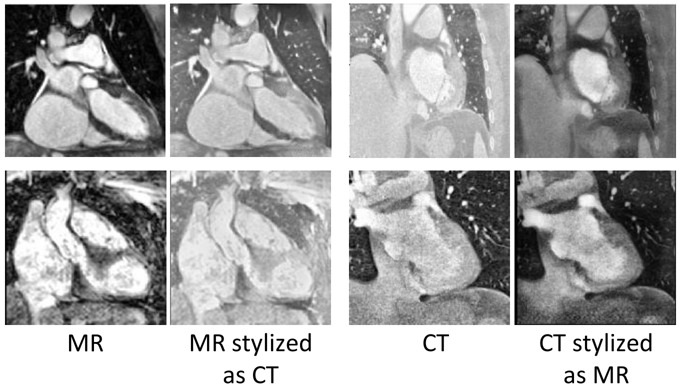


(a) CT image (b) CT label (c) Seg-CT (d) Seg-CT-noDA (e) DANN (f) ADDA (g) CycleGAN (h) Ours

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





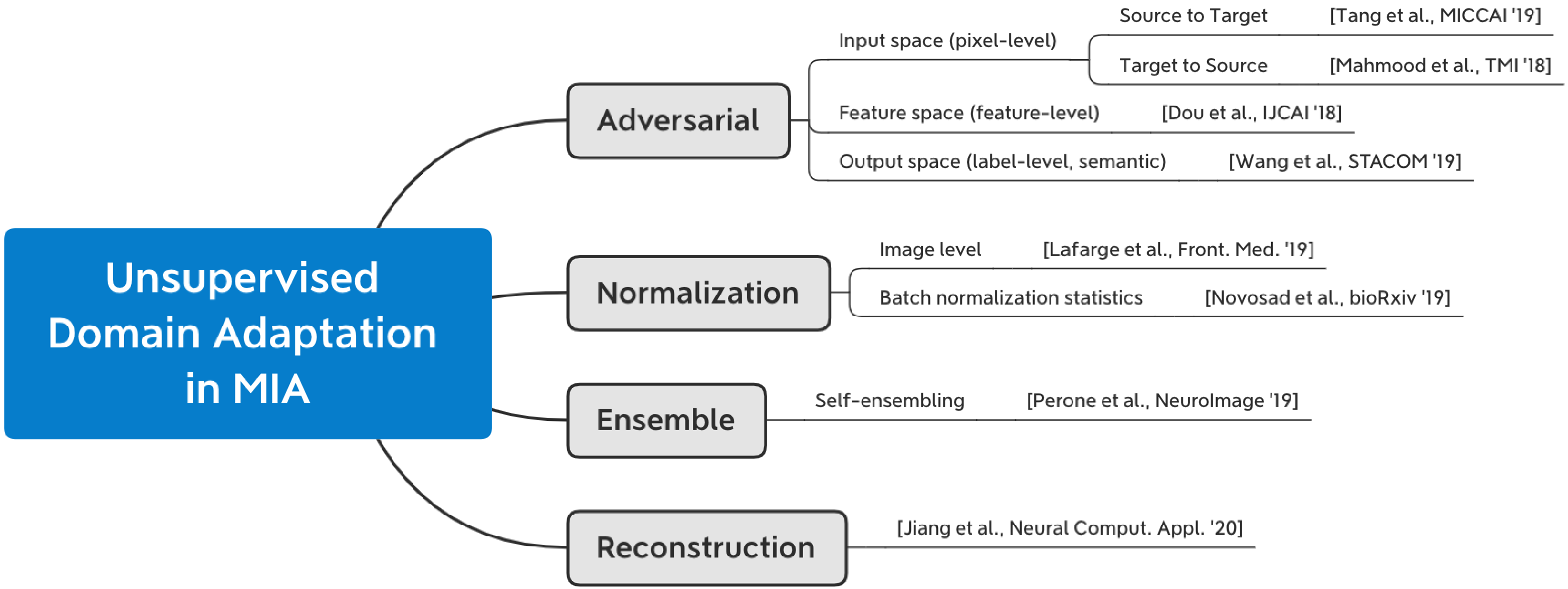
Methods	Adaptation		Dice					ASD				
	Image	Feature	AA	LAC	LVC	MYO	Average	AA	LAC	LVC	MYO	Average
W/o adaptation			28.4	27.7	4.0	8.7	17.2	20.6	16.2	N/A	48.4	N/A
DANN (Ganin et al. 2016)		✓	39.0	45.1	28.3	25.7	34.5	16.2	9.2	12.1	10.1	11.9
ADDA (Tzeng et al. 2017)		✓	47.6	60.9	11.2	29.2	37.2	13.8	10.2	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]

