AI Alignment in Medical Image Segmentation: Model-Fitting, Robustness and Reassurance

Zeju Li, PhD
FMRIB Center, Nuffield Department of Clinical Neurosciences,
University of Oxford, UK.
Email: zeju.li@ndcn.ox.ac.uk
March 2024

Statistics
3,288 citations
12 first authored peer-reviewed journal/conference

Research Experience

Education

- 2023.1 – Current: FMRIB Analysis Group
  - Senior Researcher in Brain Connectivity, University of Oxford, Oxford
- 2018.10 – 2022.12: BioMedIA Group
  - PhD in Computing, Imperial College London, London
  - Supervisors: Prof. Ben Glocker and Prof. Daniel Rueckert
  - Examiners: Dr. Juan Eugenio Iglesias and Dr. Stamatia Giannarou
- 2011.9 – 2018.7:
  - Master in Biomedical Engineering, Fudan, Shanghai
  - Bachelor in Electronic Engineering, Fudan, Shanghai

Intern

- 2019.7 – 2020.4: Computer Vision Group
  - Huawei Noah's Ark Lab, London
- 2018.7 – 2018.9: MIRACLE Group
  - Institute Of Computing Technology, Beijing

Medical Image Segmentation

- Identify groups of pixels that go together **for** medical imaging (e.g. MRI, CT, ultrasound…).
- A critical step to transfer the power of machine learning into the clinical diagnosis process.

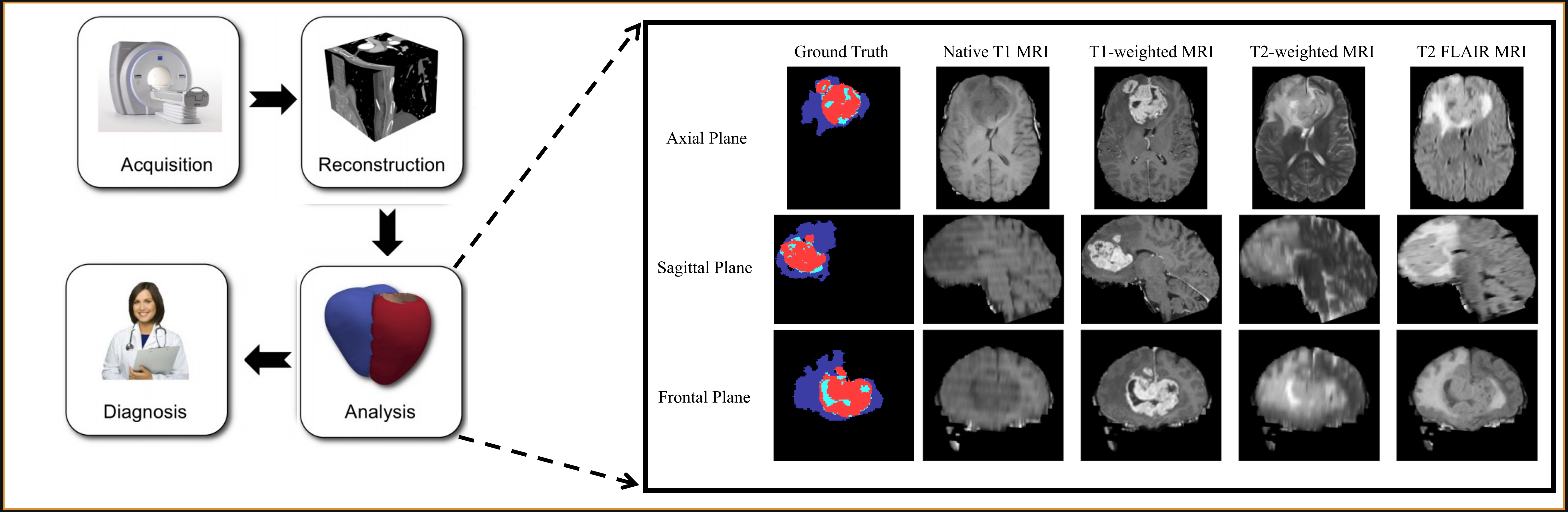


Figure credited to D. Rueckert, J.A. Schnabel. “Model-based and data-driven strategies in medical image computing“, P IEEE, 2019.

AI Alignment in Medical Image Segmentation

The model is unreliable
because of misaligned goals!

Model
Performance

Deployment
Environment

Optimization
Algorithms

Data
Distributions

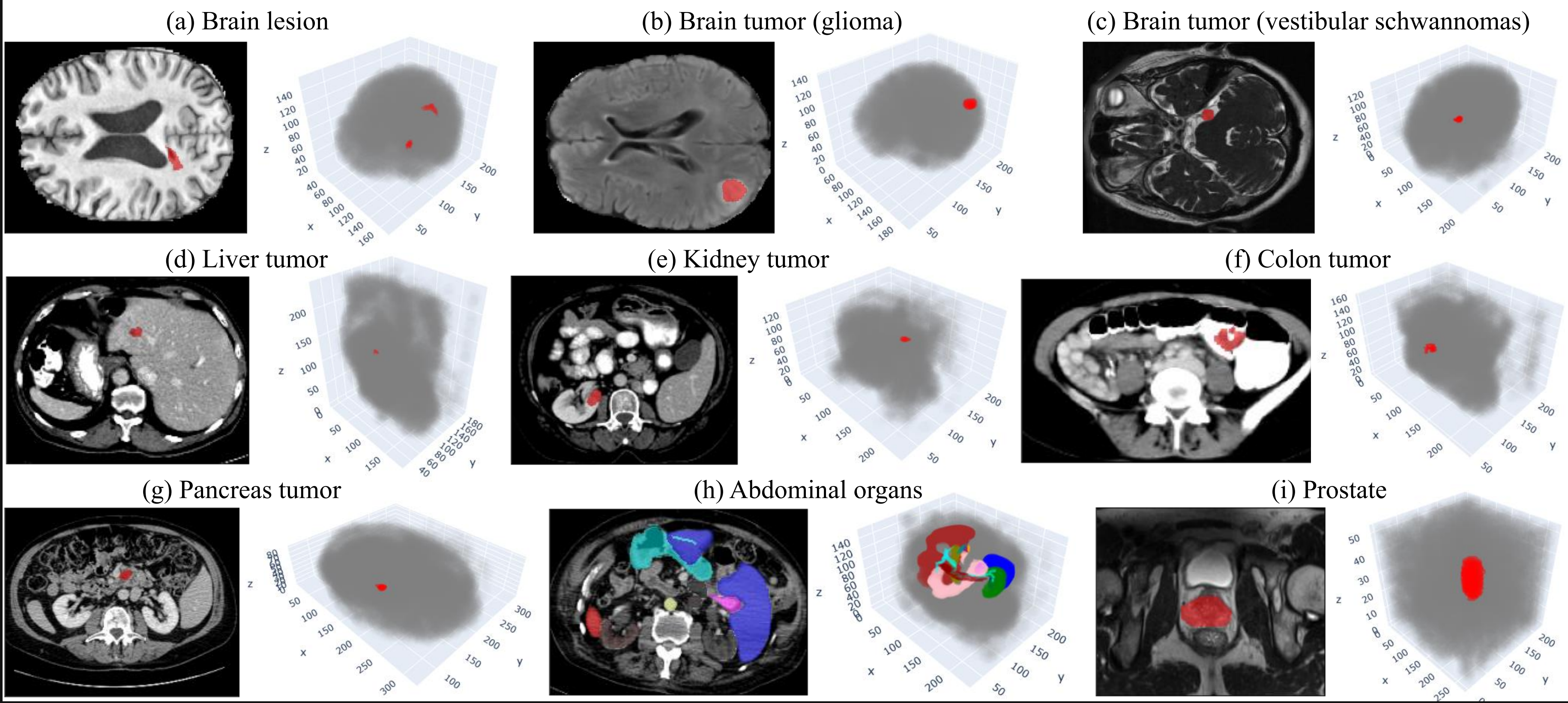
Reassurance: Expectation Over-optimism Problem.

Robustness: Goal Mis-generalization Problem.

Model-Fitting: Specification Failure Problem.

Specification Failure Reason: Class Imbalance

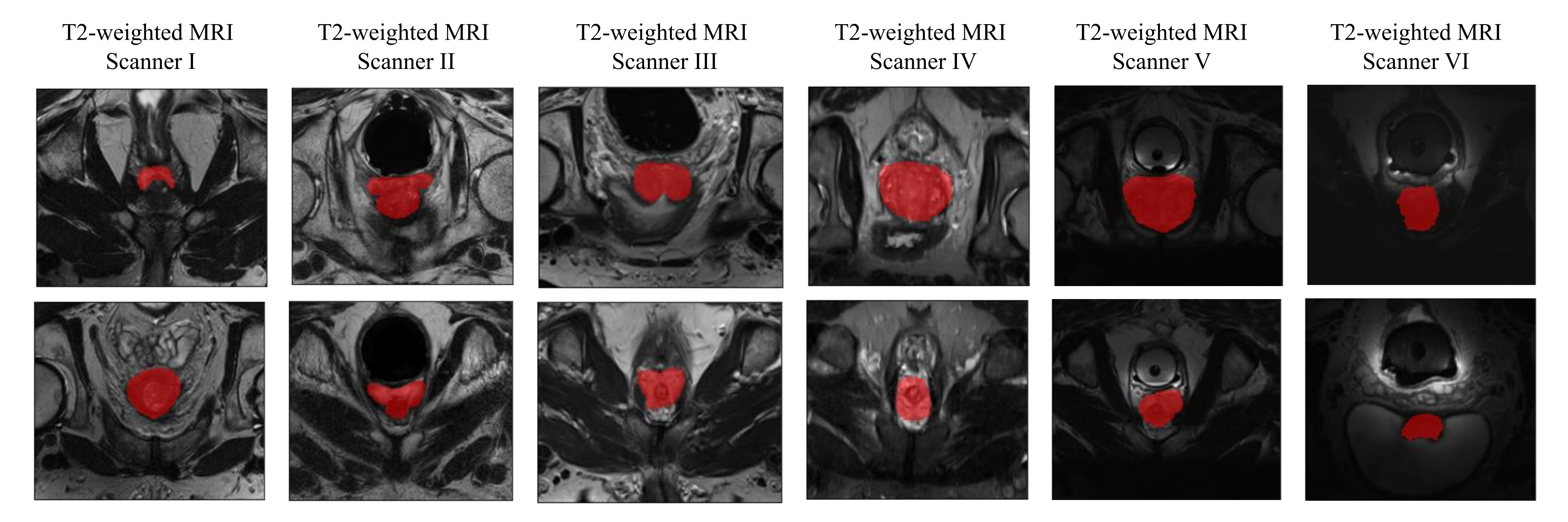
- **Class imbalance**, which refers to the imbalanced distributions of samples from different categories, cause difficulties for machine learning models to learn well.



AI Alignment in Medical Image Segmentation

Goal Mis-generalization Reason: Domain Shifts

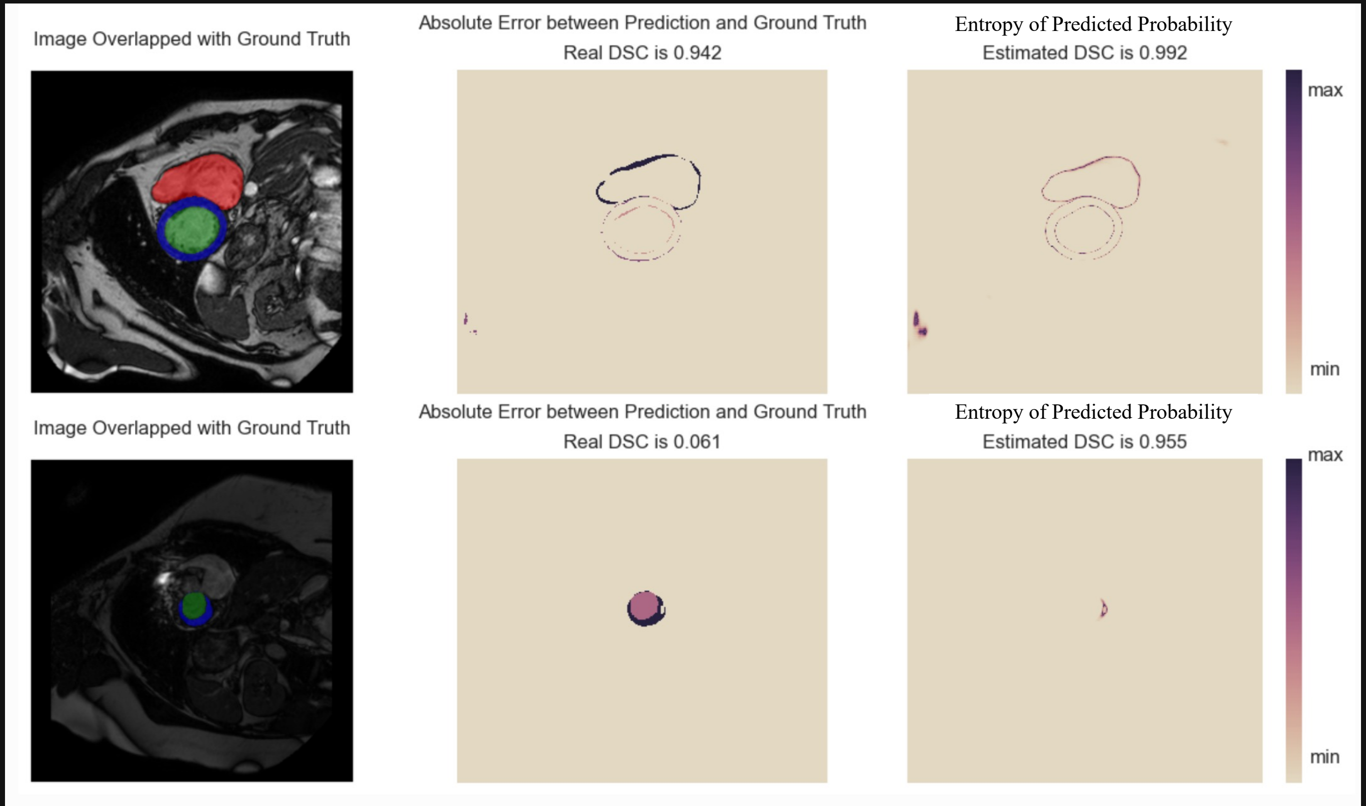
- **Domain shift is** a common issue in medical imaging as the medical images can be collected from different clinical sites. As data-driven methods, deep neural network may not generalize well under domain shifts.

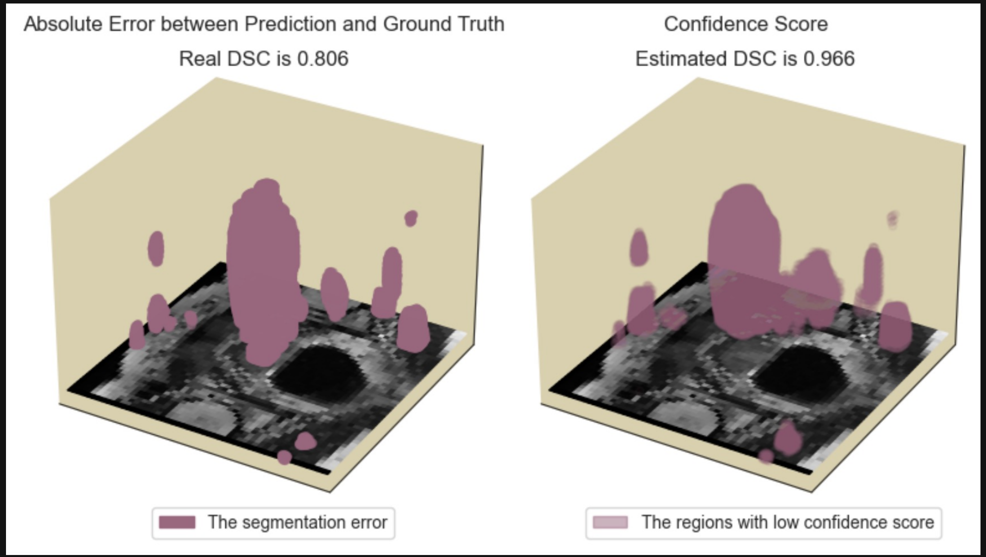


Example from Q. Liu *et al.* "Shape-aware meta-learning for generalizing prostate MRI segmentation to unseen domains ", MICCAI, 2020.

Expectation Over-optimism Reason: Model Uncalibration

- As probabilistic model, neural network produces the probability of the predictions. However, modern deep neural networks are known to **over-confident and uncalibrated** about the predictions.

