Imperial College London

BioMedIA

Improving the Generalization Capability for Medical Image Segmentation

Zeju Li
BioMedIA Group, Department of Computing, Imperial College London

May 25th 2020

Overview

Introduction

- Short Bio
- Brief Intro of Medical Image Segmentation
- Previous Researches
- Open Problems

Improving Generalization Capability

- Generalization in Deep Learning
- Overfitting under Class Imbalance
- Automatic Data Augmentation

Conclusion

BioMedIA

Short Bio

Education

- 2018.10 – Current
  - PhD in Computing, Imperial College London, London
- 2015.9 – 2018.7
  - Master in Biomedical Engineering, Fudan, Shanghai
- 2011.9 – 2015.7
  - Bachelor in Electronic Engineering, Fudan, Shanghai

Intern

- 2019.7 – 2020.4
  - Huawei Noah's Ark Lab, London
- 2018.7 – 2018.9
  - **MIRACLE**, ICT, Beijing

Imperial College London

FUDAN UNIVERSITY 1905

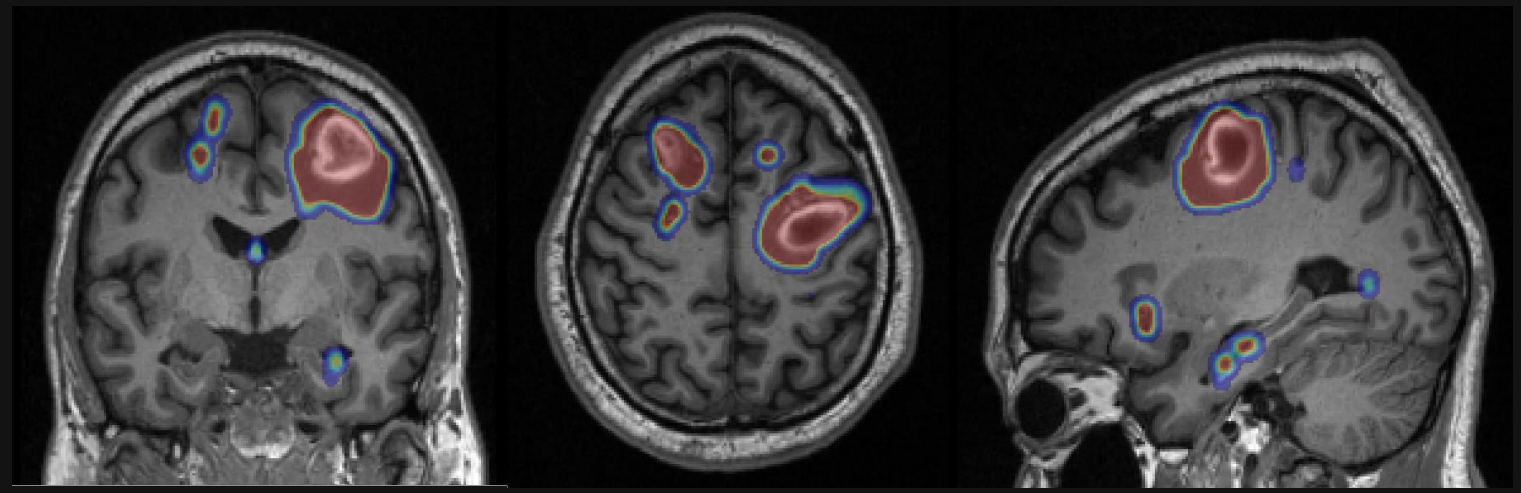
HUAWEI

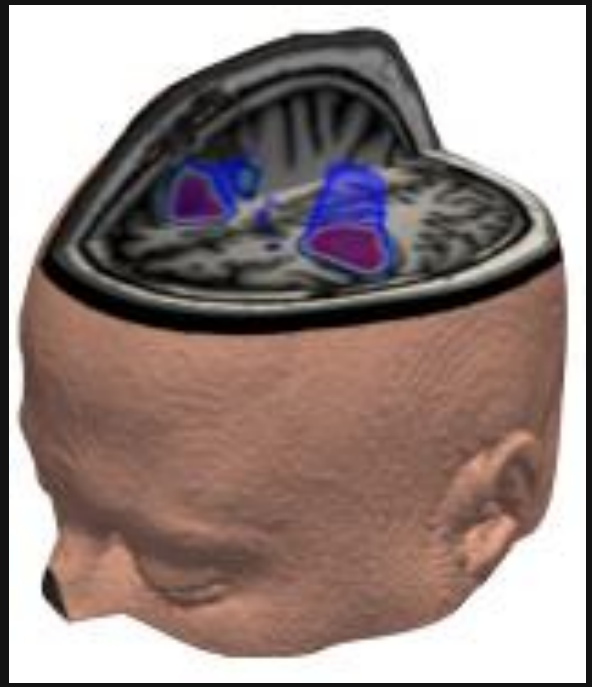
ICT

Brief Intro of Medical Image Segmentation

Goal

- Identify groups of pixels that go together.
- **Accuracy and efficiency**.