Tackling Data Heterogeneity with UDA

A brief overview of existing approaches

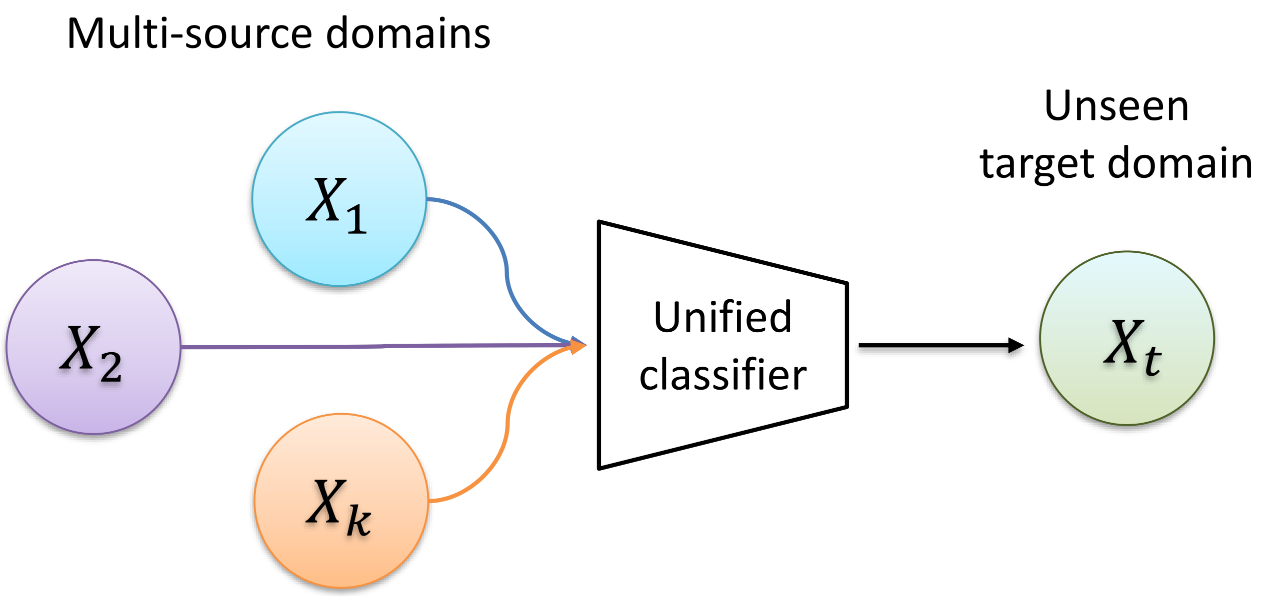


Domain Generalization

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

Regularization for generic semantic features

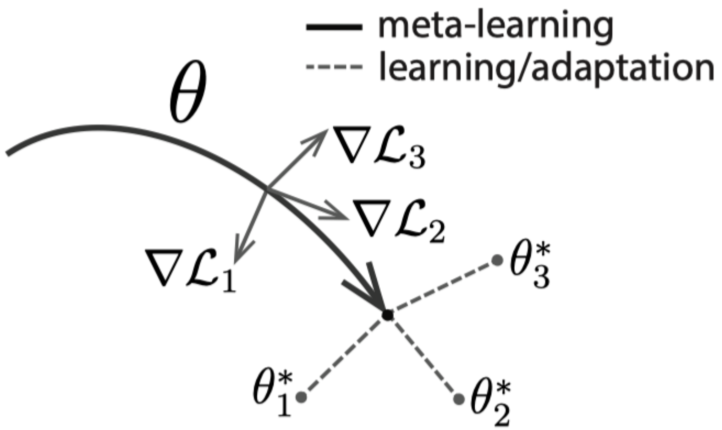
- adversarial feature alignment for domain invariance [Li et al. ECCV 2018]
- decompose networks parameters to domain-specific/invariant [Khosla ECCV 2012]
- data augmentation based methods [Shankar et al. ICLR 2018; Volpi et al. NeurIPS 2018]
- multi-task or self-supervised signals [Ghifary et al. ICCV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