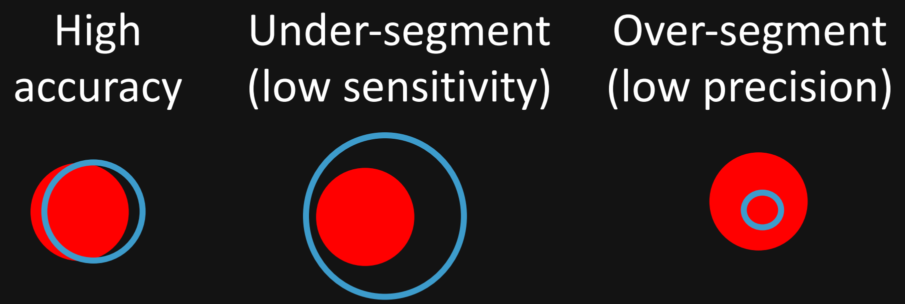


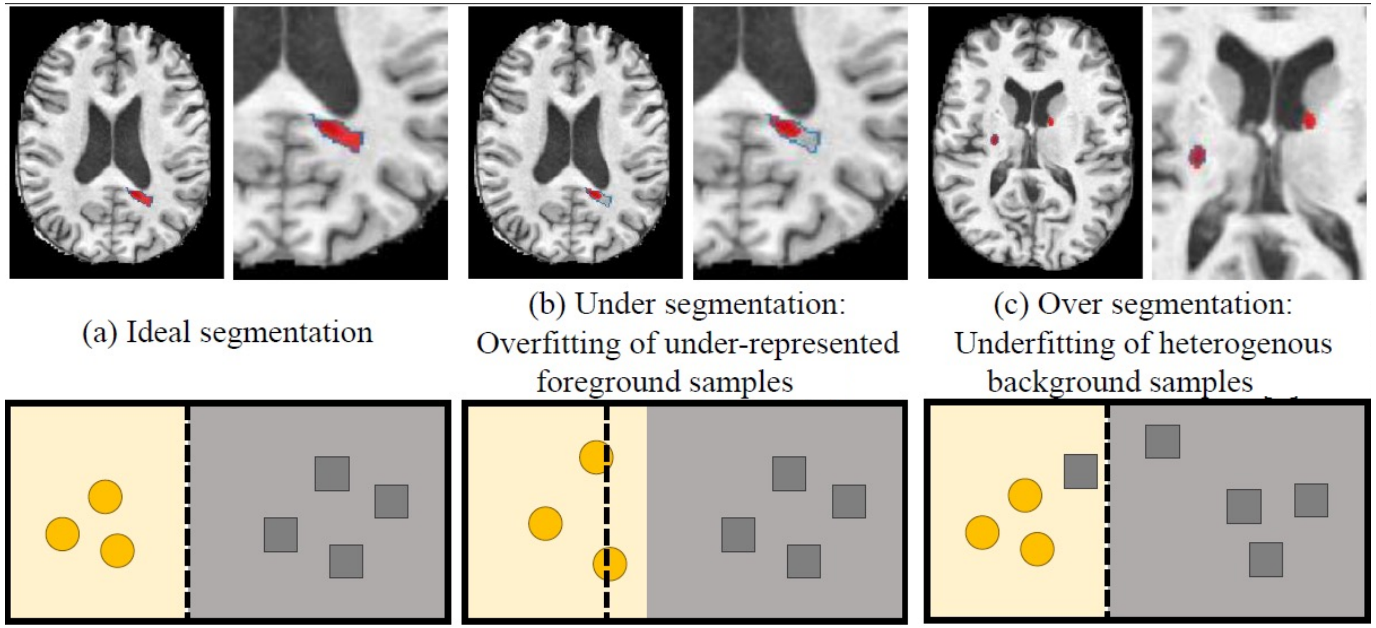


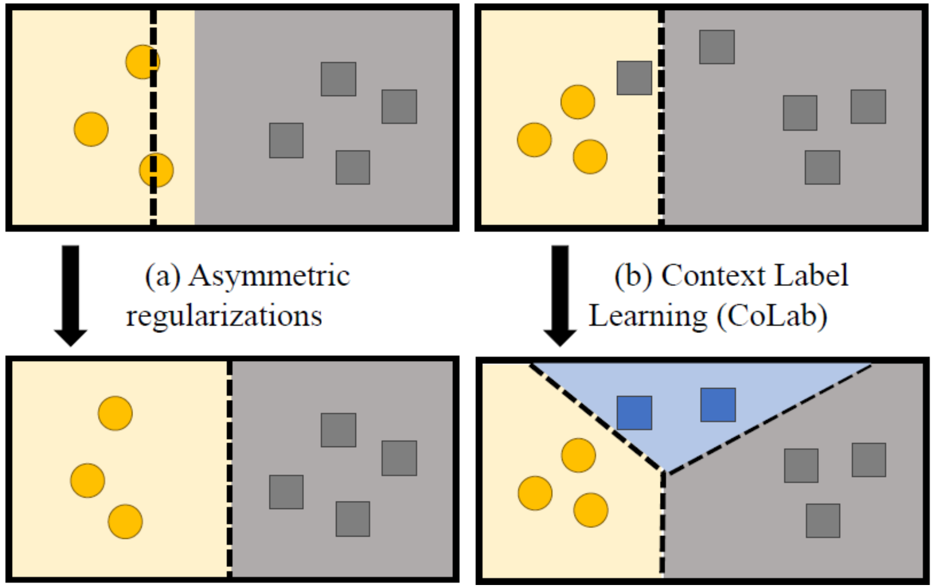
Effects of Class Imbalance

Observations

- Class imbalance causes **under- and over-segmentation**.
- Understanding the effects of class imbalance in segmentation.



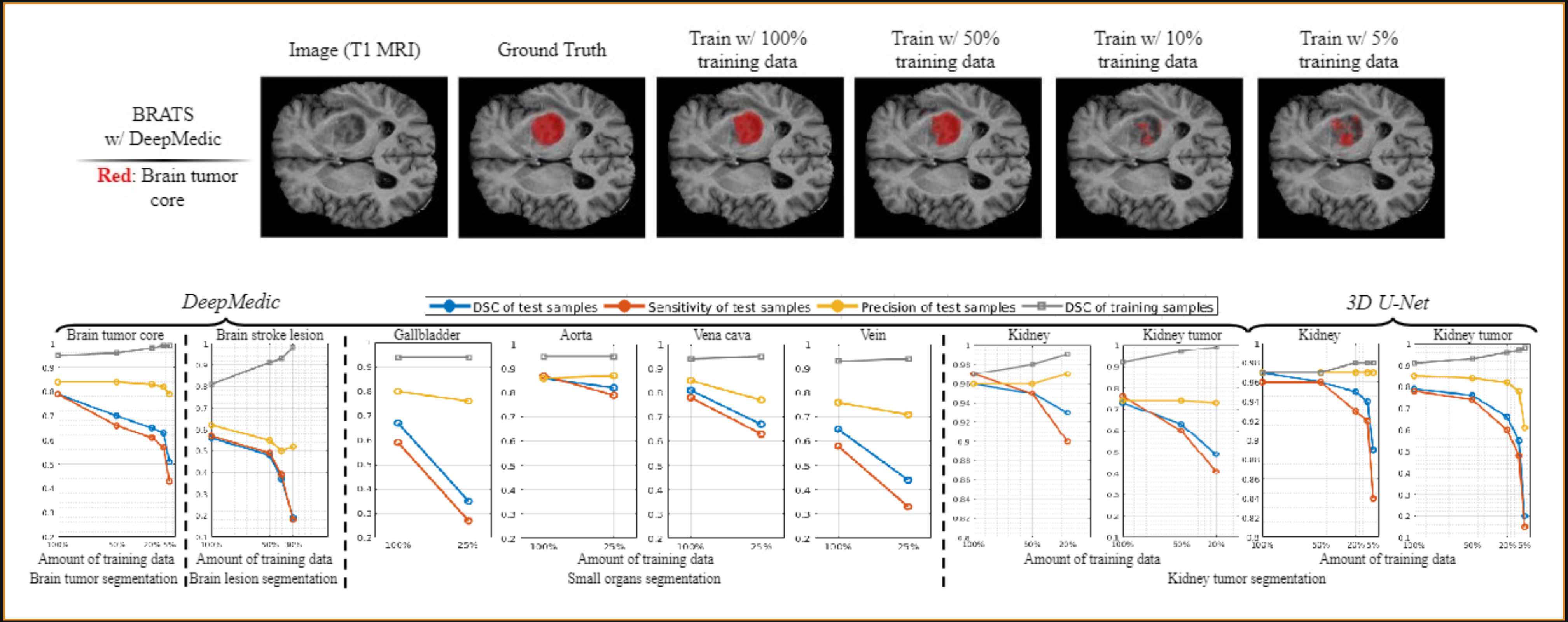




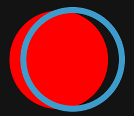
Overfitting under Class Imbalance

Analysis

➢ With less training data, performances decline due to the drastic **reduction of sensitivity**, while precision is retained.



High accuracy



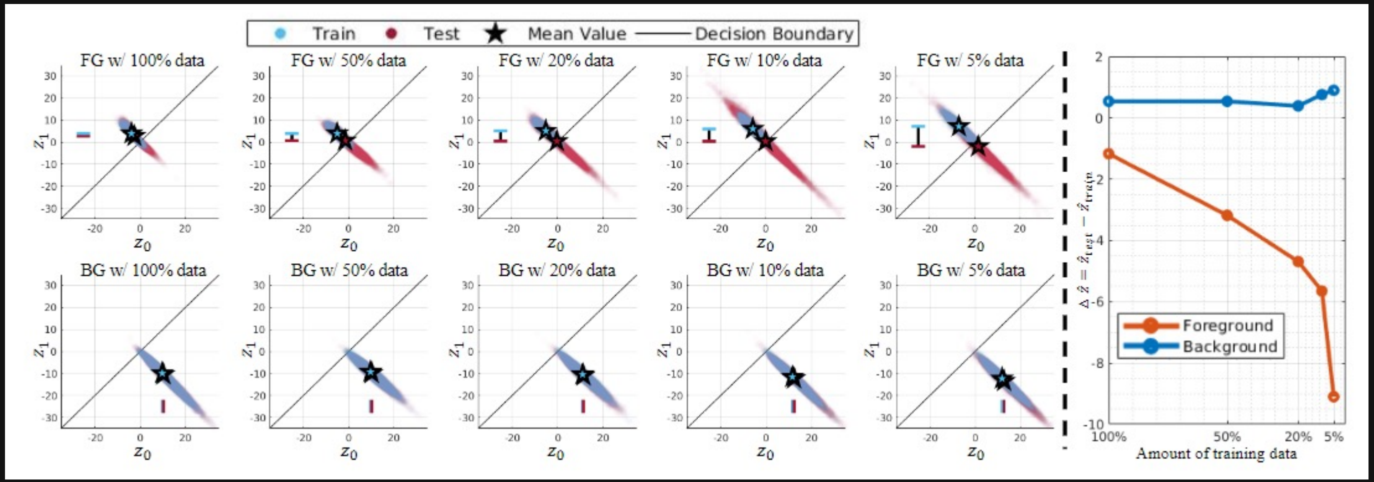
Under-segment (low sensitivity)

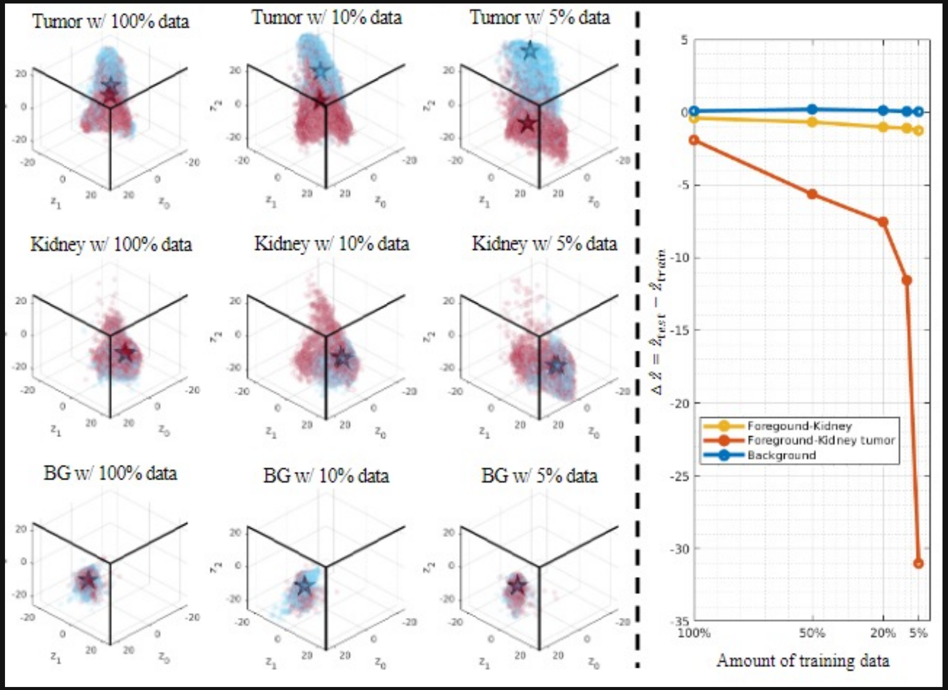
Z. Li *et al.* "Analyzing overfitting under class imbalance in neural networks for image segmentation", TMI, 2020. [MICCAI Student Travel Award]

Overfitting under Class Imbalance

Analysis

- CNN maps training and testing samples of the background class to similar logit values.
- However, mean activation for testing data shifts significantly for the foreground class towards and sometimes across the decision boundary.



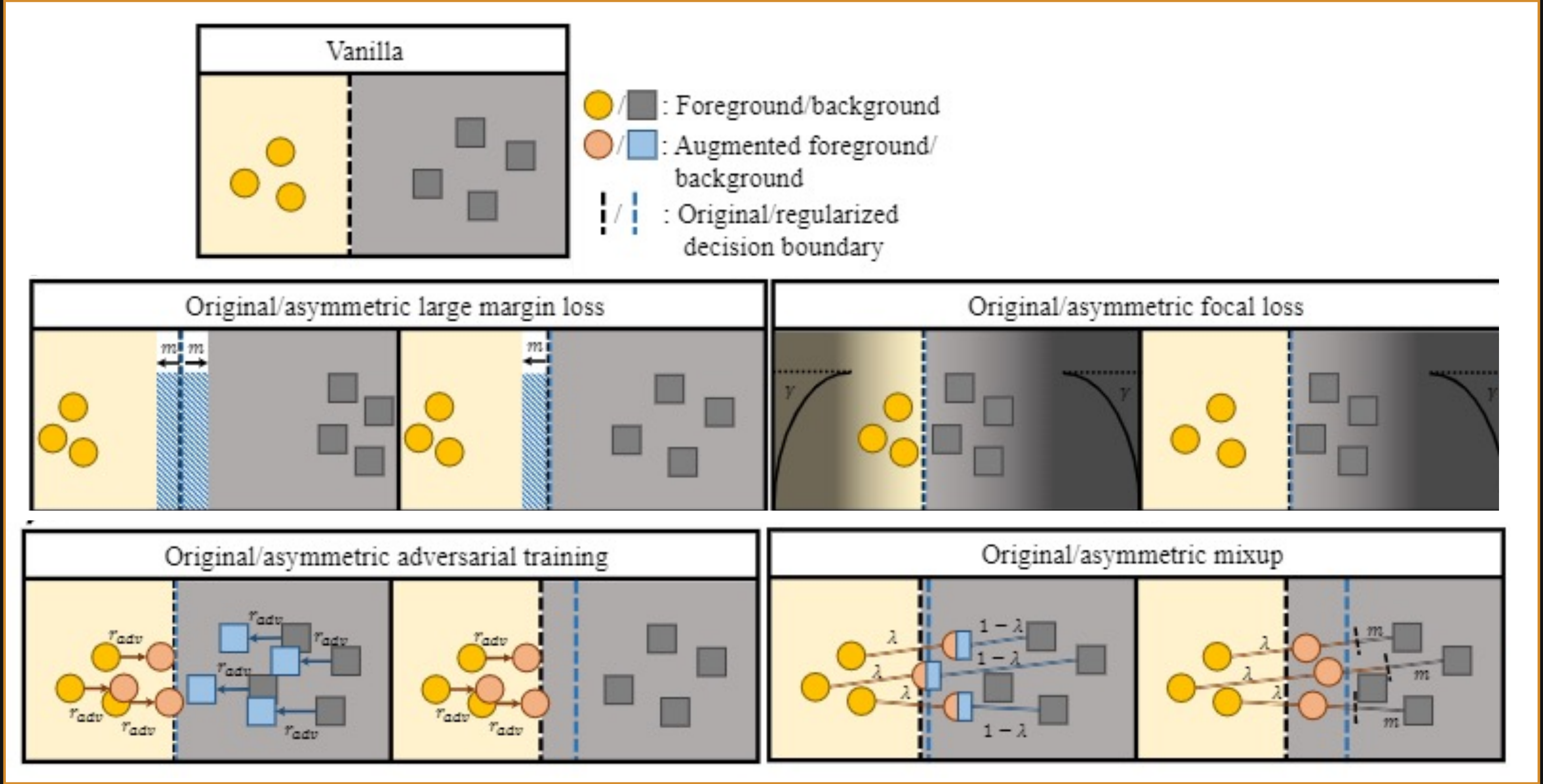


Z. Li *et al.* "Analyzing overfitting under class imbalance in neural networks for image segmentation", TMI, 2020. [MICCAI Student Travel Award]

Overfitting under Class Imbalance

Method and Results

➢ We make the logit activations of foreground class far away from the decision boundary by **setting bias for the foreground class** in different ways.