https://biomedia.doc.ic.ac.uk/software/deepmedic/

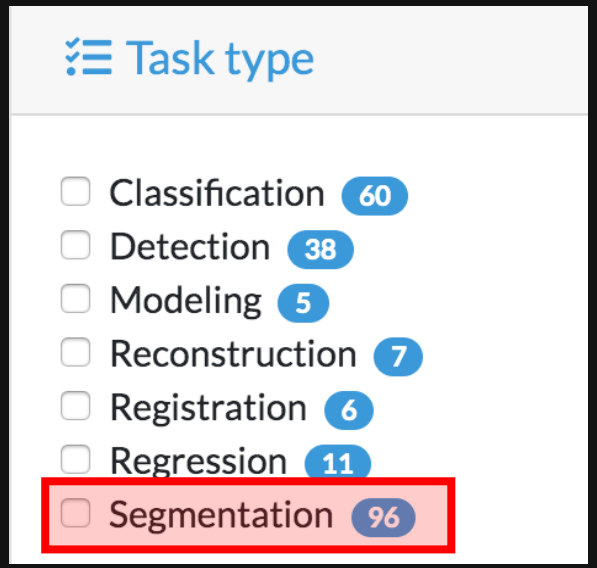
Brief Intro of Medical Image Segmentation

Importance

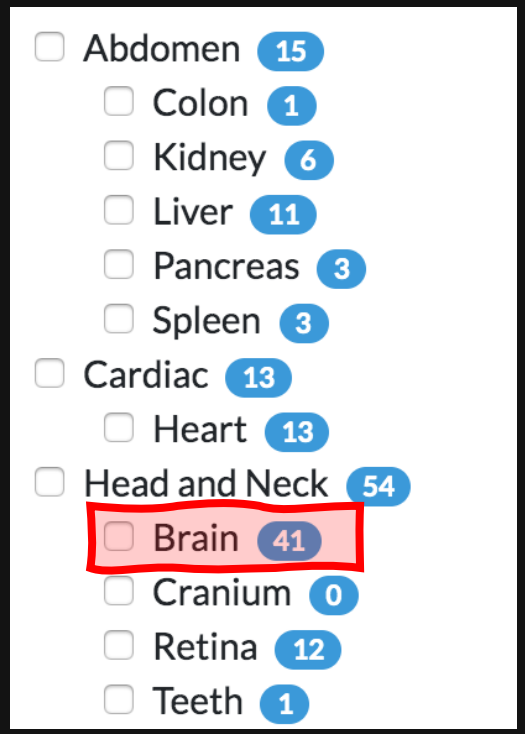
- The most popular medical imaging task.
- **40% of MICCAI papers** are about Medical Image Segmentation.

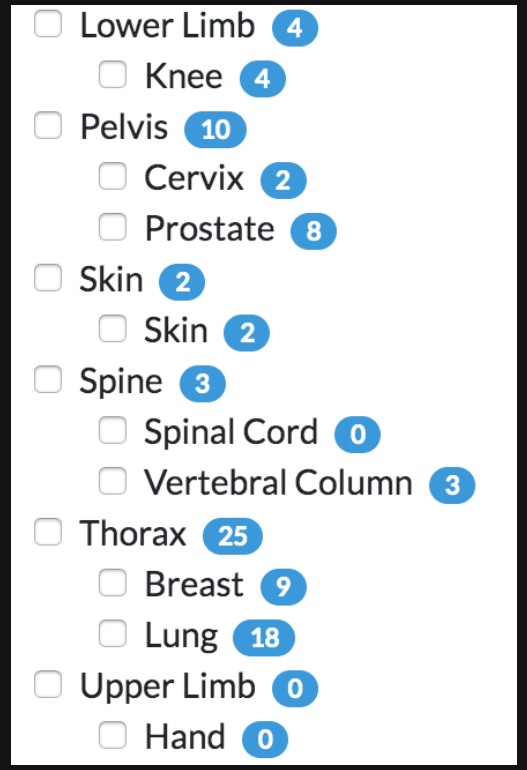
Applications

- Reduce tedious annotations.
- A prerequisite for following-up tasks.
- **MRI**/ CT/ Xray/ Ultrasound/ Histology/ Fundus Photography.



https://grand-challenge.org/challenges/





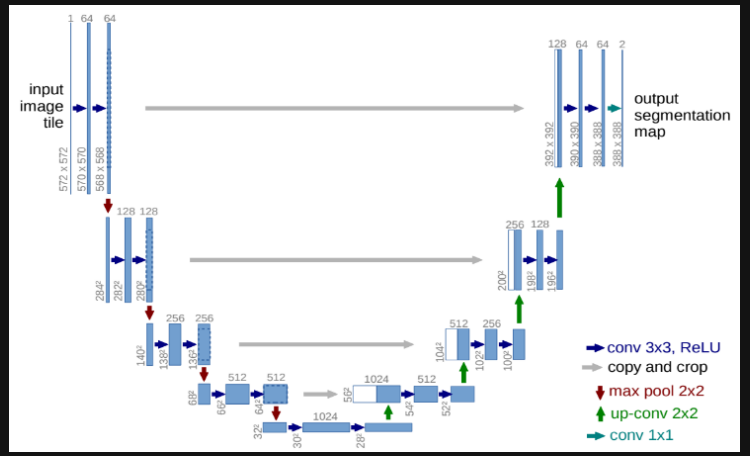
Brief Intro of Medical Image Segmentation