Applying to domain generalization:

- 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]

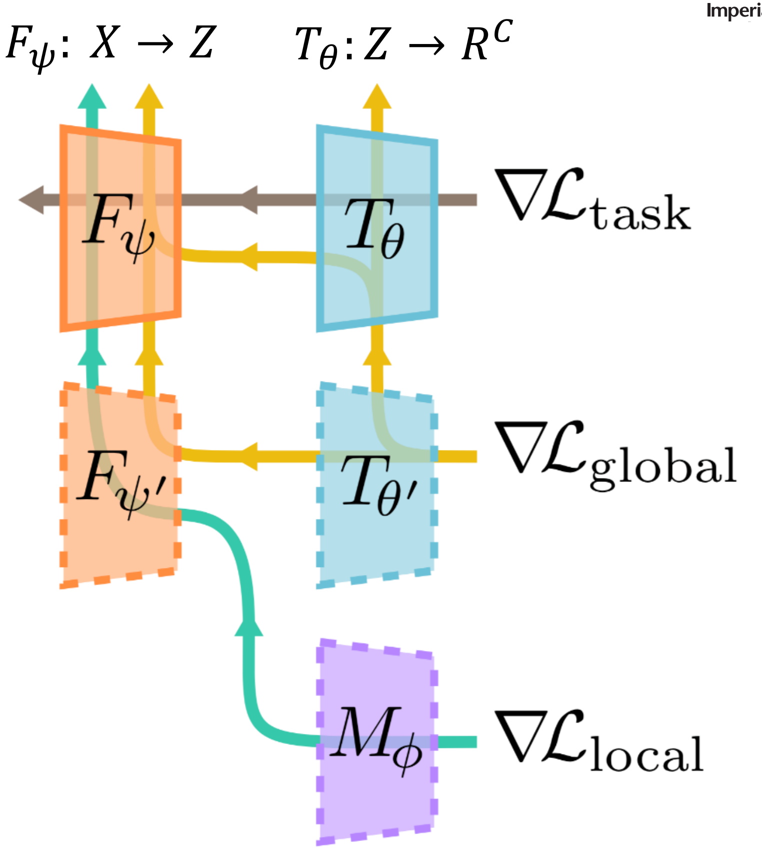
Q. Dou, D. Castro, et al. "Domain Generalization via Model-Agnostic Learning of Semantic Features", NeurIPS, 2019.

MASF: Model-Agnostic Learning of Semantic Features

Episodic training paradigm

Available domains: $D = \{D_{1,} D_{2,} \dots, D_{K,}\}$

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.

Global Class Alignment

Inter-class relationships concept is domain-invariant and transferable

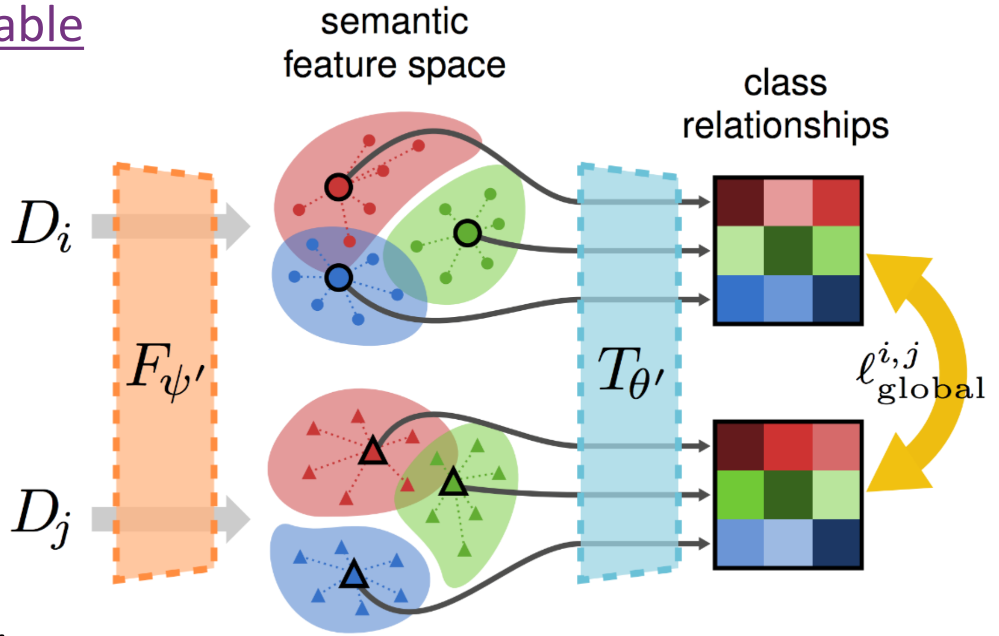
- In each domain, compute class-specific mean feature vector:

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

- 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:

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



(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:

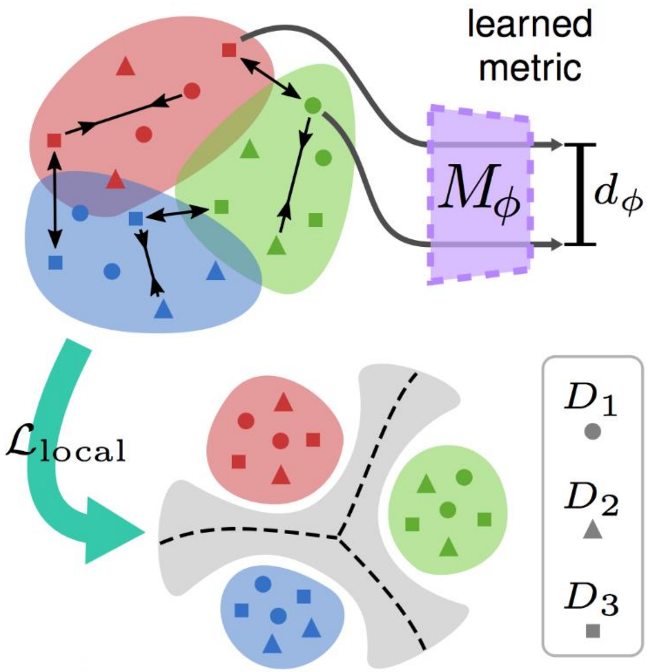
$$d_\phi(\mathbf{z}_n, \mathbf{z}_m) = \|\mathbf{e}_n - \mathbf{e}_m\|_2 = \|M_\phi(\mathbf{z}_n) - M_\phi(\mathbf{z}_m)\|_2$$

- 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]:

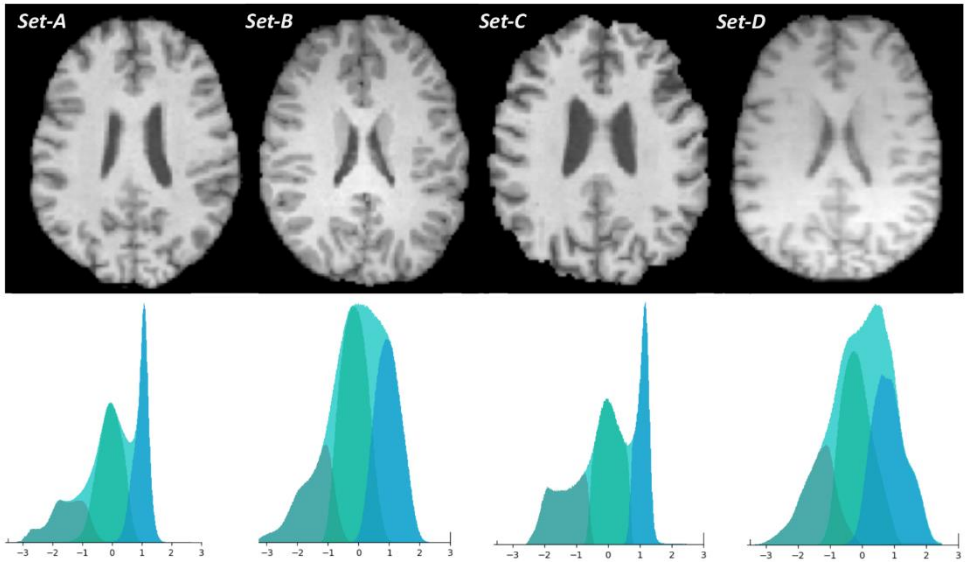
$$\ell_{\text{tri}}^{a,p,n} = \max\{0,\ d_\phi(\mathbf{z}_a, \mathbf{z}_p)^2 - d_\phi(\mathbf{z}_a, \mathbf{z}_n)^2 + \xi\}$$