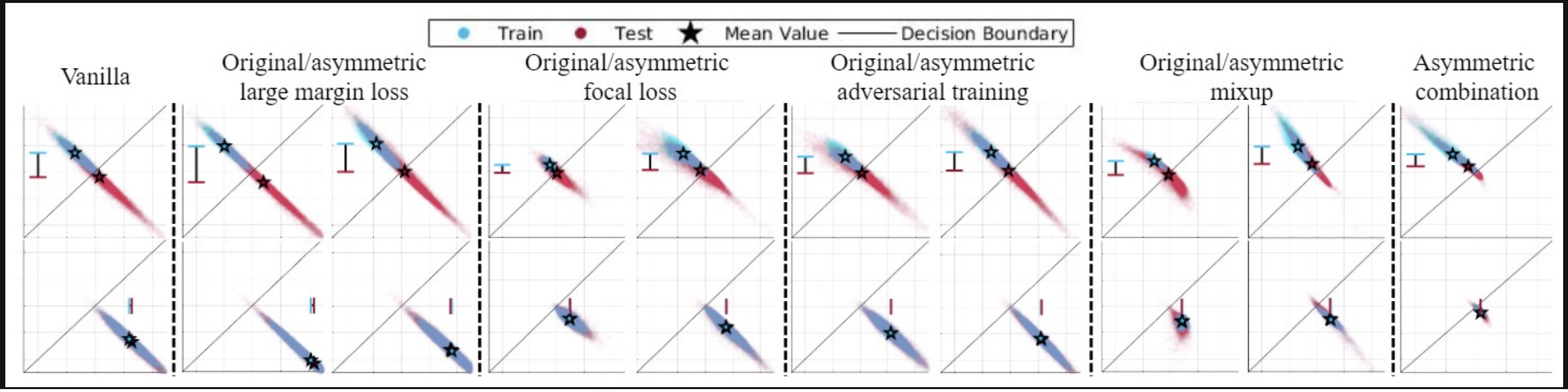
Method	5% training			
	DSC	SEN	PRC	HD
Vanilla - CE [20]	50.4	41.0	83.5	18.0
Vanilla - CE - 80% tumor	45.5	36.0	86.7	17.8
Vanilla - F1 (DSC)	47.2	37.4	86.6	15.9
Vanilla - F2 [14]	45.8	36.9	81.9	17.9
Vanilla - F4 [14]	51.6	42.5	83.8	18.1
Vanilla - F8 [14]	47.4	38.7	83.1	19.6
Large margin loss [31]	44.5	35.9	82.8	20.2
Asymmetric large margin loss	56.8	48.9	83.4	15.0
Focal loss [29]	54.0	44.8	82.6	16.0
Asymmetric focal loss	58.8	51.4	81.6	15.0
Adversarial training [12]	53.2	44.6	85.0	19.2
Asymmetric adversarial training	58.5	50.8	80.1	16.2
Mixup [47]	49.7	40.9	83.0	19.6
Asymmetric mixup	**59.8**	56.8	74.7	17.7
Symmetric combination	50.0	42.0	84.6	21.1
Asymmetric combination	63.4	63.1	75.9	**15.1**

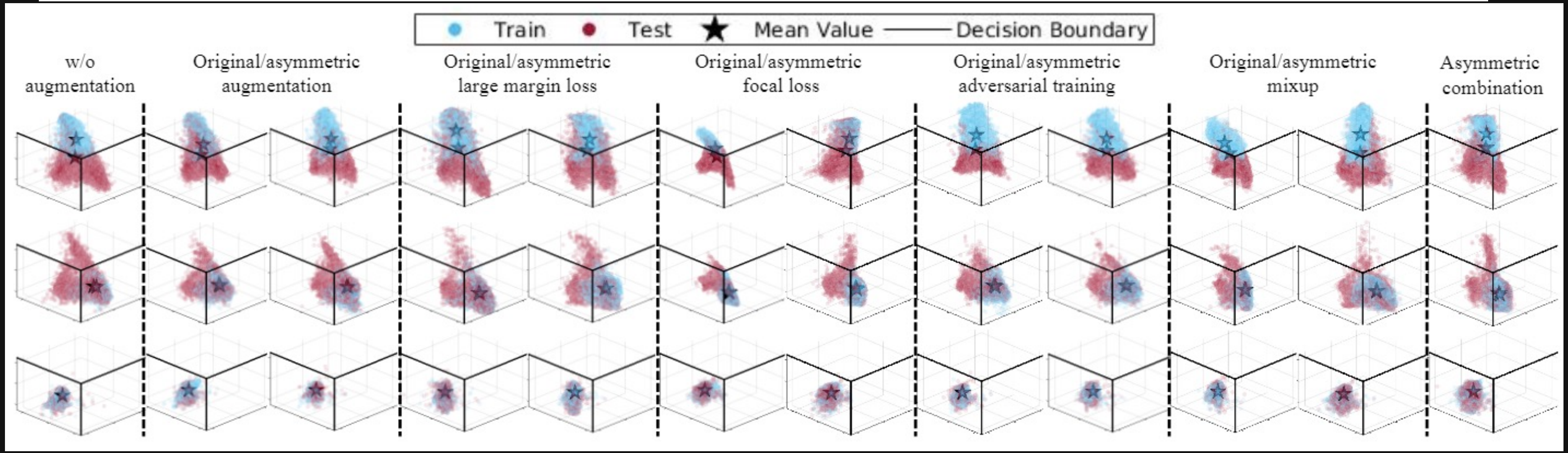
Z. Li *et al.* "Analyzing overfitting under class imbalance in neural networks for image segmentation", TMI, 2020. [MICCAI Student Travel Award]

Overfitting under Class Imbalance

Results

- Asymmetric modifications lead to better separation of the logits of unseen foreground samples.



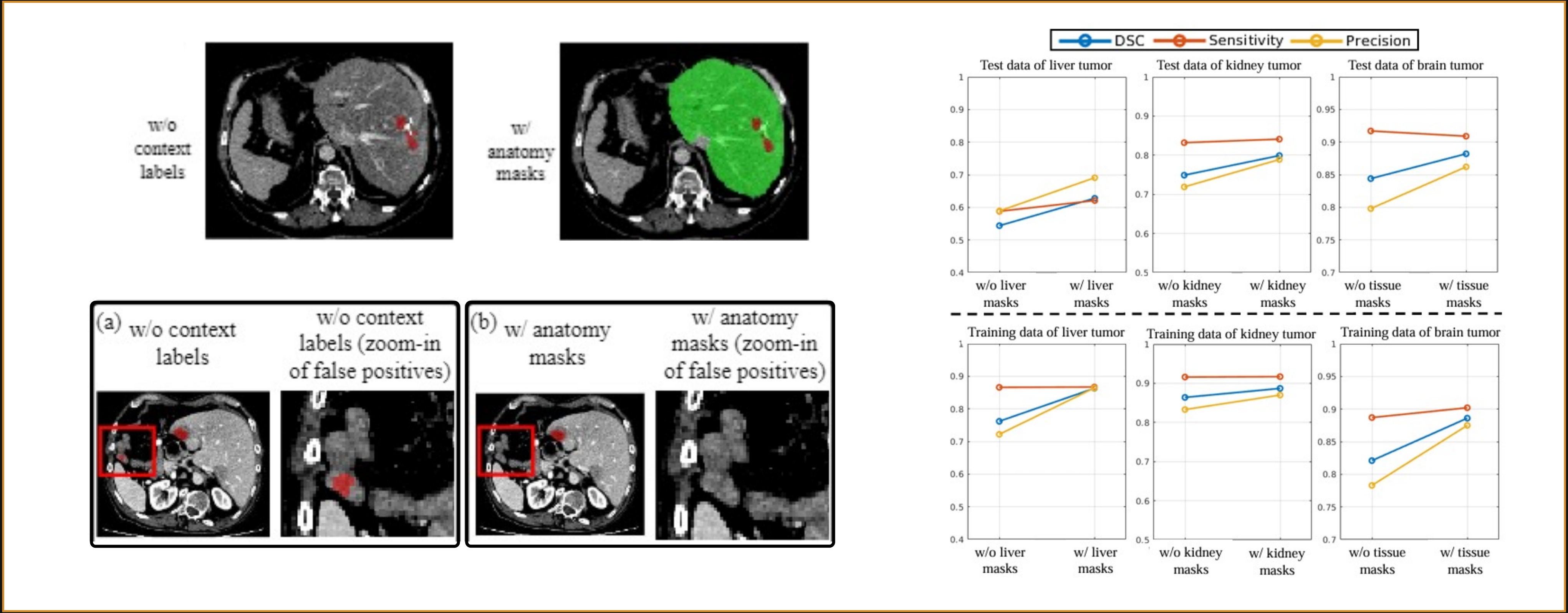


Z. Li *et al.* "Analyzing overfitting under class imbalance in neural networks for image segmentation", TMI, 2020. [MICCAI Student Travel Award]

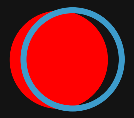
Underfitting under Class Imbalance

Analysis

➢ With heterogeneous background, performances decline due to the drastic **reduction of precision**, while sensitivity is retained.



High accuracy



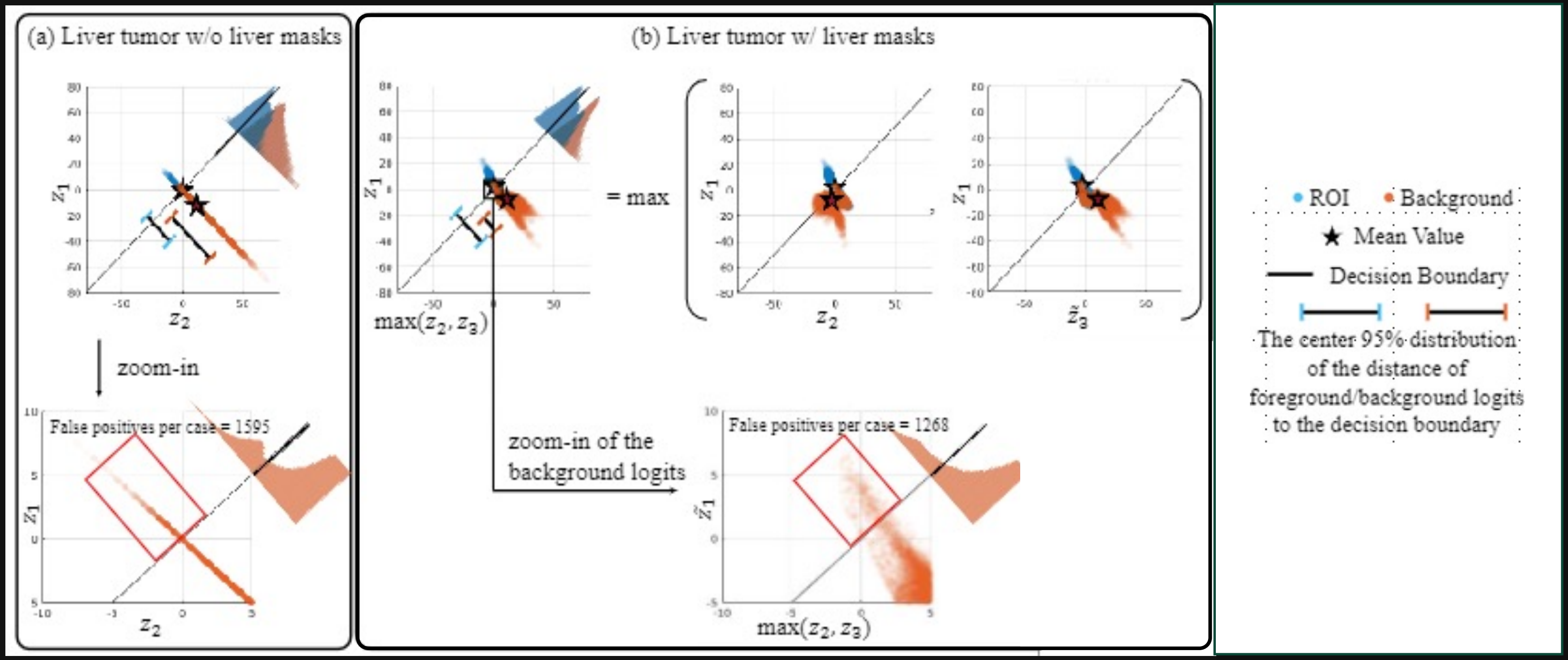
Over-segment (low precision)

Z. Li *et al.* “Context label learning: improving background class representations in semantic segmentation“, TMI, 2023.

Underfitting under Class Imbalance

Analysis

- Neural networks could not map the heterogeneous background samples to **compact clusters** in feature space.
- As a result, the logit activations of background would approach and even move across the decision boundary.

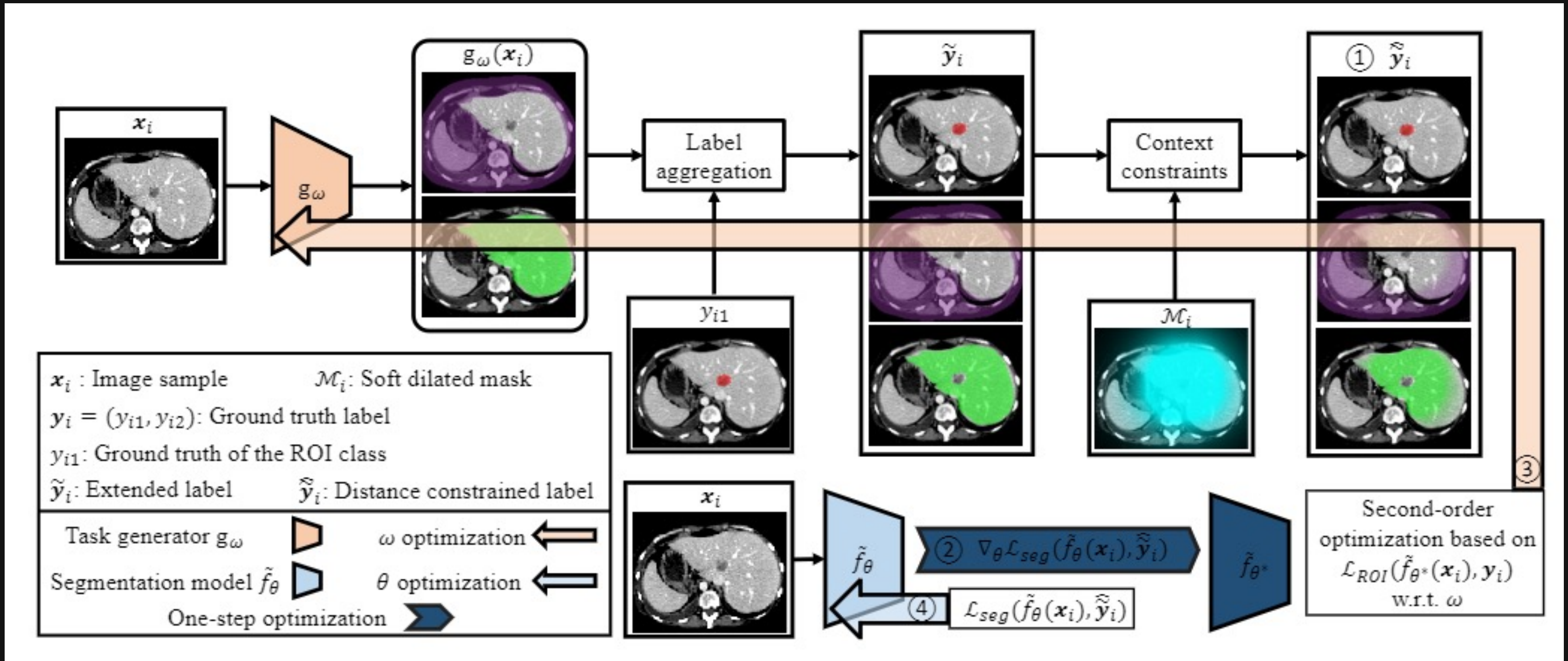


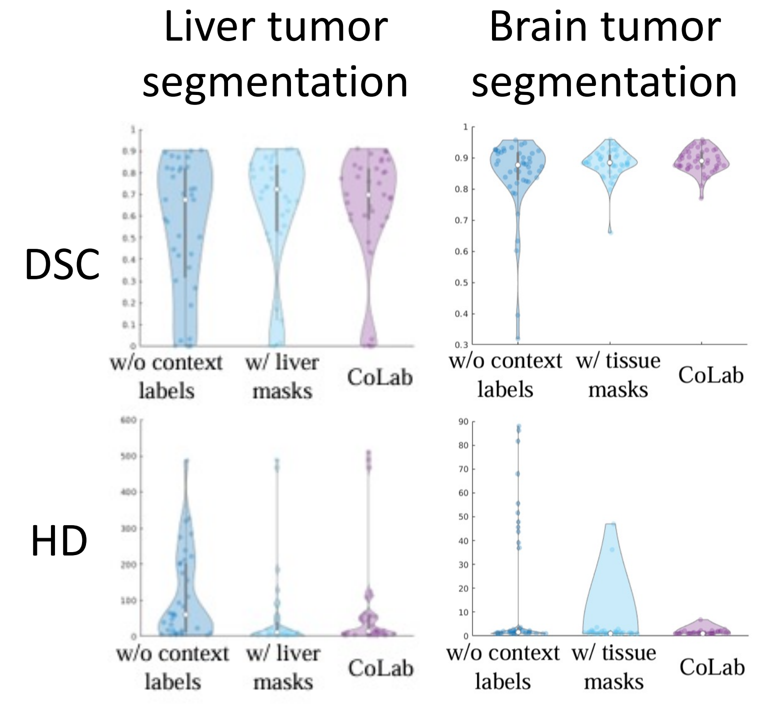
Z. Li *et al.* "Context label learning: improving background class representations in semantic segmentation", TMI, 2023.