Methodologies

- Deformable models based segmentation [T. McInerney and D. Terzopoulos D. MedIA 1996]
  - Snake [C. Xu, J. Prince TIP 1998]
  - Level-set [C. Li et. al. CVPR 2005]
- Statistical inference based segmentation [D. Pham, C. Xu, and J. Prince AnnuRev Biomed Eng 2000]
  - Markov random field [Y. Zhang, M. Brady, S. Smith TMI 2001]
  - Graph cut [Y. Boykov, O. Veksler, R. Zabih TPAMI 2001]
- Registration based segmentation [J. Iglesias and M. Sabuncu MedIA 2015]
  - Multi-atlas [P. Aljabar Neuroimage 2009]
- Discriminative classifier based segmentation

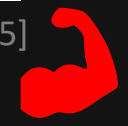
$$\begin{cases} Z = F_\theta(X) \\ Y = G_\varphi(Z) \end{cases} \qquad Y = G_\varphi\big(F_\theta(X)\big)$$

  - Random forests [L. Breiman Machine learning 2001]
  - Sparse representation [J. Wright TPAMI 2008]
  - **Deep learning** [O. Ronneberger, P. Fischer, T. Brox MICCAI 2015]



[O. Ronneberger, P. Fischer, T. Brox MICCAI 2015]

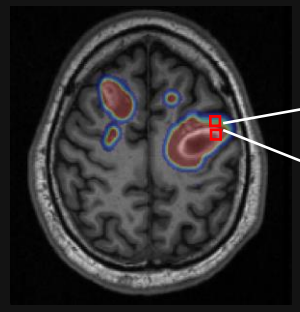
Most cited paper in the MICCAI history



Previous Researches: Part one

Segmentation coherence: problem

- The pixel segmentation of neural network is independent of another pixel.

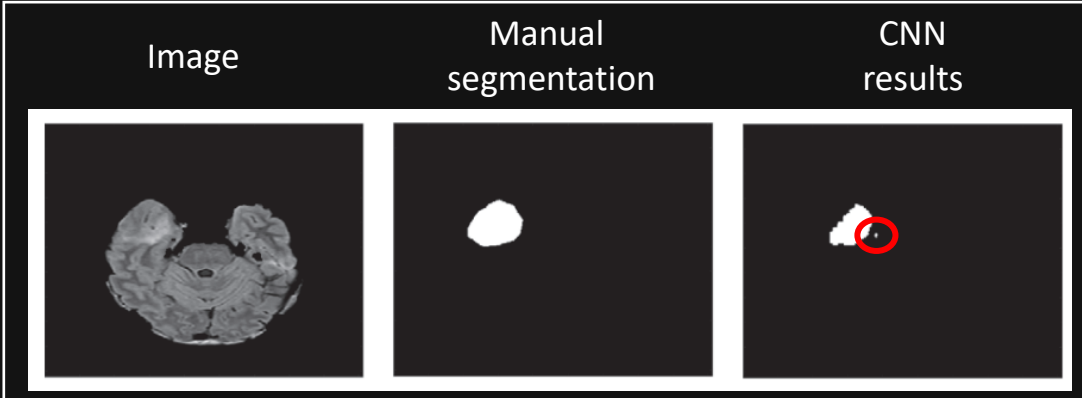


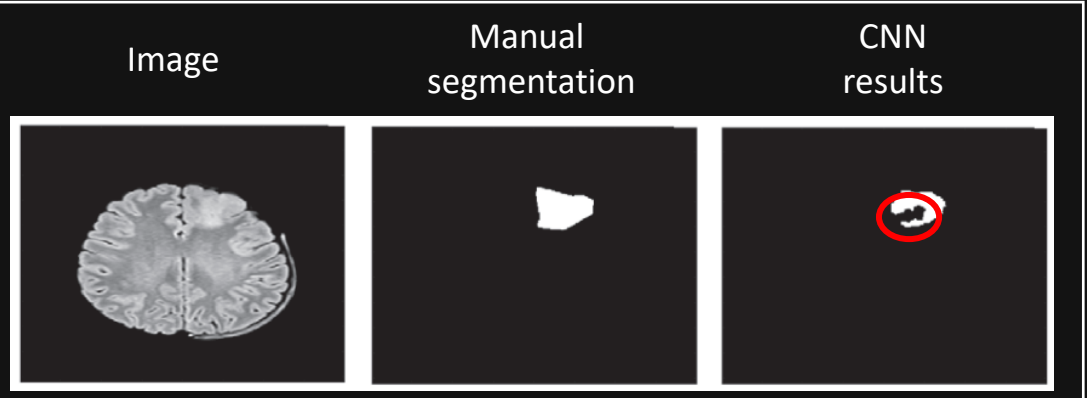
$$y_1 = G_\varphi(F_\theta(x_1))$$

$$y_2 = G_\varphi(F_\theta(x_2))$$

Do not have a strong constrain to maintain similar class

- The segmentation results **lack space continuity**/ shape prior.