MASF: Model-Agnostic Learning of Semantic Features

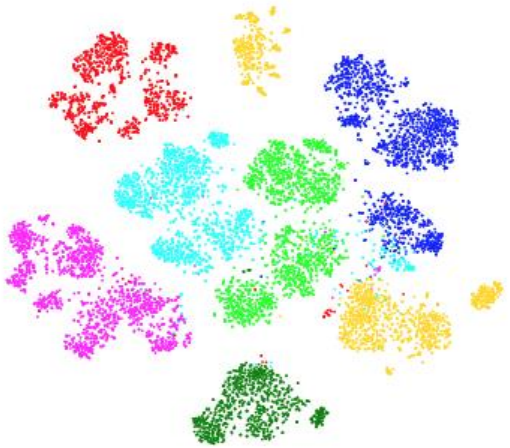
Medical application of brain tissue segmentation

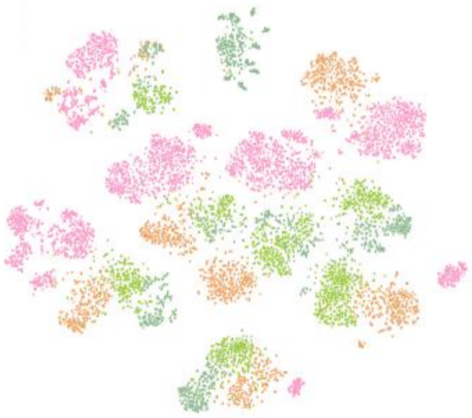
- data acquisition differences in scanners, imaging protocols, and many other factors
- posing severe limitations for translating learning-based methods in clinical practice
- segmentation of 3 brain tissues: white matter, gray matter and cerebrospinal fluid
- 4 domains corresponding to 4 hospitals