Underfitting under Class Imbalance

Method and Results

- We propose **Context label learning (CoLab)**.
- We train an auxiliary network as a task generator, along with the primary segmentation model, to automatically generate context labels that positively affect the ROI segmentation accuracy.





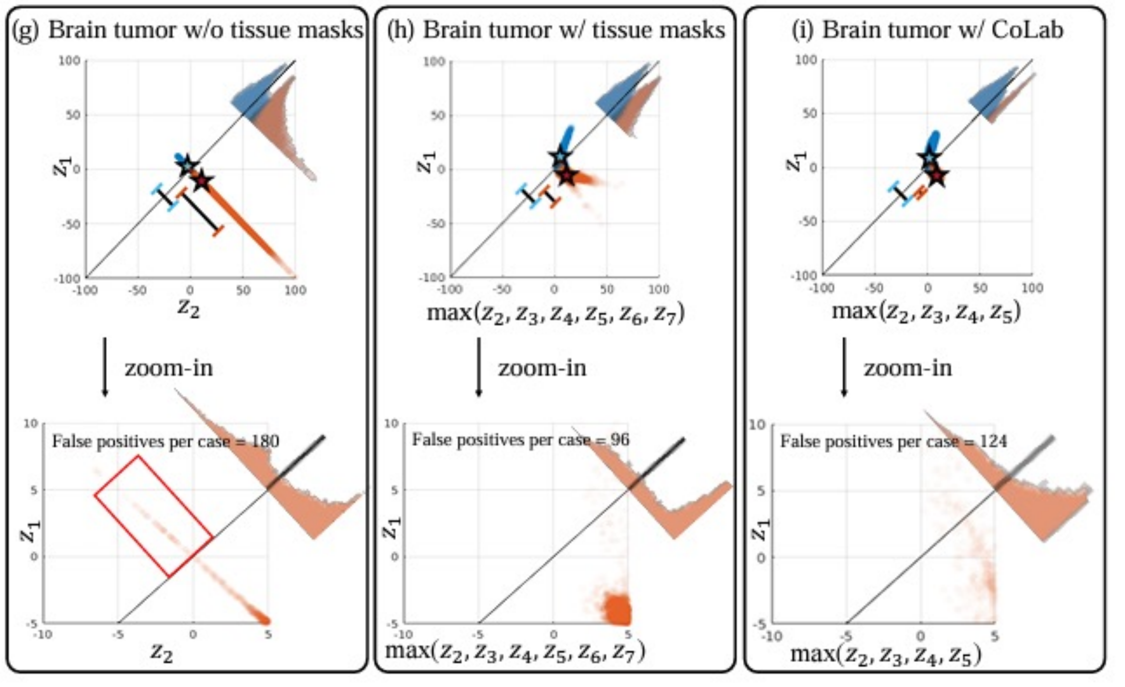
Z. Li *et al.* "Context label learning: improving background class representations in semantic segmentation", TMI, 2023.

Underfitting under Class Imbalance

Results

➢ Similar and sometimes better effect in improving segmentation accuracy when compared with human-defined context labels.

Task	Method	t	DSC	SEN	PRC	HD
Liver tumor [5]	w/o liver masks	1	54.4	58.8	58.9	111.1
	K-means [1]	2	**61.4**	**61.4**	67.0	71.9
	Dilated masks [26]	2	60.7	59.8	68.0	67.6
	CoLab	2	62.5	62.8	**67.3**	69.4
	CoLab	4	57.3	60.5	62.3	**56.3**
	CoLab	6	59.7	60.3	65.2	43.6
	w/ model-predicted liver masks [16]	2	62.4	61.6	70.6	44.1
	w/ liver masks [5]	2	62.8	62.1	69.1	53.5
Kidney tumor [12]	w/o kidney masks	1	74.9	83.2	71.9	120.4
	K-means [1]	2	**76.8**	**83.5**	74.3	87.1
	Dilated masks [26]	2	76.4	83.9	73.1	95.3
	CoLab	2	78.5	82.2	77.7	**75.7**
	CoLab	4	76.4	80.6	**76.5**	63.7
	CoLab	6	74.9	81.0	73.3	79.4
	w/ model-predicted kidney masks [16]	2	79.2	81.3	82.7	38.1
	w/ kidney masks [12]	2	79.9	84.1	78.9	54.7
Brain tumor [33]	w/o tissue masks	1	84.3	91.2	80.5	15.2
	K-means [1]	6	85.0	91.3	80.3	9.4
	Dilated masks [26]	2	84.8	91.6	81.1	8.8
	CoLab	2	85.2	90.4	82.5	7.8
	CoLab	4	89.0	**91.7**	86.7	1.4
	CoLab	6	**87.9**	92.0	**84.9**	**2.5**
	w/ tissue masks [17], [33]	6	88.2	90.9	86.2	3.1

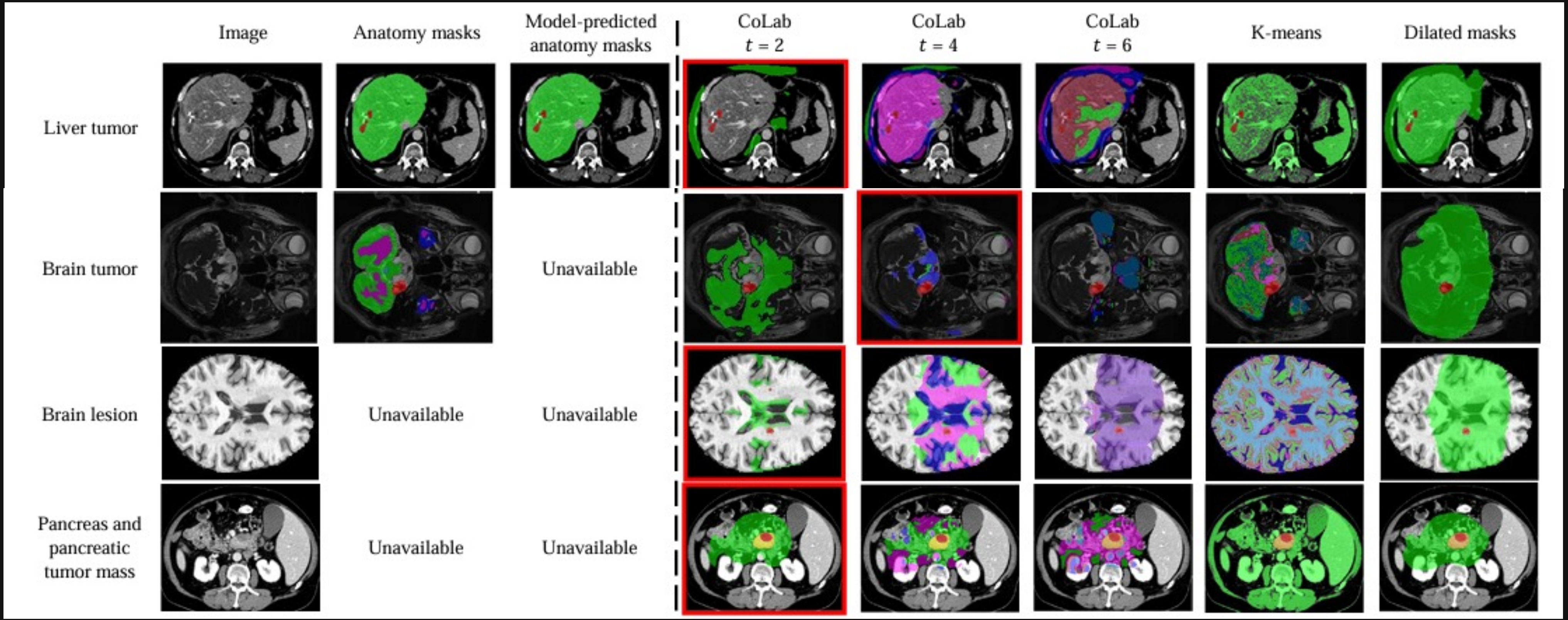


Z. Li *et al.* “Context label learning: improving background class representations in semantic segmentation“, TMI, 2023.

Underfitting under Class Imbalance

Results

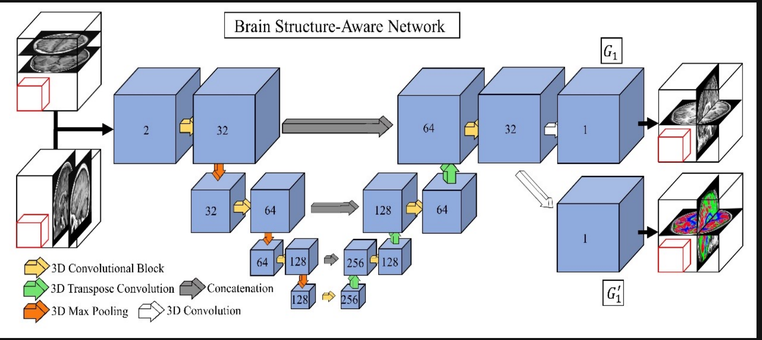
- Examples of context labels generated could inform us on how to design optimal contextual tasks.



Z. Li *et al.* “Context label learning: improving background class representations in semantic segmentation“, TMI, 2023.

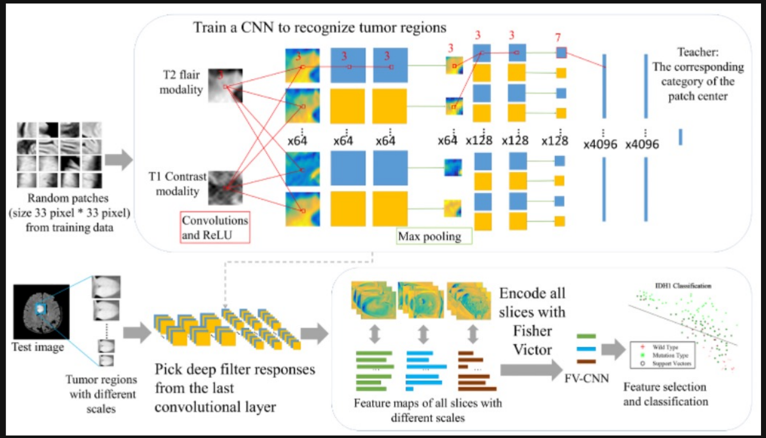
Other Related Works to Improve Model-Fitting

Improved Model-Fitting with **Multi-Task Learning**



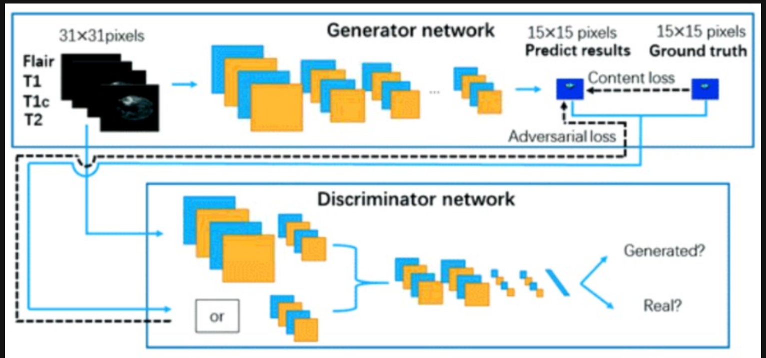
Z. Li *et al.* "Deepvolume: brain structure and spatial connection-aware network for brain MRI super-resolution", TCybern, 2019.

Feature Re-using for Downstream Applications



Z. Li *et al.* "Deep learning based radiomics (DLR) and its usage in noninvasive idh1 prediction for low grade glioma", Sci. Rep., 2017.

Generative Model based Segmentation

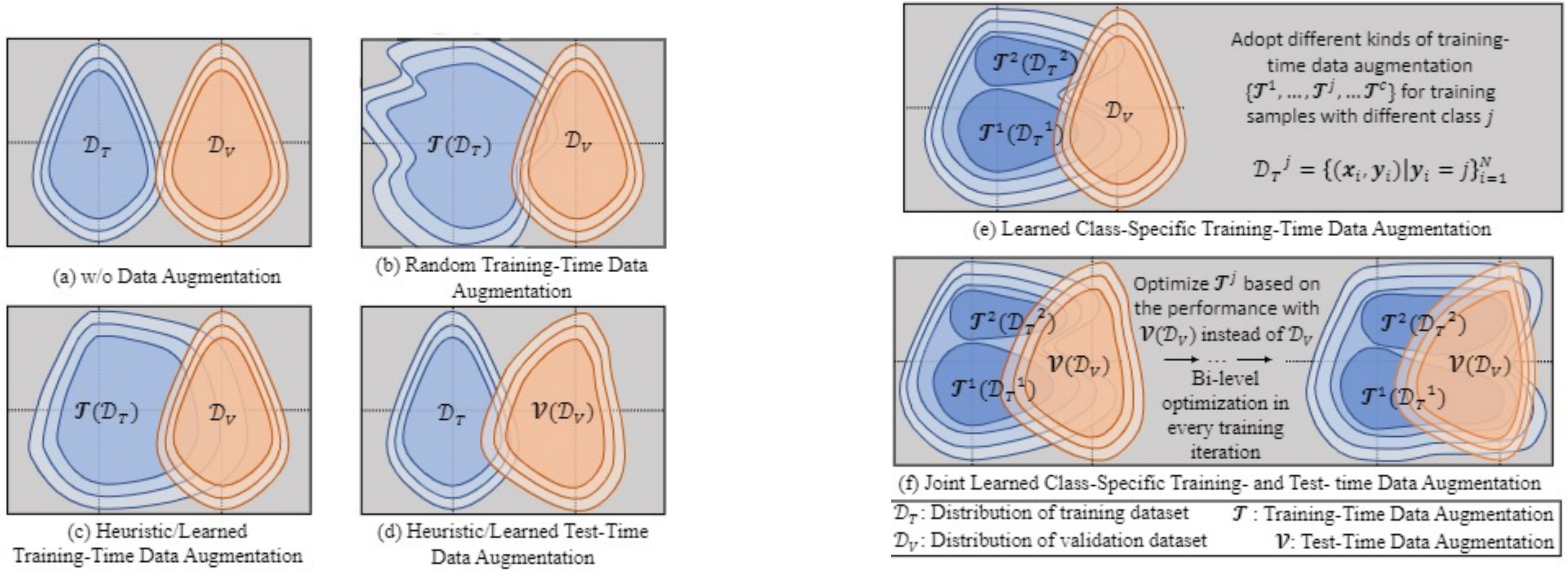


Z. Li *et al.* "Brain tumor segmentation using an adversarial network", MICCAI-Brainlesion workshop, 2017.

Joint Optimization of Class-Specific Training- and Test-time data augmentation

Motivation

- Training-time data augmentation (TRA) and test-time data augmentation (TEA) are closely connected as both aim to **align the training and test data distribution**.