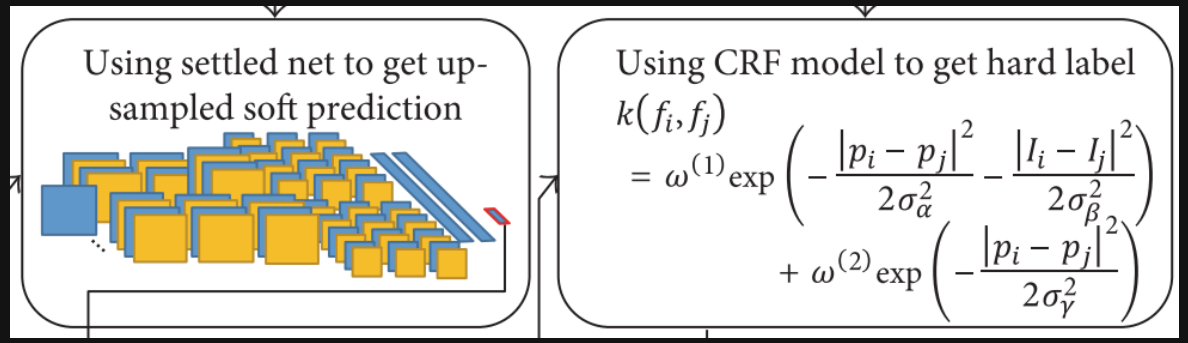


[**Z. Li**, et al. J Healthc Eng 2017]

Previous Researches: Part one

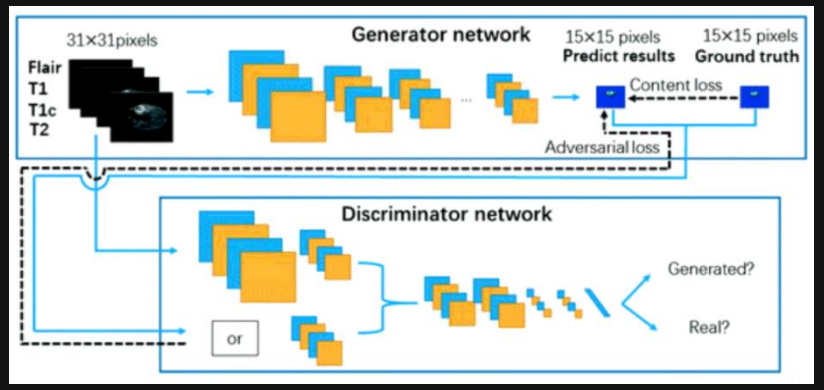
Segmentation coherence: methods

- Integrate **conditional random field** (CRF)



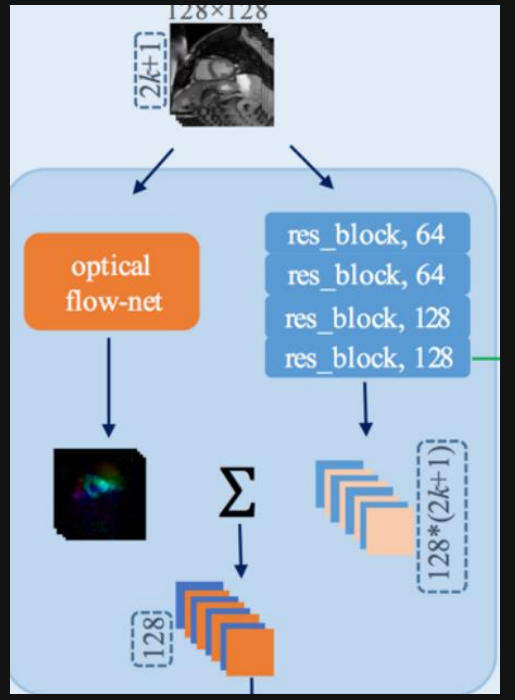
[**Z. Li**, et al. J Healthc Eng 2017]

- Adapt **adversarial training** to encourage high-order consistency



[**Z. Li**, et al. MICCAI-BrainLesion 2017]