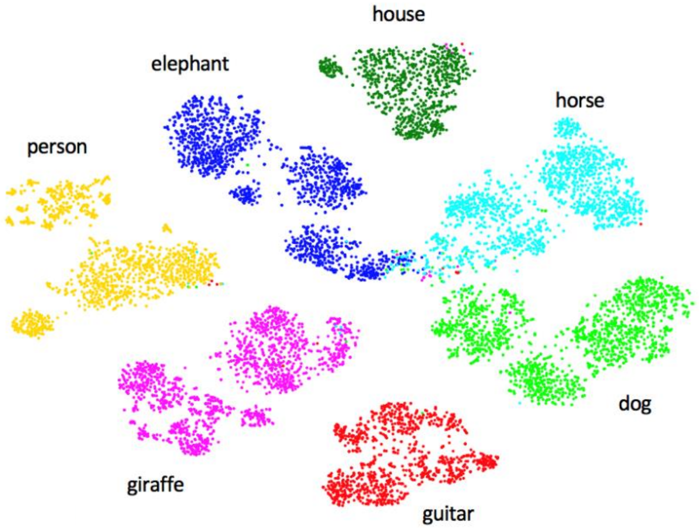


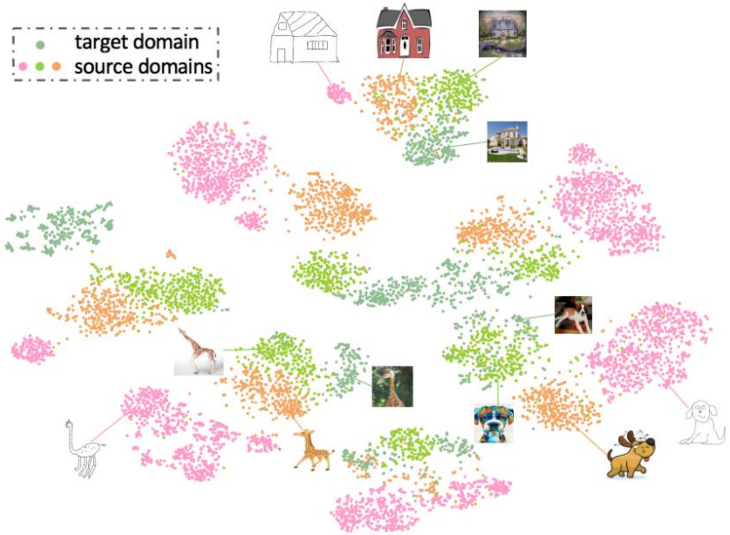
Test \ Train	Set-A	Set-B	Set-C	Set-D	DeepAll	MASF
Set-A	90.62	88.91	88.81	85.03	89.09	89.82
Set-B	85.03	94.22	81.38	88.31	90.41	91.71
Set-C	93.14	92.80	95.40	88.68	94.30	94.50
Set-D	76.32	88.39	73.50	94.29	88.62	89.51

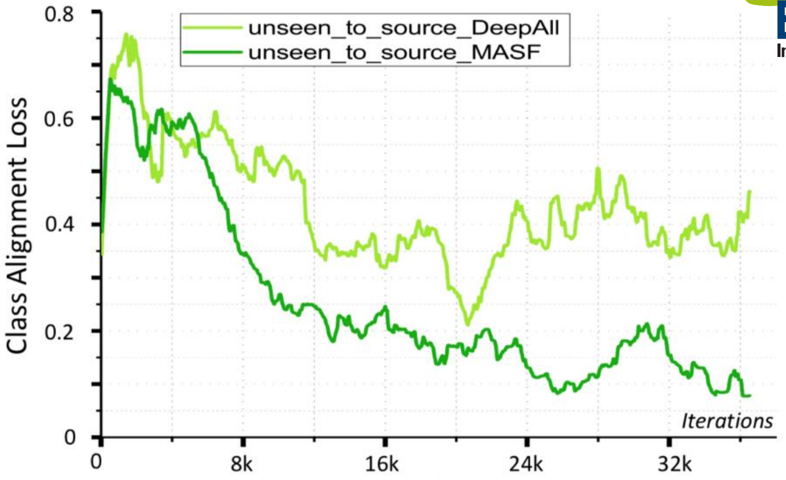
MASF: Model-Agnostic Learning of Semantic Features

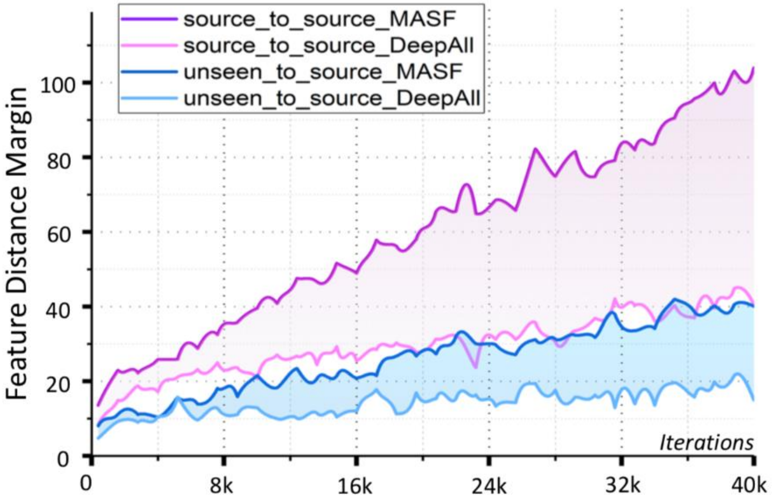












Datasets

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

Acknowledgement

Cheng Chen

Quande Liu

Cheng Ouyang

Daniel C. Castro

Konstantinos Kamnitsas

Paper & Code Available at:

Thanks for your attention!
Q & A