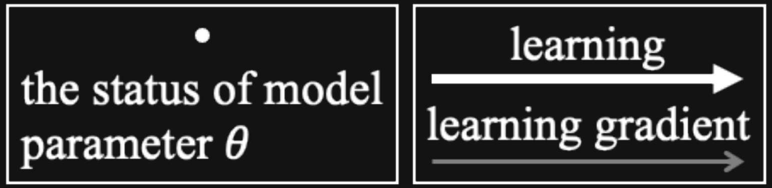


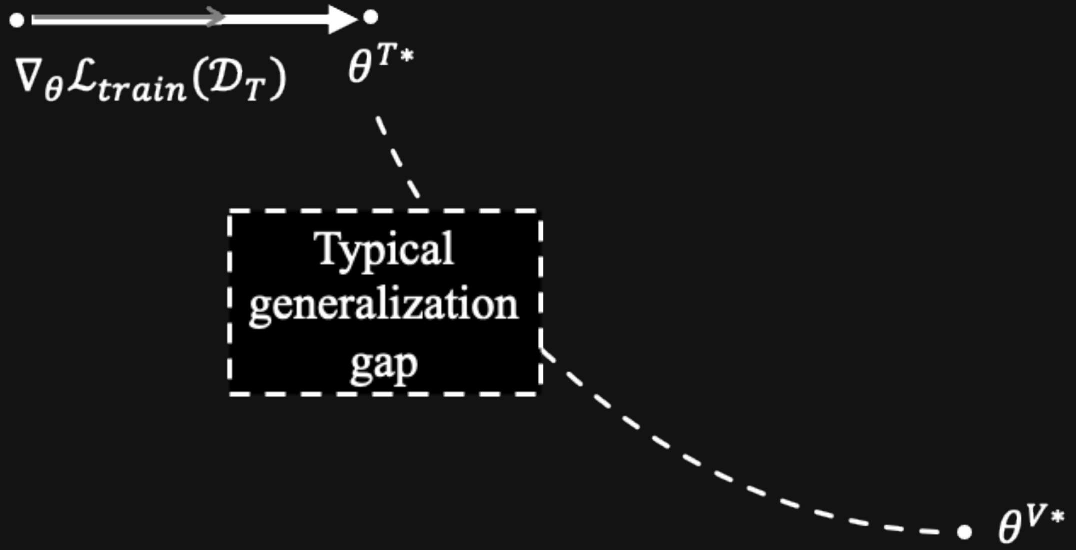
Z. Li *et al.* "Joint optimization of class-specific training- and test-time data augmentation in segmentation", TMI, 2023.

Joint Optimization of Class-Specific Training- and Test-time data augmentation

Method

- A meta-learning based data augmentation framework, taking test-time transformations into account.





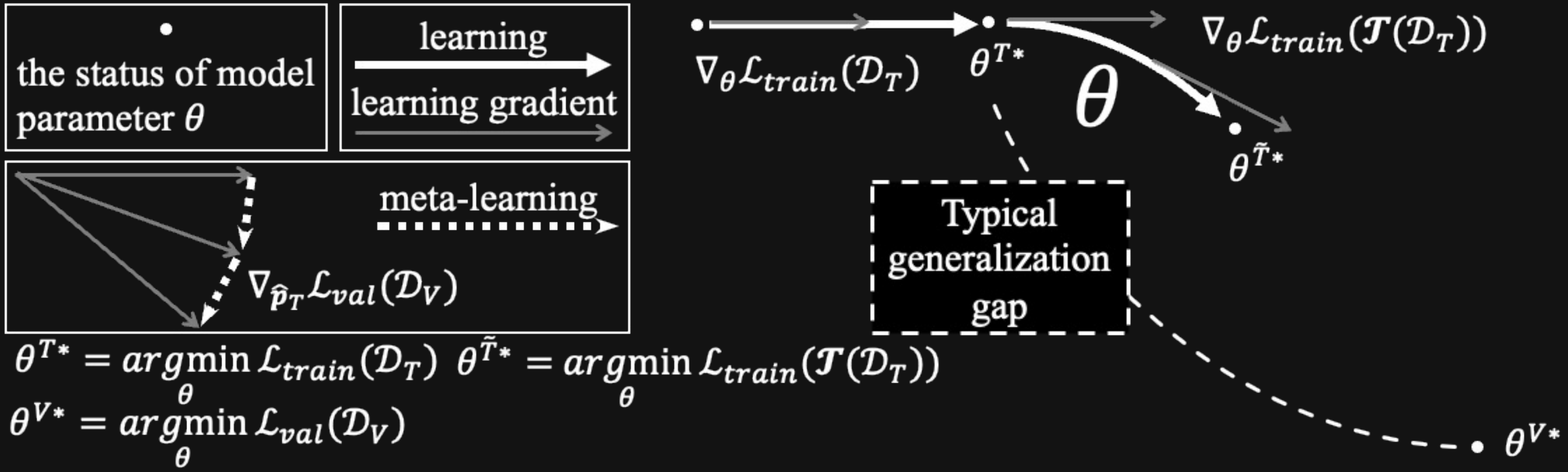
$$\theta^{T*} = \underset{\theta}{argmin}\, \mathcal{L}_{train}(\mathcal{D}_T)$$
$$\theta^{V*} = \underset{\theta}{argmin}\, \mathcal{L}_{val}(\mathcal{D}_V)$$

Z. Li *et al.* “Joint optimization of class-specific training- and test-time data augmentation in segmentation“, TMI, 2023.

Joint Optimization of Class-Specific Training- and Test-time data augmentation

Method

- A meta-learning based data augmentation framework, taking test-time transformations into account.

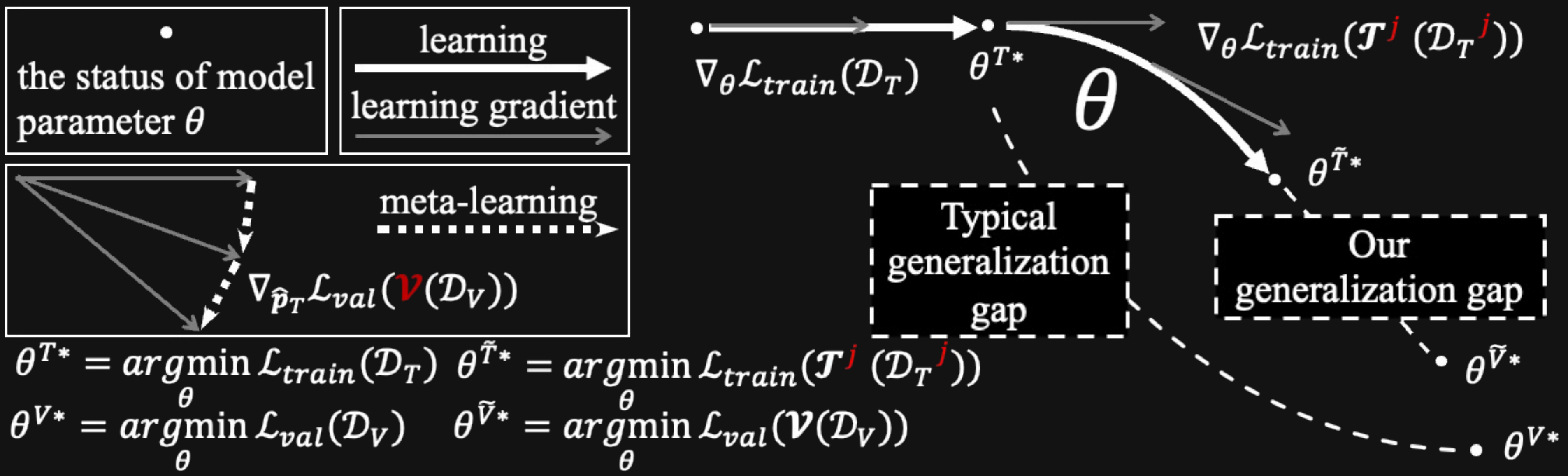


Z. Li *et al.* "Joint optimization of class-specific training- and test-time data augmentation in segmentation", TMI, 2023.

Joint Optimization of Class-Specific Training- and Test-time data augmentation

Method

- A meta-learning based data augmentation framework, taking test-time transformations into account.



Z. Li *et al.* "Joint optimization of class-specific training- and test-time data augmentation in segmentation", TMI, 2023.

Joint Optimization of Class-Specific Training- and Test-time data augmentation

Results

- ➢ Consistently improve segmentation performance in various applications.
- ➢ Potential to **replace the heuristically chosen augmentation policies** currently used in most previous works.

Class-specific TRA improves brain tumor segmentation

Model	Training-time augmentation	Test-time augmentation	50% training data DSC ↑	SEN ↑	PRC ↑	HD ↓
3D U-Net [6]	None	None	54.6	56.3	**67.2**	**32.6**
	Heuristic [18]	None	58.4	66.9	61.4	39.0
	Heuristic† [18]	None	58.8	**67.8**	59.8	52.2
	Learned [8], [29], [32]	None	**59.3**	66.6	61.1	40.9
	Learned Class-Specific	None	**62.0** (+3.6)**	**68.8**	66.2	**37.8**
	Heuristic [18]	Heuristic [18]	61.7	**67.0**	69.6	22.0
	Learned Class-Specific	Heuristic [18]	61.8	66.4	**70.2**	**20.3**
	Learned Class-Specific	Learned [22], [40]	**62.2**	66.9	**69.9**	**20.5**
	Joint Learned Class-Specific		**62.3** (+0.6)~	**67.4**	69.2	28.9

$^{*}p$-value < 0.05; $^{**}p$-value < 0.01; $^{\sim}p$-value ≥ 0.05 (compared to Heuristic‡ TRA w/o TEA or Heuristic‡ TRA w/ Heuristic TEA)
†We pretrain these models with training data from site A and fine-tune with validation data from site B.
‡We train these models with both training data from site A and validation data from site B.

Joint optimization of TRA and TEA improves cross-domain prostate segmentation

Model	Site A Training-time data augmentation	Site B Test-time data augmentation	Site B DSC	SEN	PRC	HD
DeepMedic [18]	None	None	14.9	11.6	45.3	42.6
	Heuristic [18]	None	46.4	43.2	59.4	26.9
	Heuristic† [18]	None	56.7	46.4	77.5	**9.4**
	Heuristic‡ [18]	None	**69.3**	**67.2**	73.5	**15.1**
	Learned‡ [7], [24], [27]	None	65.8	62.8	**75.1**	21.9
	Learned Class-Specific‡	None	**70.0** (+0.7)~	**68.0**	**75.9**	18.7
	Heuristic‡ [18]	Heuristic [16]	69.4	66.3	76.5	**8.0**
	Learned Class-Specific‡	Heuristic [16]	69.9	66.3	**80.0**	**8.0**
	Learned Class-Specific‡	Learned [20], [32]	**70.2**	67.7	**77.6**	15.3
	Joint Learned Class-Specific‡		**72.8** (+3.4)**	**71.0**	76.6	**7.9**

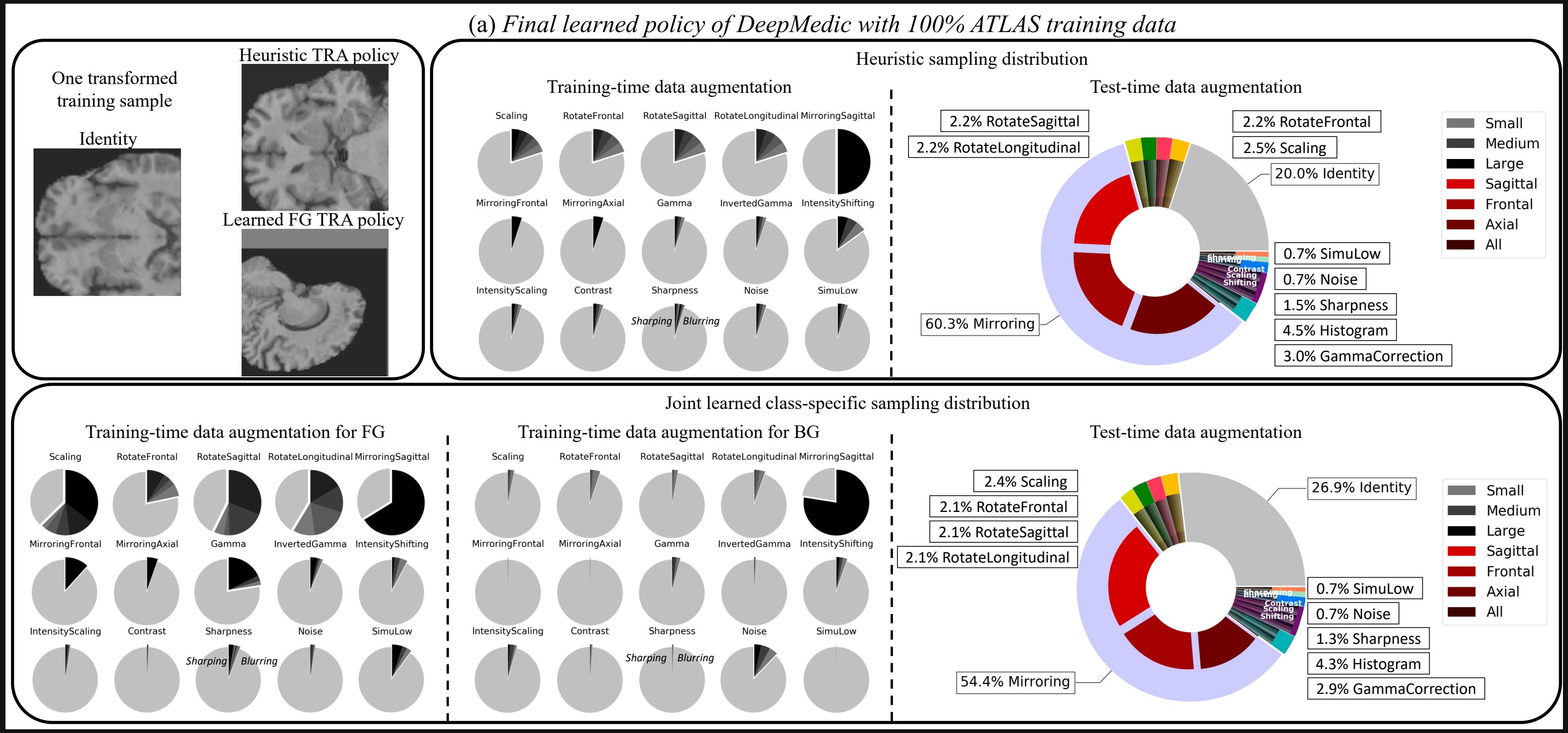
$^{*}p$-value < 0.05; $^{**}p$-value < 0.01; $^{\sim}p$-value ≥ 0.05 (compared to Heuristic‡ TRA w/o TEA or Heuristic‡ TRA w/ Heuristic TEA)
†We pretrain these models with training data from site A and fine-tune with validation data from site B.
‡We train these models with both training data from site A and validation data from site B.

Z. Li *et al.* "Joint optimization of class-specific training- and test-time data augmentation in segmentation", TMI, 2023.

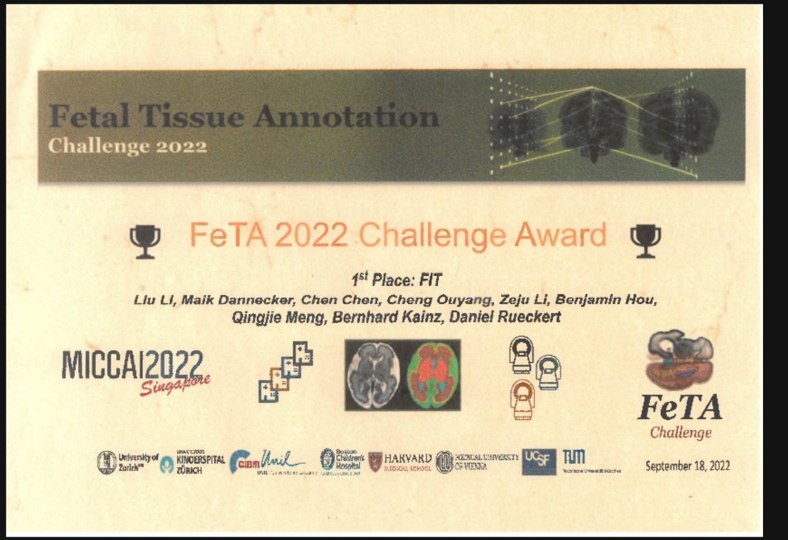
Joint Optimization of Class-Specific Training- and Test-time data augmentation

Results

- The learned policies would adopt larger transformations to the foreground than the background samples, **implicitly alleviating the class imbalance issue**.



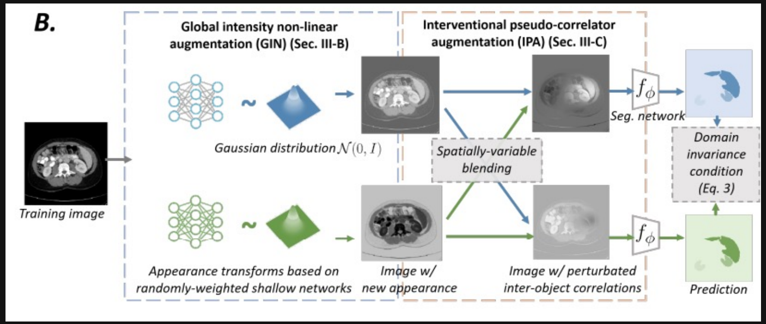
Serve as a key component for winning a robust learning challenge



Z. Li *et al.* "Joint optimization of class-specific training- and test-time data augmentation in segmentation", TMI, 2023.