- Add **optical-flow** for high dimension data

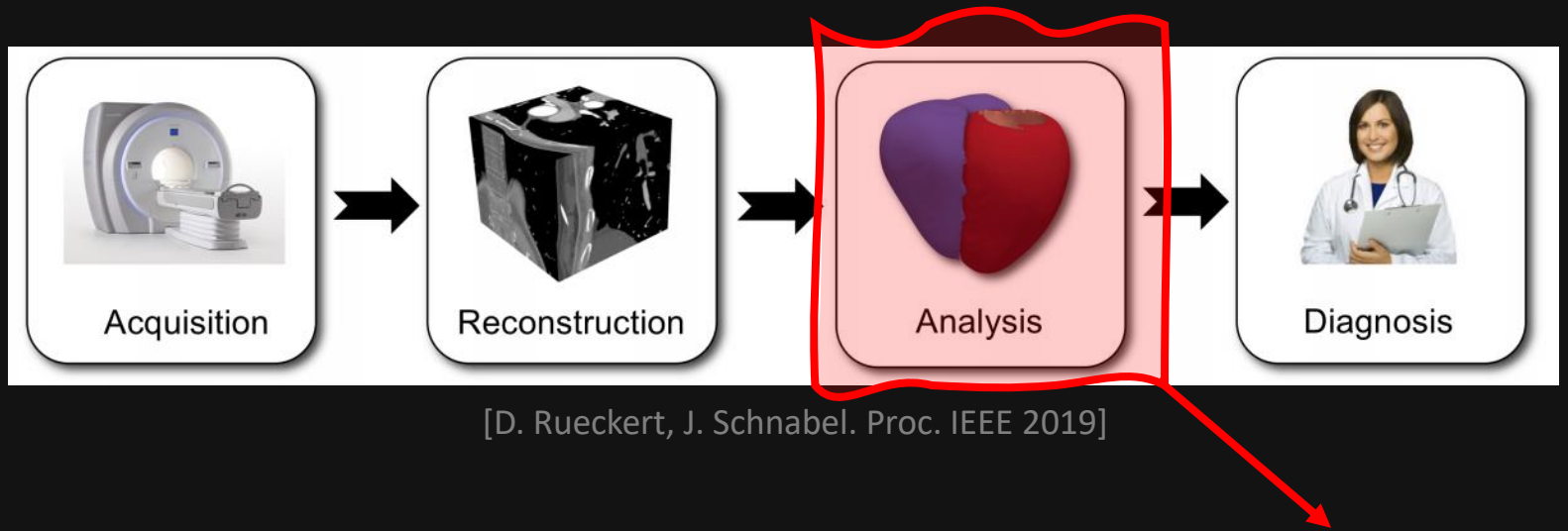


[W. Yan, Y. Wang, **Z. Li**, et al. MICCAI 2018]

Previous Researches: Part two

Integrated segmentation: problem

- Medical imaging pipeline consists of **many processes**, which are optimized independently.
- There could be some **errors** and **information loss** among different processes.

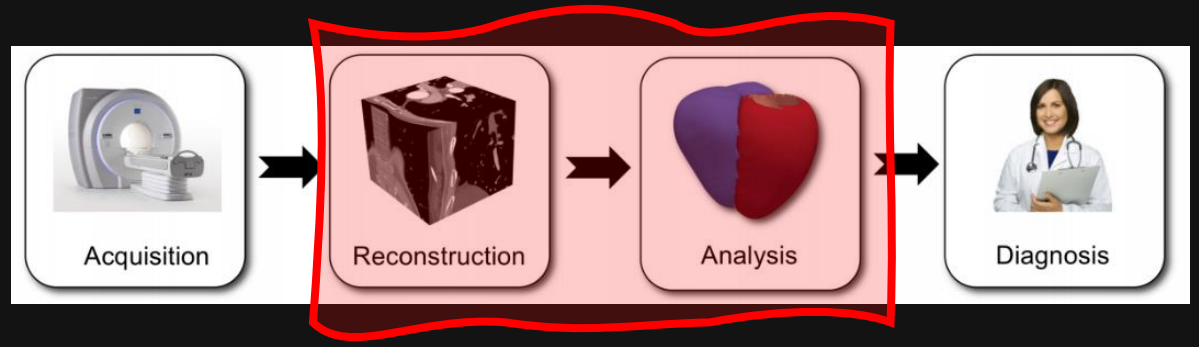


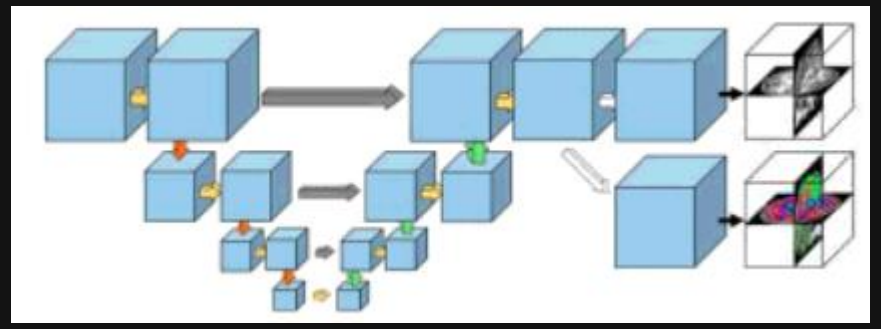
[D. Rueckert, J. Schnabel. Proc. IEEE 2019]