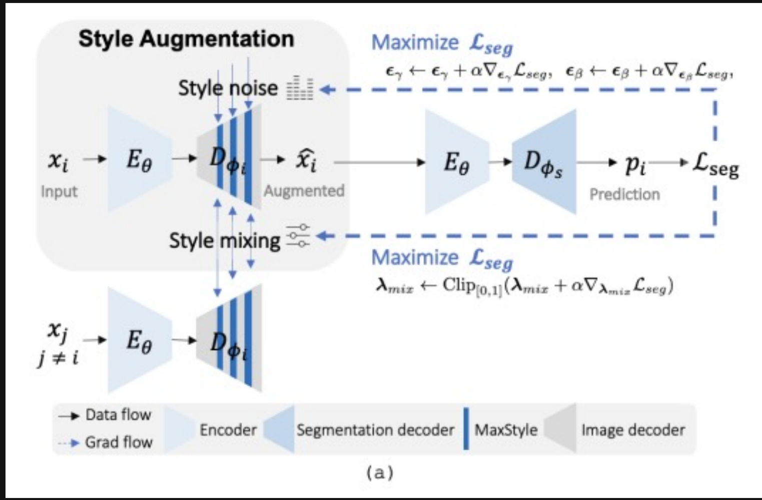
Other Related Works to Improve Robustness

Domain Generalization with Random Kernels



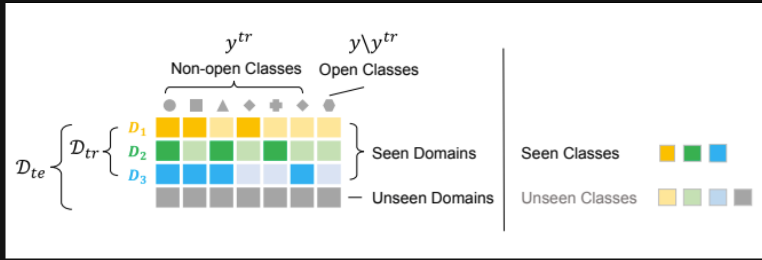
C. Ouyang, C. Chen, S. Li, **Z. Li** *et al.* "Causality-inspired single-source domain generalization for medical image segmentation", TMI, 2022.

Domain Generalization with Adversarial Training



C. Chen, **Z. Li** *et al.* "MaxStyle: Adversarial style composition for robust medical image segmentation", MICCAI, 2022.

Domain Generalization for Long-Tailed Classification



X. Gu, Y. Guo, **Z. Li** *et al.* "Tackling long-tailed category distribution under domain shifts", ECCV, 2022.

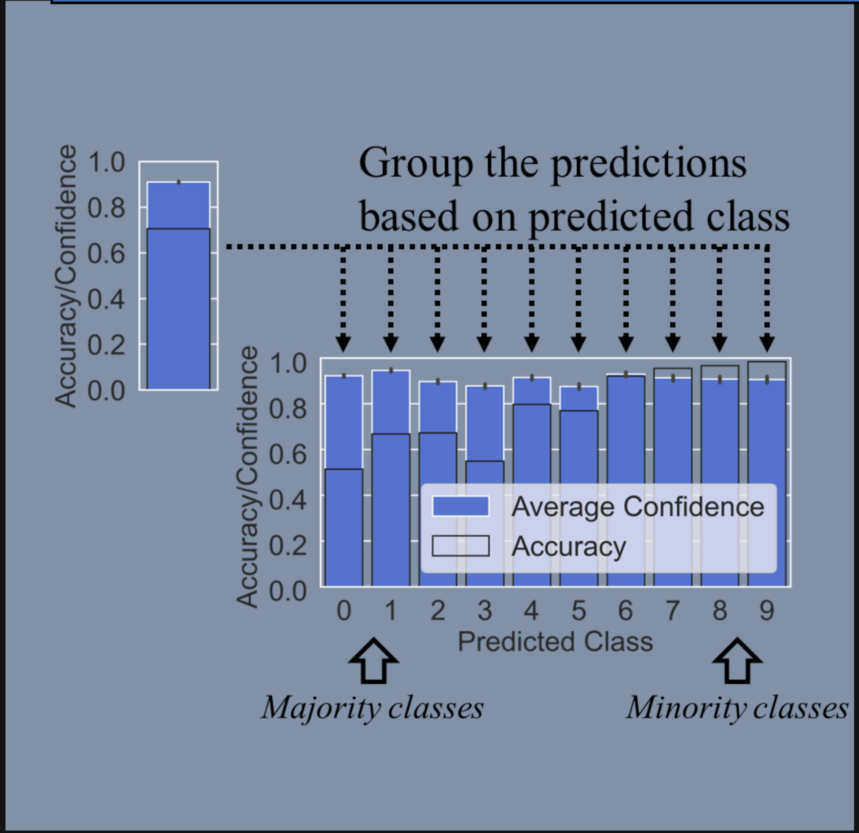
Performance Estimation under Domain Shifts

Confidence-based estimation

- Effect of **class imbalance** on confidence-based model evaluation methods.

Optimization with validation set $\mathcal{D}^{\mathrm{V}} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$

Deployment with test set $\mathcal{D}^{Te} = \{\boldsymbol{x}_i{}^{Te}\}_{i=1}^{M}$



Z. Li *et al.* "Estimating model performance under domain shifts with class-specific confidence scores", MICCAI, 2022.

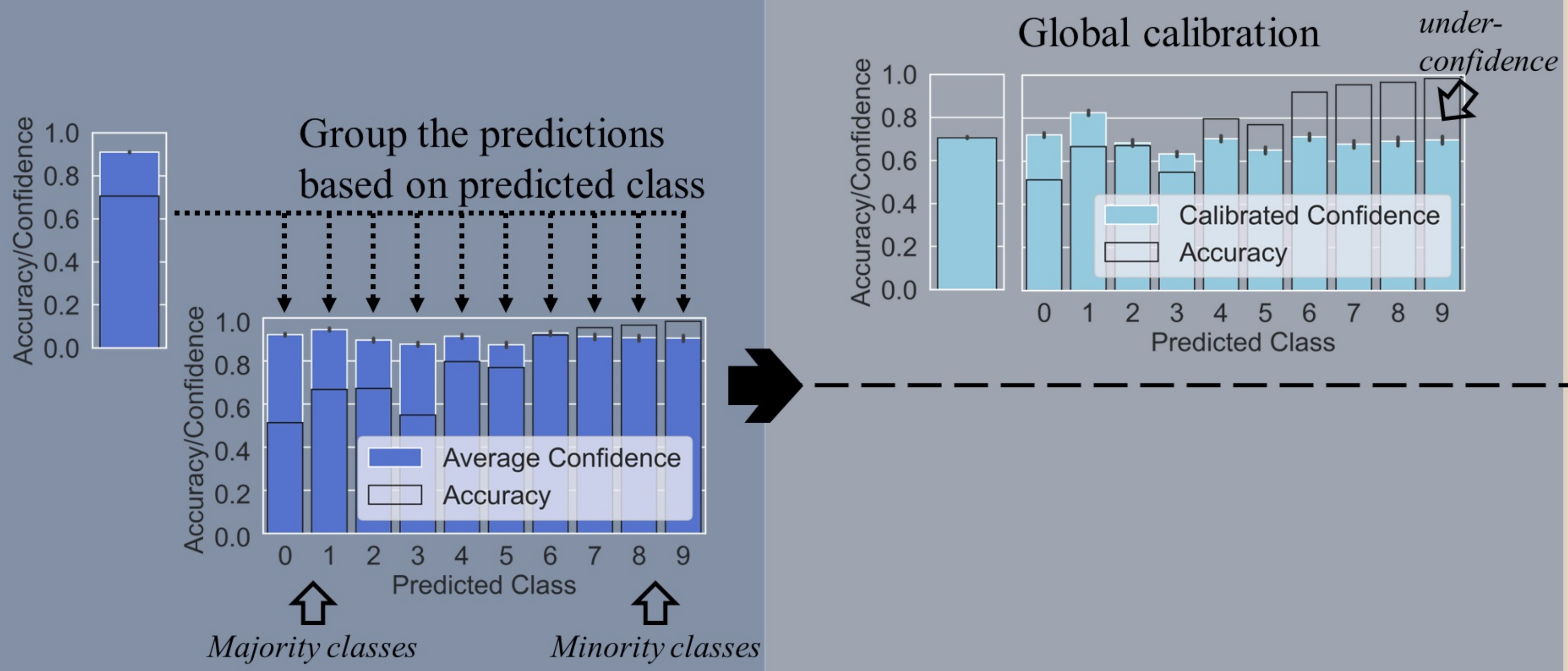
Performance Estimation under Domain Shifts

Confidence-based estimation

- Effect of **class imbalance** on confidence-based model evaluation methods.

Optimization with validation set $\mathcal{D}^{\mathrm{V}} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$

Deployment with test set $\mathcal{D}^{Te} = \{\boldsymbol{x}_i{}^{Te}\}_{i=1}^{M}$



Z. Li *et al.* "Estimating model performance under domain shifts with class-specific confidence scores", MICCAI, 2022.

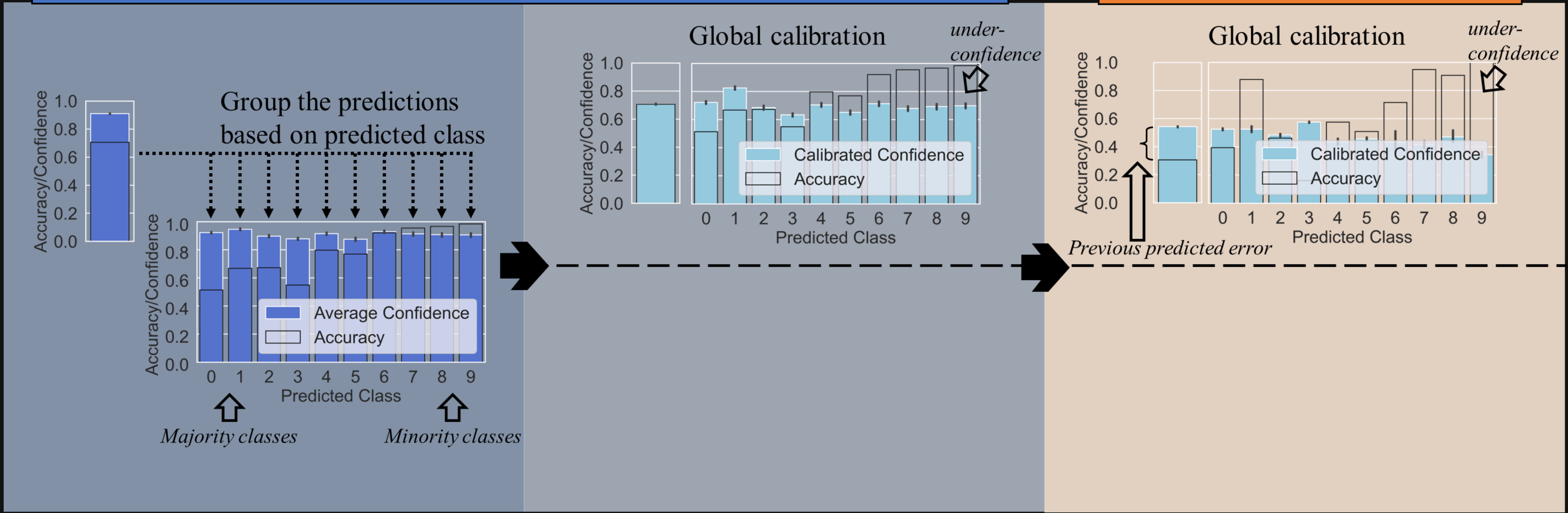
Performance Estimation under Domain Shifts

Confidence-based estimation

- Effect of class imbalance on confidence-based model evaluation methods.

Optimization with validation set $\mathcal{D}^{V} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$

Deployment with test set $\mathcal{D}^{Te} = \{\boldsymbol{x}_i{}^{Te}\}_{i=1}^{M}$



Z. Li *et al.* "Estimating model performance under domain shifts with class-specific confidence scores", MICCAI, 2022.

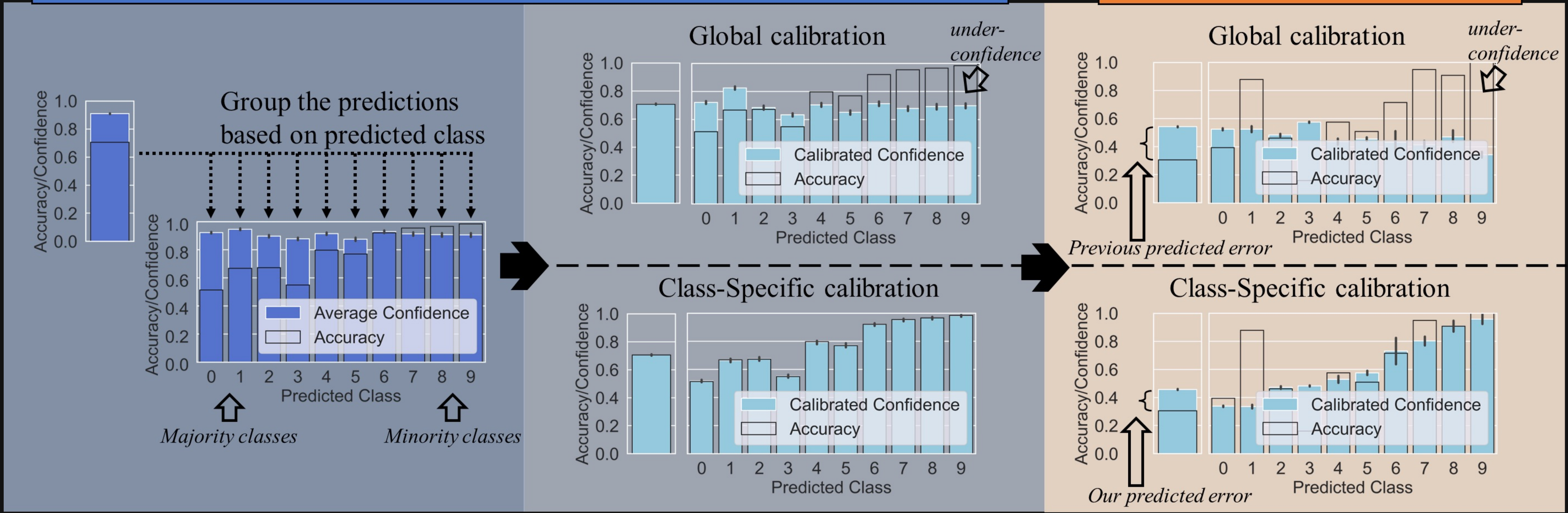
Performance Estimation under Domain Shifts

Confidence-based estimation

- Effect of class imbalance on confidence-based model evaluation methods.

Optimization with validation set $\mathcal{D}^{V} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$

Deployment with test set $\mathcal{D}^{Te} = \{\boldsymbol{x}_i^{Te}\}_{i=1}^{M}$

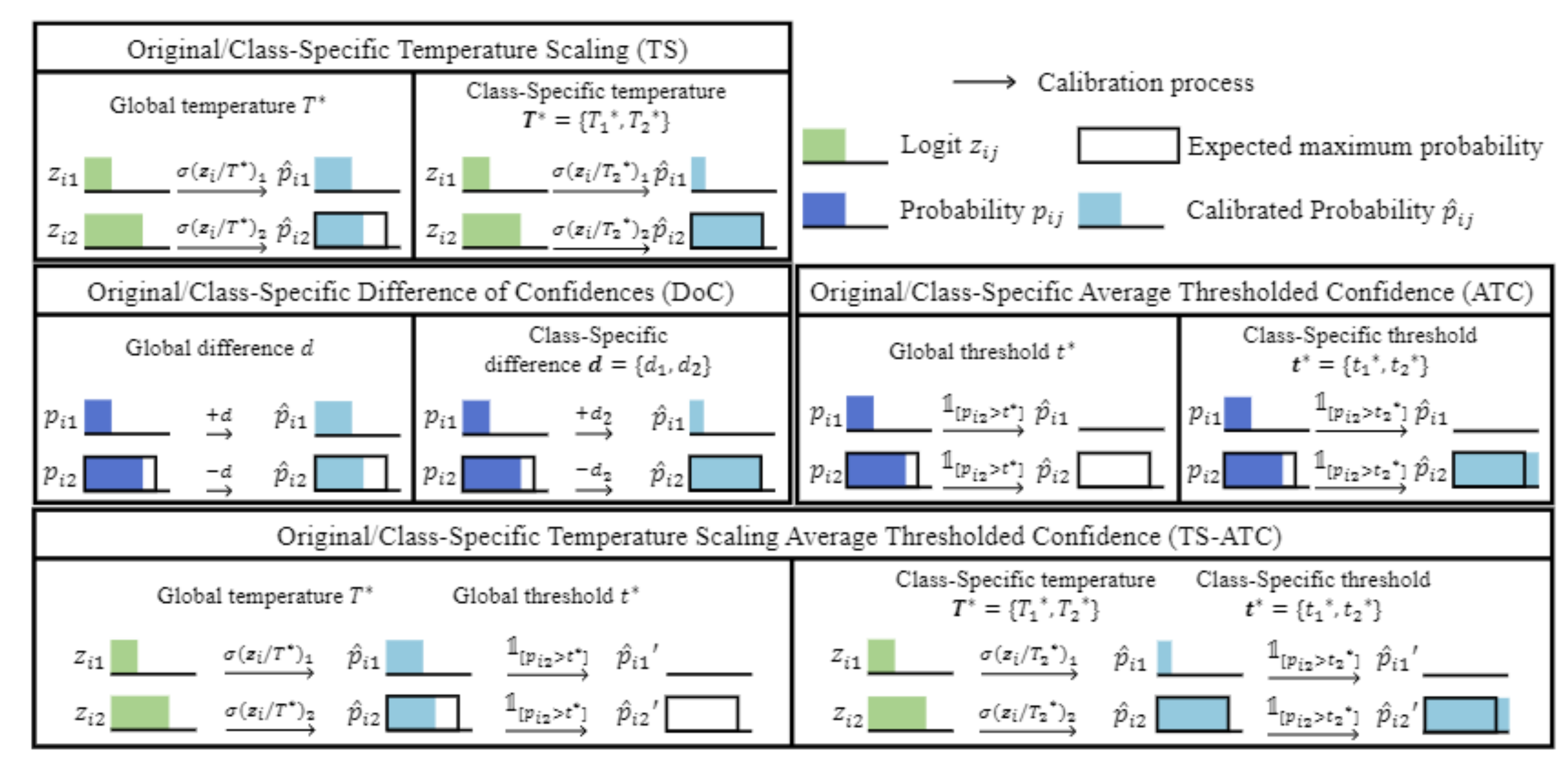


Z. Li *et al.* "Estimating model performance under domain shifts with class-specific confidence scores", MICCAI, 2022.

Performance Estimation under Domain Shifts

Method

➢ Introduce class-wise calibration within the framework of performance estimation for imbalanced datasets.



Z. Li *et al.* "Estimating model performance under domain shifts with class-specific confidence scores", MICCAI, 2022.

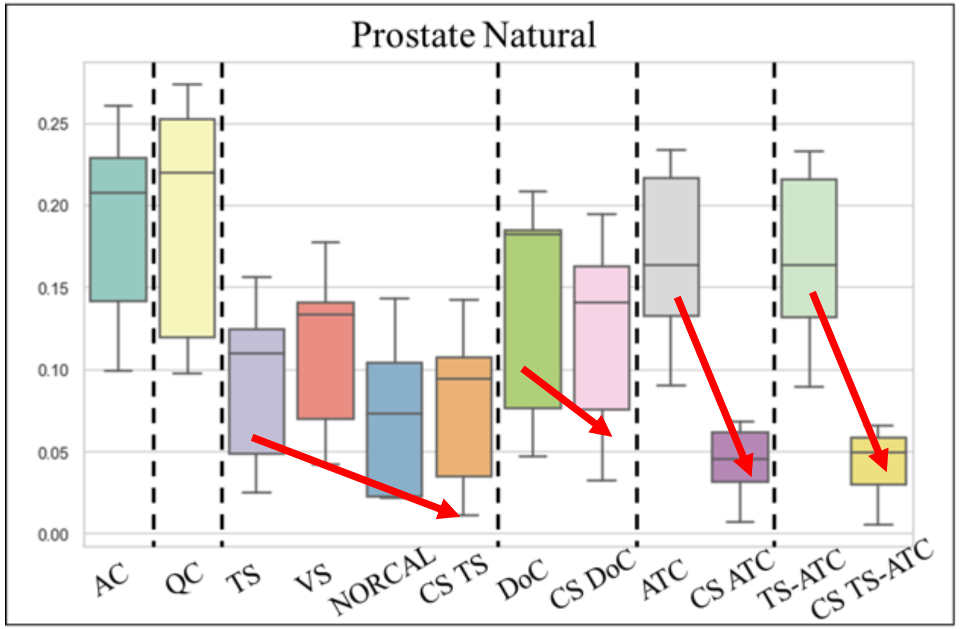
Performance Estimation under Domain Shifts

Results

➢ Consistently improve model estimation accuracy, especially for **segmentation tasks**.

Task	Classification			Segmentation		
Training dataset	CIFAR-10	HAM10000		ATLAS	Prostate	
Test domain shifts	Synthetic	Synthetic	Natural	Synthetic	Synthetic	Natural
AC	31.3 ± 8.2	12.3 ± 5.1	20.1 ± 13.4	35.6 ± 2.1	8.7 ± 4.9	18.7 ± 5.9
QC [30]	—	—	—	3.0 ± 1.7	5.2 ± 6.6	19.3 ± 7.1
TS [12]	5.7 ± 5.6	3.9 ± 4.3	12.1 ± 8.3	9.7 ± 2.5	3.7 ± 5.4	9.2 ± 4.9
VS [12]	3.8 ± 2.1	4.2 ± 4.2	13.6 ± 9.6	11.4 ± 2.5	4.8 ± 5.1	11.2 ± 4.9
NORCAL [29]	7.6 ± 3.8	4.2 ± 4.6	13.7 ± 9.6	6.7 ± 2.4	5.8 ± 5.7	7.3 ± 4.7
CS TS	$5.5^{\sim}$ ± 5.6	$3.7^{\sim}$ ± 4.0	$11.9^{\sim}$ ± 8.0	1.6^{**} ± 1.8	3.0^{**} ± 5.7	7.8^{**} ± 4.8
DoC [11]	10.8 ± 8.2	4.6 ± 5.0	15.3 ± 9.7	4.2 ± 3.2	3.7 ± 5.8	13.9 ± 6.5
CS DoC	9.4^{**} ± 7.2	$4.5^{\sim}$ ± 4.9	14.7^{*} ± 9.2	**1.3^{**} ± 1.9**	3.5^{*} ± 6.1	12.1^{*} ± 5.9
ATC [10]	4.6 ± 4.4	3.4 ± 3.9	7.1 ± 6.3	30.4 ± 1.8	8.6 ± 3.3	16.7 ± 5.3
CS ATC	2.8^{**} ± 2.9	**$3.3^{\sim}$ ± 4.8**	**$5.8^{\sim}$ ± 7.6**	1.6^{**} ± 1.5	**1.1^{**} ± 1.7**	4.3^{**} ± 2.2
TS-ATC [10,12]	5.3 ± 3.9	4.2 ± 4.2	7.3 ± 7.1	30.4 ± 1.8	8.5 ± 3.3	16.7 ± 5.3
CS TS-ATC	**2.7^{**} ± 2.3**	$4.2^{\sim}$ ± 5.4	$5.9^{\sim}$ ± 8.4	**1.3^{**} ± 1.4**	1.2^{**} ± 1.7	**4.2^{**} ± 2.2**

*p-value < 0.05; **p-value < 0.01; $^{\sim}$p-value ≥ 0.05 (compared with their class-agnostic counterparts)



Z. Li *et al.* “Estimating model performance under domain shifts with class-specific confidence scores“, MICCAI, 2022.

MOVAL

Software

- Based on our algorithm, we develop an open-source software, MOVAL, to help practitioners evaluate their model performance.

```
>>> import moval
>>> import numpy as np
>>> logits = np.random.randn(1000, 10)
>>> gt = np.random.randint(0, 10, (1000))
>>> moval_model = moval.MOVAL()
>>> moval_model.fit(logits, gt)
>>> estim_acc = moval_model.estimate(logits)
```

Reliable

MOVAL facilitates the assessment of pre-trained models across diverse scenarios, aiding practitioners in estimating performance.

Effective

MOVAL expands existing performance estimation algorithms and demonstrates state-of-the-art results, particularly on real-world long-tailed datasets.

Universal

MOVAL accommodates various applications within a unified framework, encompassing tasks such as classification, 2D segmentation, and 3D segmentation.

Versatile

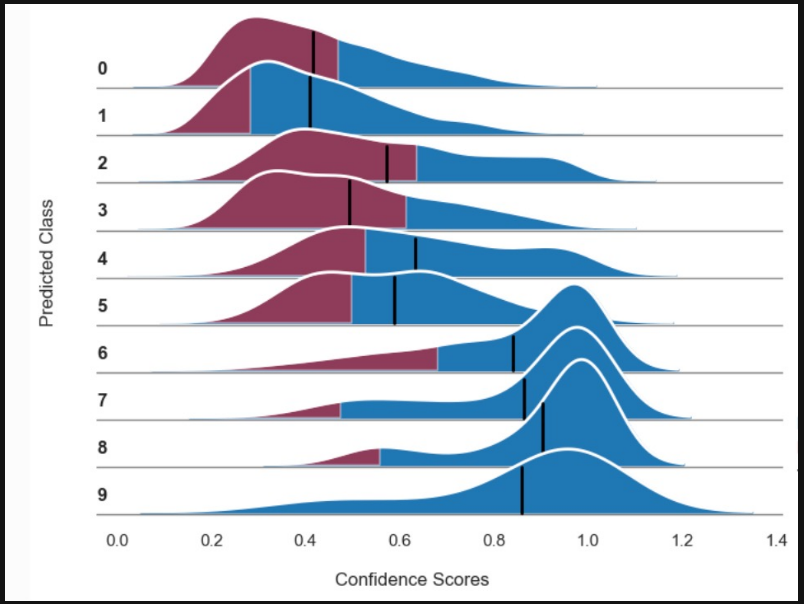
MOVAL not only calculates but also calibrates various confidence scores beyond the maximum class probability.

User-Friendly

MOVAL can be effortlessly installed as a Python module and supports the NumPy array data format.

Modular

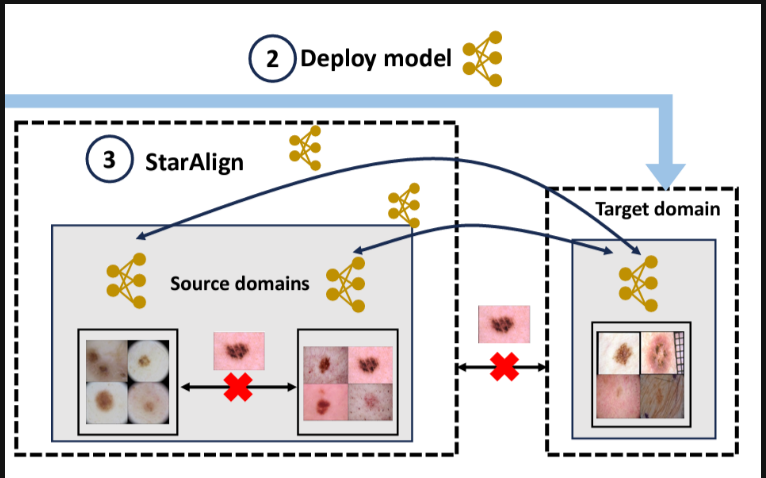
MOVAL comprises distinct modules for confidence score calculation, calibration, and optimization, allowing for easy extension and customization.



https://github.com/ZerojumpLine/MOVAL/

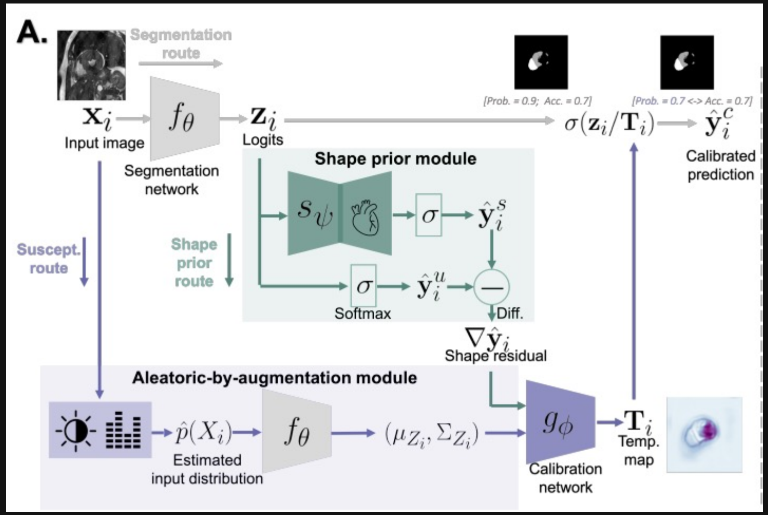
Other Related Works to Improve Reassurance

Post-deployment Adaption under Federated Setting



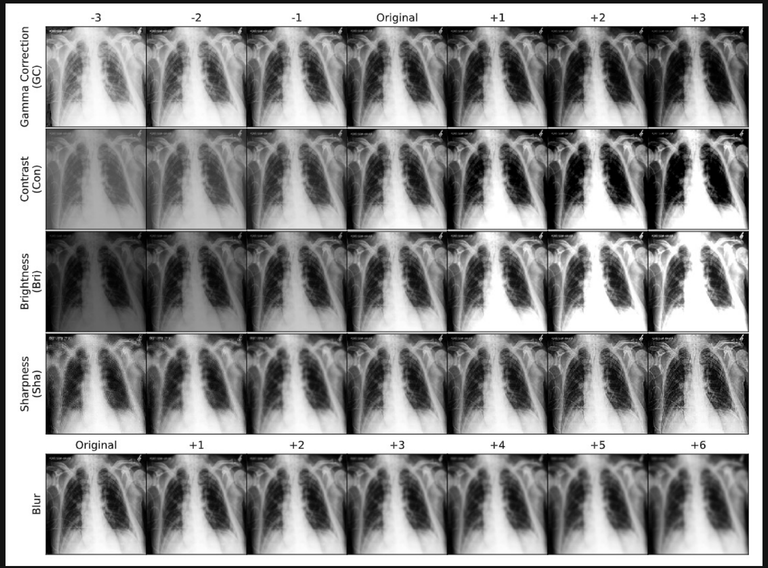
F. Wagner, **Z. Li** *et al*. "Post-deployment adaptation with access to source data via federated learning and source-target remote gradient alignment", MICCAI-MLMI workshop, 2023.

Model Calibration for Image Segmentation



C. Ouyang, S. Wang, C. Chen, **Z. Li** *et al.* "Improved post-hoc probability calibration for out-of-domain MRI segmentation", MICCAI-UNSURE workshop, 2022.

Stress Testing with Fairness Concerns

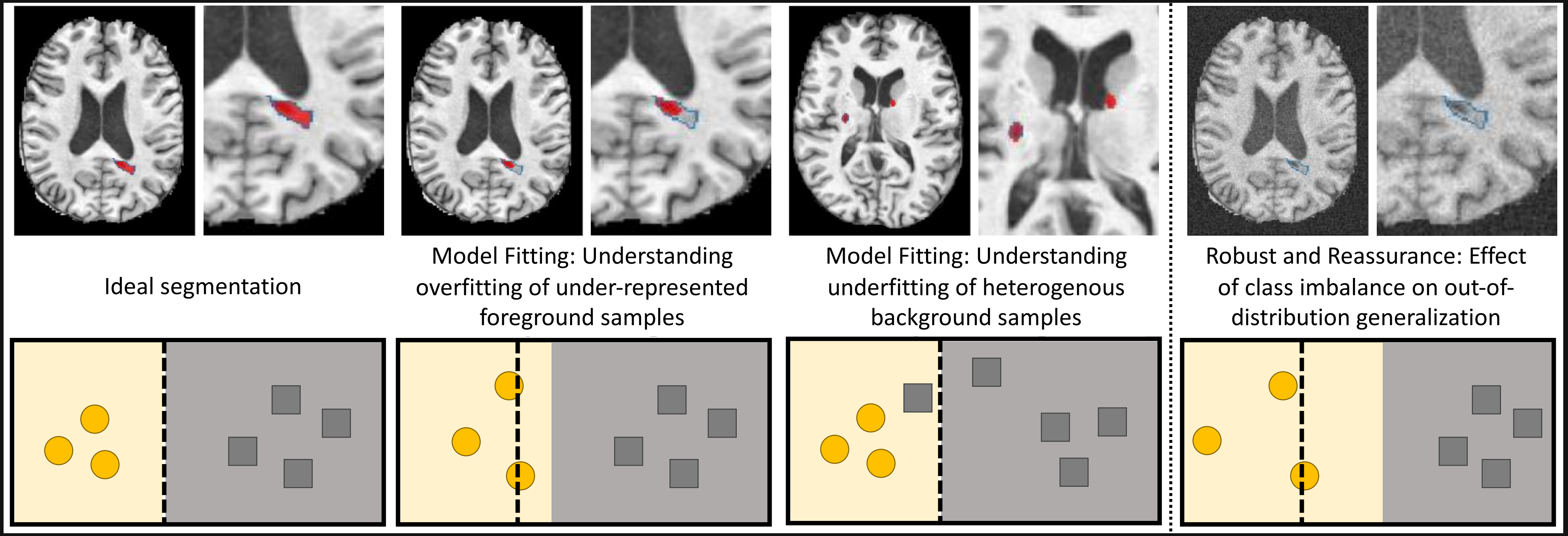


M. Islam, **Z. Li** *et al.* "Progressive stress testing of model robustness in medical image classification", MICCAI-UNSURE workshop, 2023.