Previous Researches: Part two

Integrated segmentation: methods

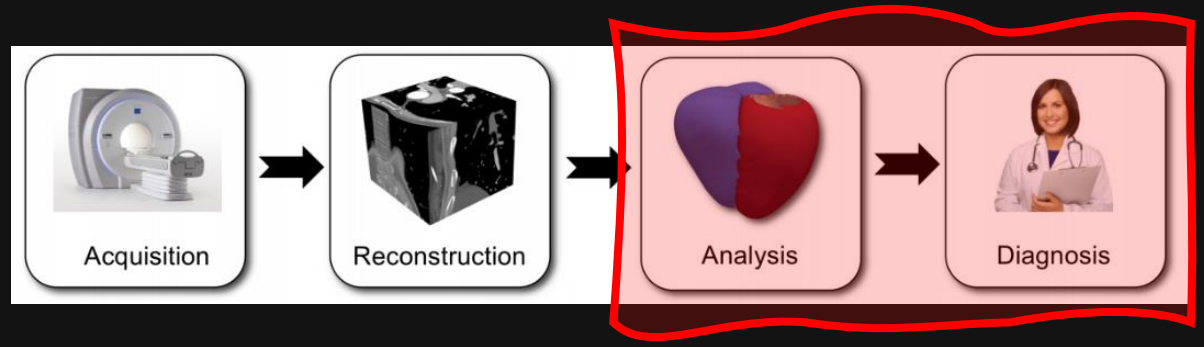
- Combine the process of **reconstruction and segmentation**

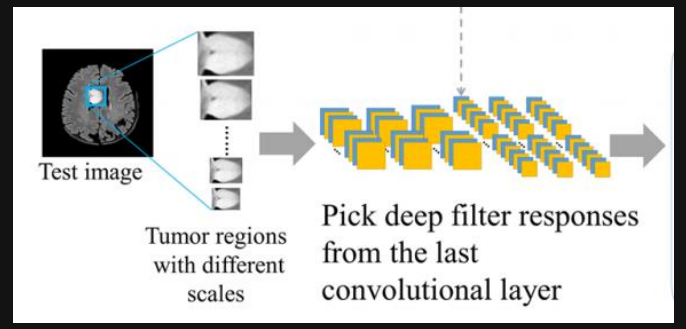




[**Z. Li**, et al. TCybern 2019]

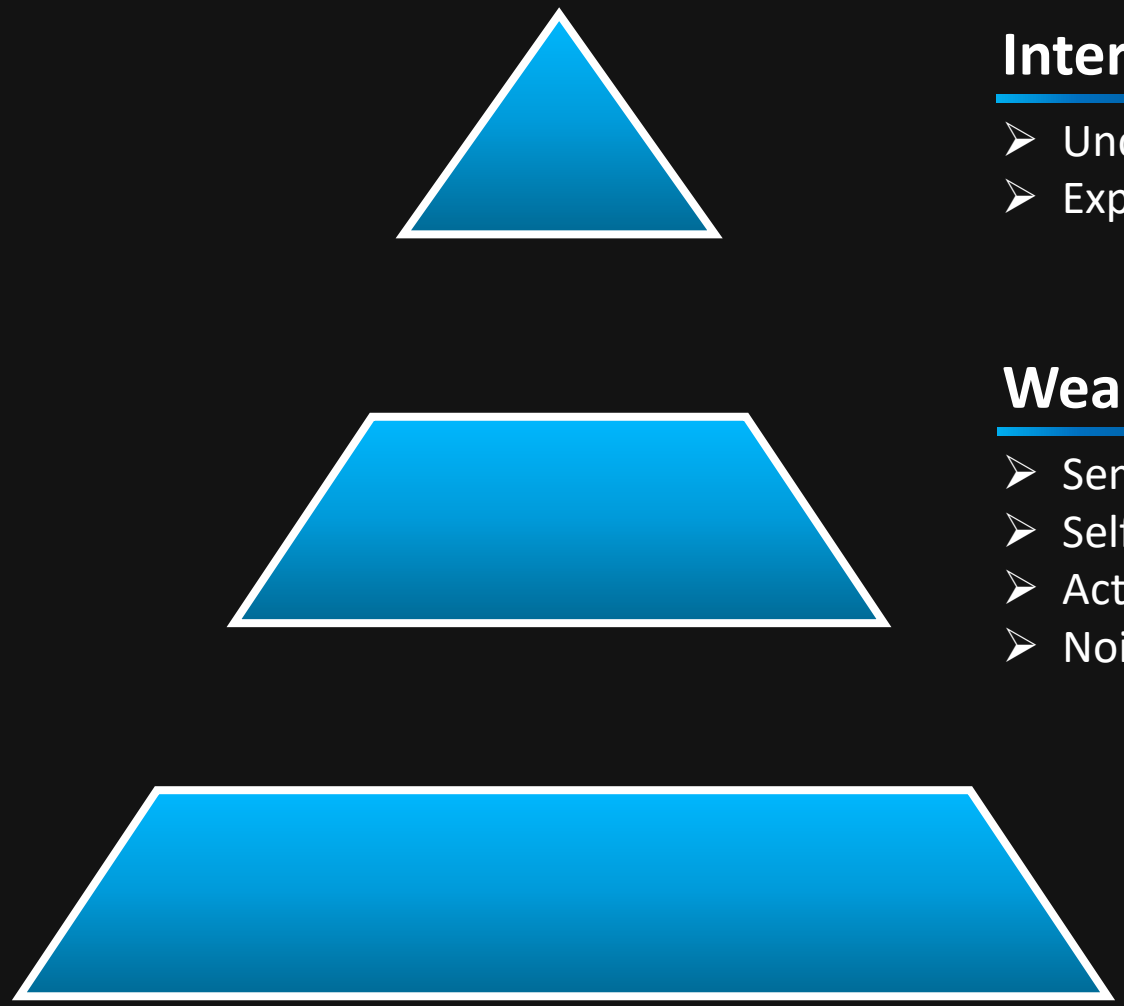
- Combine the process of **segmentation and diagnosis**