Conclusion

Summary

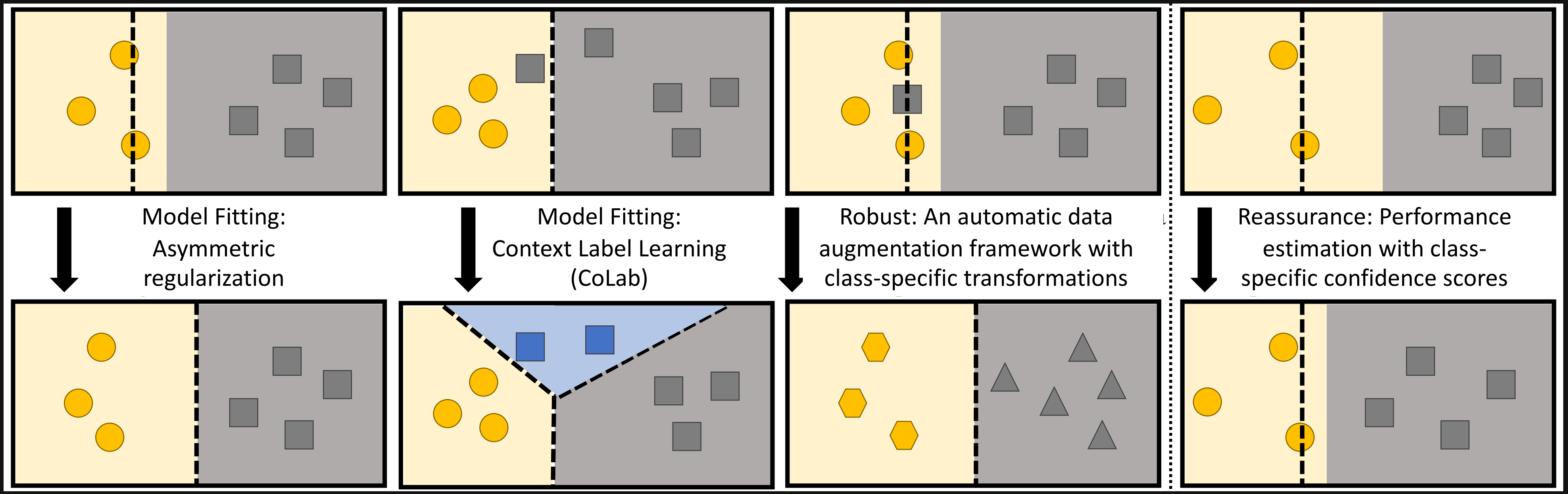
- **Class imbalance** and **domain shifts**, which exist simultaneously in real world datasets, limit the performance of modern machine learning models when deployed to medical image segmentation.
- Class imbalance cause under-segmentation because of **overfitting foreground samples**, while over-segmentation because of **underfitting background samples**.



Conclusion

Summary

- ➢ **Asymmetric** loss functions and regularization techniques help counter overfitting under class imbalance by enlarging foreground sample variances.
- ➢ **Context labels** help alleviate underfitting under class imbalance.
- ➢ **Class-specific parameters** are beneficial for improving data augmentation and tackling domain shifts.



Take Home Message

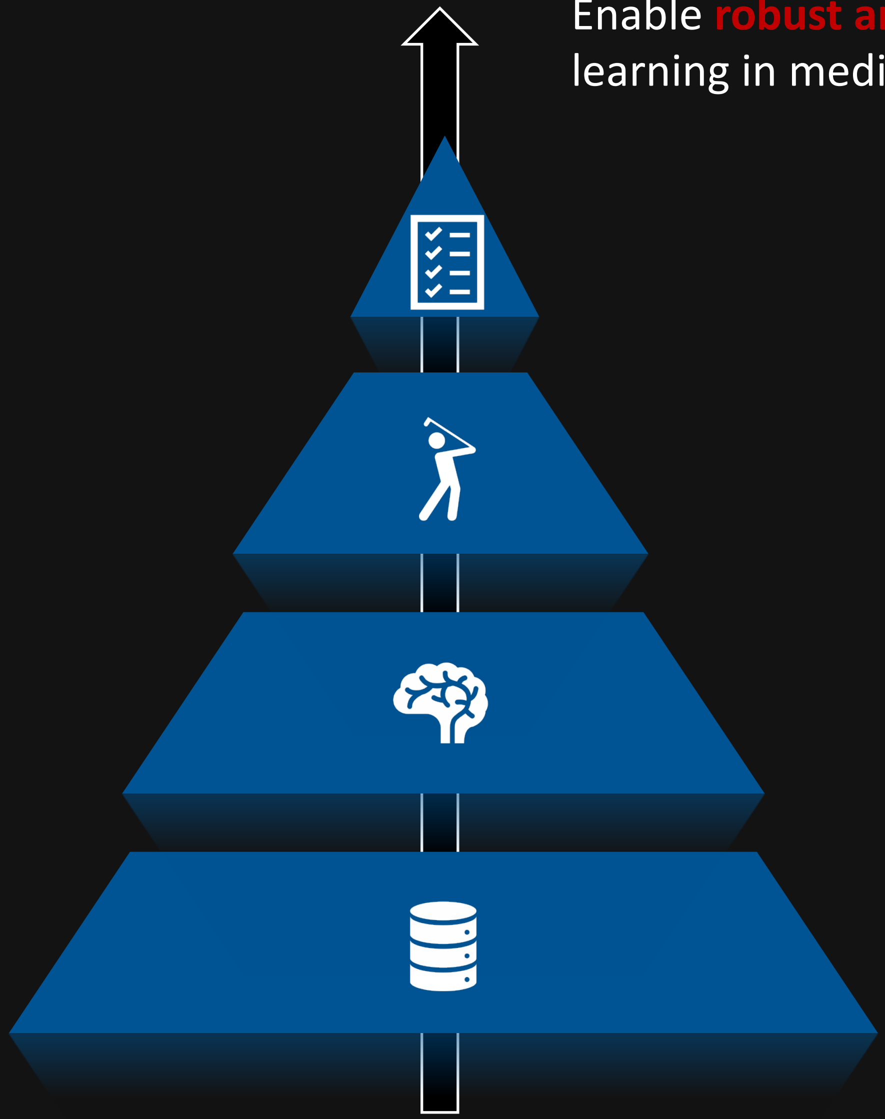
Enable robust and safe deployment of machine learning in medical image segmentation!

Model Performance

Deployment Environment

Optimization Algorithms

Data Distributions



Reassurance: Calibrating model prediction with to match class-wise performance.

Robustness: Aligning the training and test data distribution with data augmentation.

Model-Fitting: Alleviating class imbalance with advanced learning strategies.

Reference

- Model-Fitting:

1. **Z. Li** *et al.* "Analyzing overfitting under class imbalance in neural networks for image segmentation", TMI, 2020.
2. **Z. Li** *et al.* "Context label learning: improving background class representations in semantic segmentation", TMI, 2023.
3. **Z. Li** *et al.* "Deep learning based radiomics (DLR) and its usage in noninvasive idh1 prediction for low grade glioma", Sci. Rep., 2017.
4. **Z. Li** *et al.* "Deepvolume: brain structure and spatial connection-aware network for brain MRI super-resolution", TCybern, 2019.
5. **Z. Li** *et* al. "Brain tumor segmentation using an adversarial network.", MICCAI-brainlesion workshop, 2017.

- Robustness:

1. **Z. Li** *et al.* "Joint optimization of class-specific training- and test-time data augmentation in segmentation", TMI, 2023.
2. C. Ouyang, C. Chen, S. Li, **Z. Li** *et al.* "Causality-inspired single-source domain generalization for medical image segmentation", TMI, 2022.
3. C. Chen, **Z. Li** *et al.* "MaxStyle: Adversarial style composition for robust medical image segmentation", MICCAI, 2022.
4. X. Gu, Y. Guo, **Z. Li** *et al.* "Tackling long-tailed category distribution under domain shifts", ECCV, 2022.
5. C. Chen, C. Ouyang, **Z. Li** *et al.*"Enhancing mr image segmentation with realistic adversarial data augmentation", MedIA, 2022

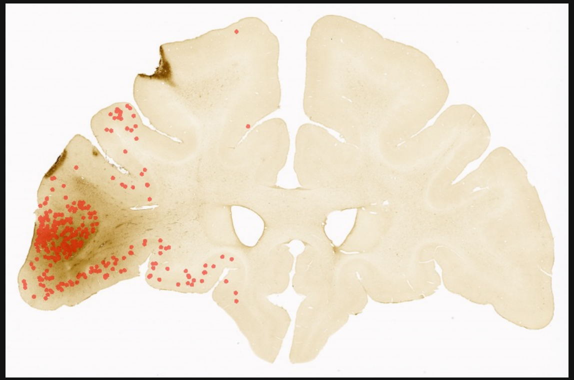
- Reassurance:

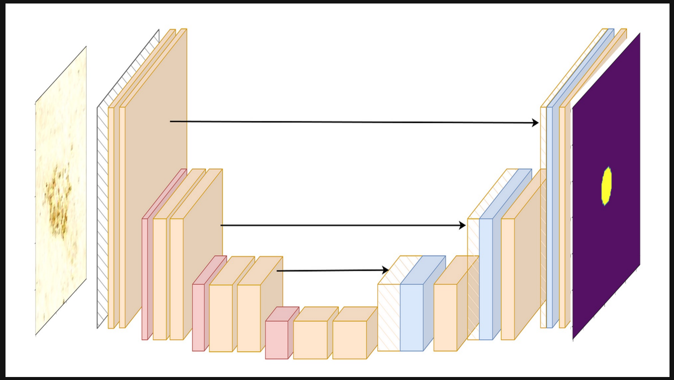
1. **Z. Li** *et al.* "Estimating model performance under domain shifts with class-specific confidence scores", MICCAI, 2022.
2. F. Wagner, **Z. Li** *et al.* "Post-deployment adaptation with access to source data via federated learning and source-target remote gradient alignment", MICCAI-MLMI workshop, 2023.
3. C. Ouyang, S. Wang, C. Chen, **Z. Li** *et al.* "Improved post-hoc probability calibration for out-of-domain MRI segmentation", MICCAI-UNSURE workshop, 2022.
4. M. Islam, **Z. Li** *et al.* "Progressive stress testing of model robustness in medical image classification", MICCAI-UNSURE workshop, 2023.
5. **Z. Li** *et al.* "Encoding ct anatomy knowledge for unpaired chest x-ray image decomposition", MICCAI, 2019.

Post-doc Working on Building Macaque Connectivity Atlas

Cell Segmentation

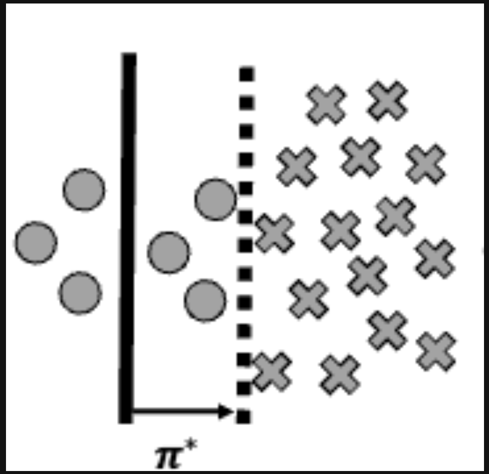
End-to-end neural networks such as U-Net / Transformer.





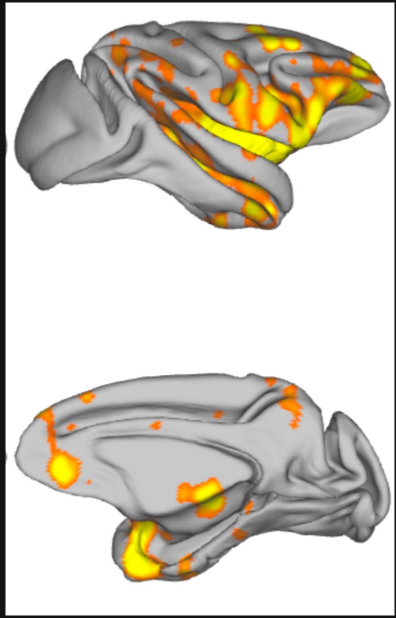
Semi-supervised Learning

Propose an algorithm that enhances the quality of pseudo-labelling for imbalanced semi-supervised learning.

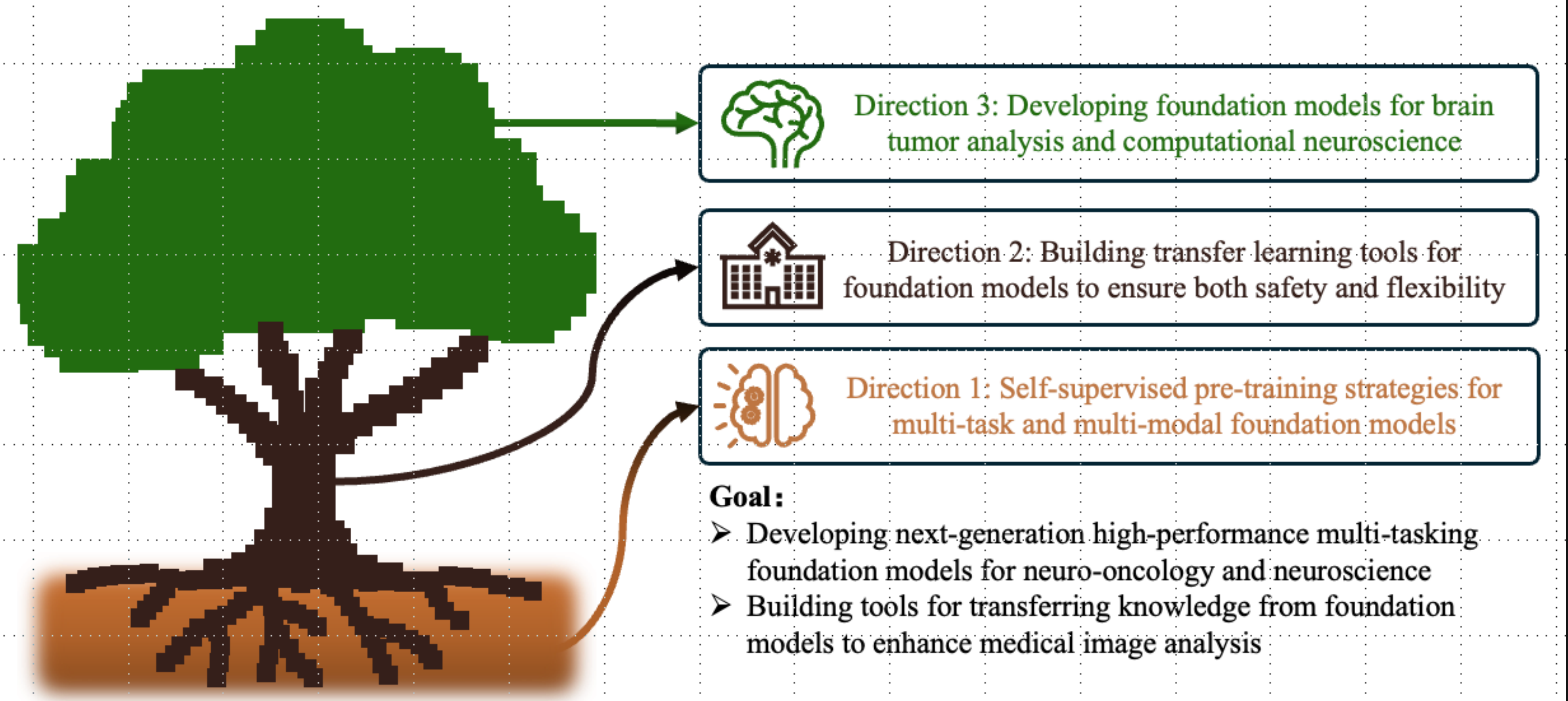


Registration to structure MRI

The first tracer atlas that is both fine-grained and quantitative



Future Vision: Towards Building Foundation Models for Neuro-oncology and Neuroscience



Paper & code

Thank you!

Zeju Li, PhD
FMRIB Center, Nuffield Department of Clinical Neurosciences,
University of Oxford, UK.
Email: zeju.li@ndcn.ox.ac.uk
March 2024

Statistics
3,288 citations
12 first authored peer-reviewed journal/conference

Academic Service
Area Chair, MICCAI 2024
Reviewer, CVPR/MICCAI, TMI/MedIA