[**Z. Li**, et al. Sci Rep 2017]

Open Problems



Interpretability

- Uncertainty estimation
- Explainable deep learning

Clinical Relevance

- Large-scale validation
- Beyond imaging data

Weak Label Usage

- Semi-supervised/ unsupervised learning
- Self-supervised learning
- Active learning
- Noise-label learning

Robust Learning

- Domain adaptation
- Domain shifts
- Data heterogeneity

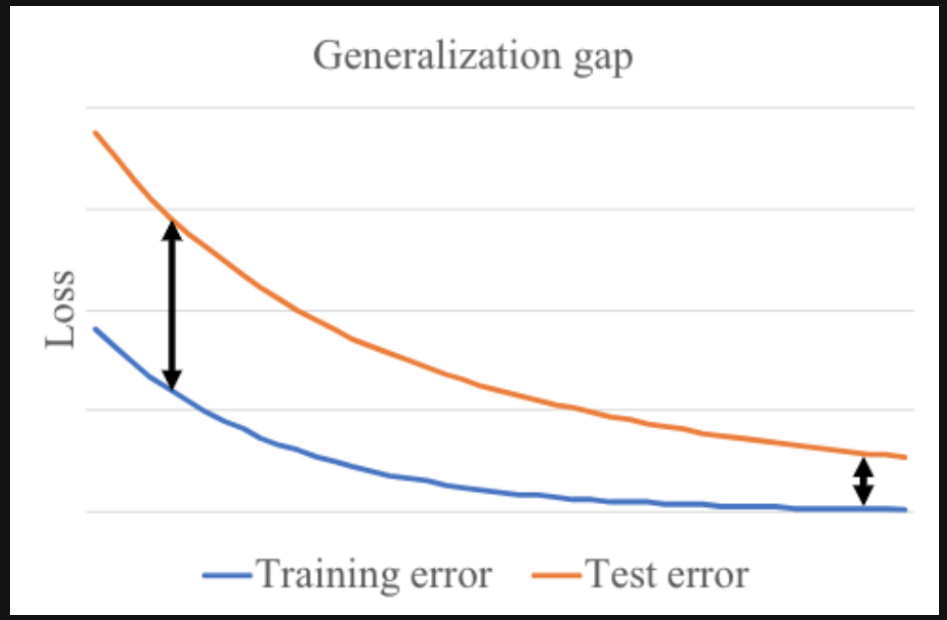
Generalization Capacity

- Transfer learning
- Model design
- Limited data training
- Data augmentation

Generalization in Deep Learning

Generalization gap

- Generalization errors impede deep learning applications.

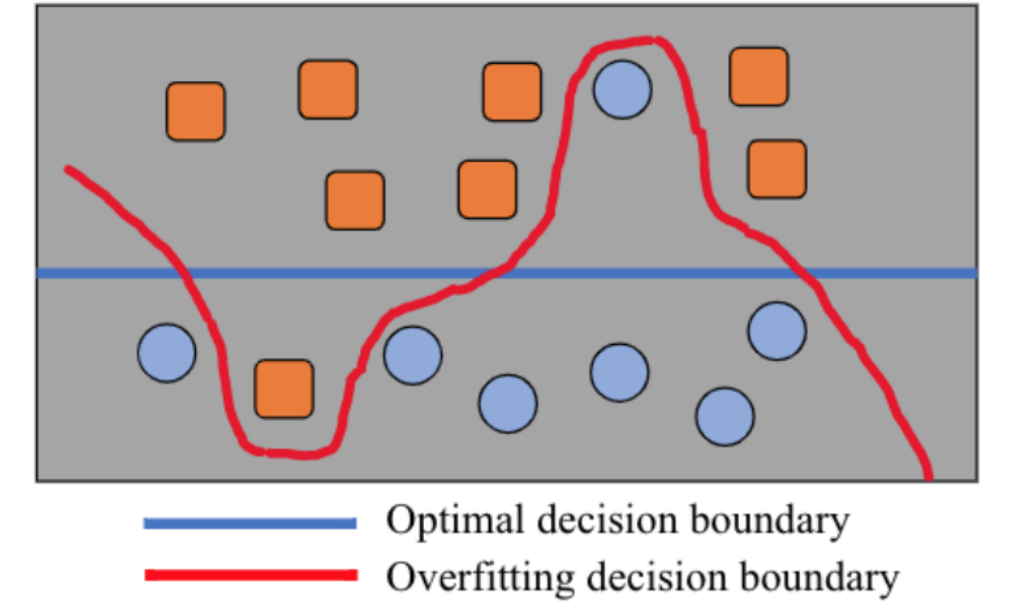


$$G_{M,S} = E_{test} - E_{train},$$

Generalization in Deep Learning

Generalization error

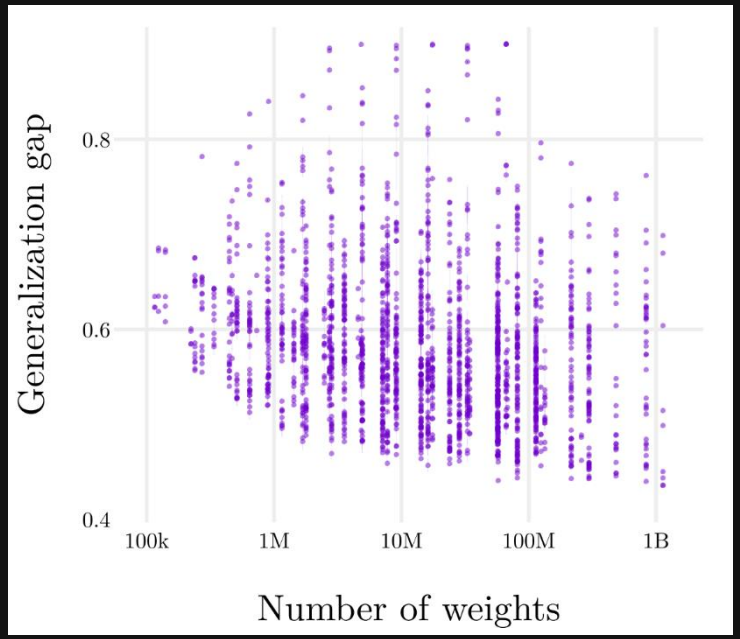
- Generalization errors in deep learning come from **overfitting** instead of underfitting.
- Generalization is believed to be smaller than the penalties of model and dataset complexities.
- As the effect of model in unclear, we want to improve the generalization capability in the **data perspective**.



[C. Zhang et al. ICLR 2017]

$$G_{M,S} \leq P_M + P_S,$$

[K. Kawaguchi, L. Kaelbling, Y. Bengio arxiv:1710.05468]

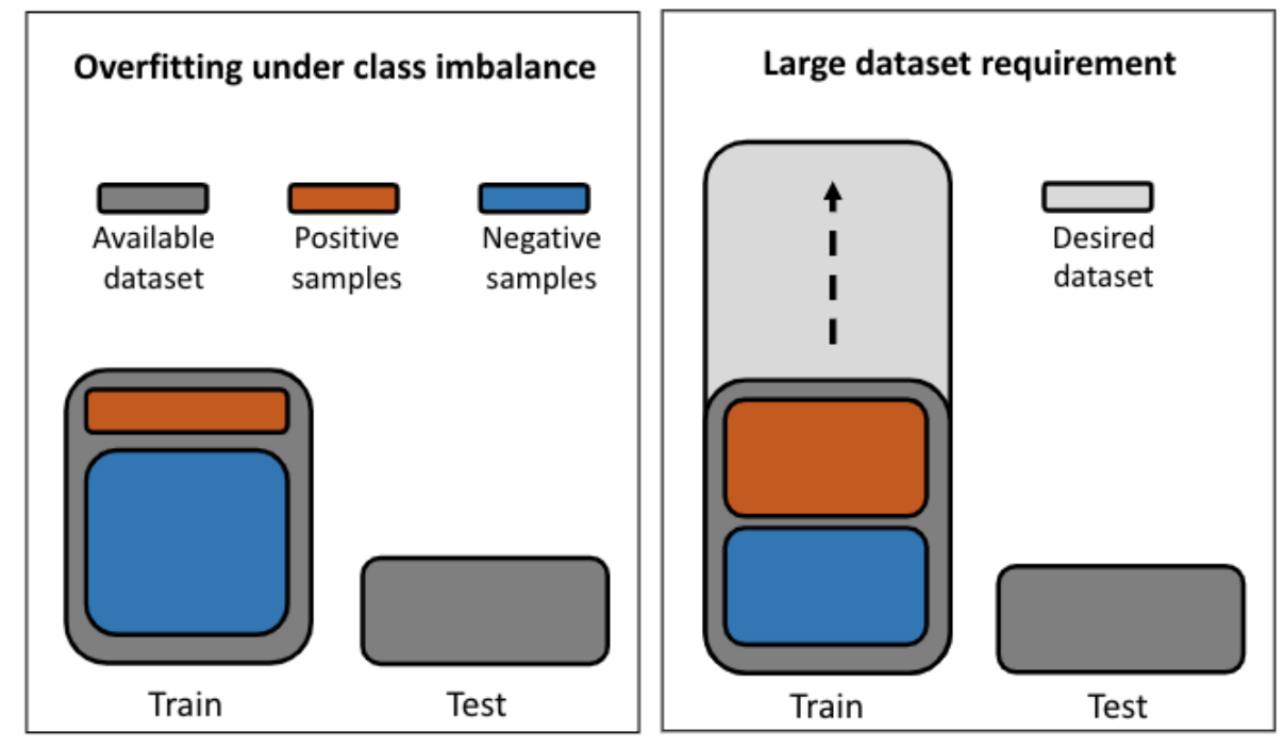


[R Novak, *et al*. ICLR 2018]

Generalization in Deep Learning

Challenges for medical imaging: a data perspective

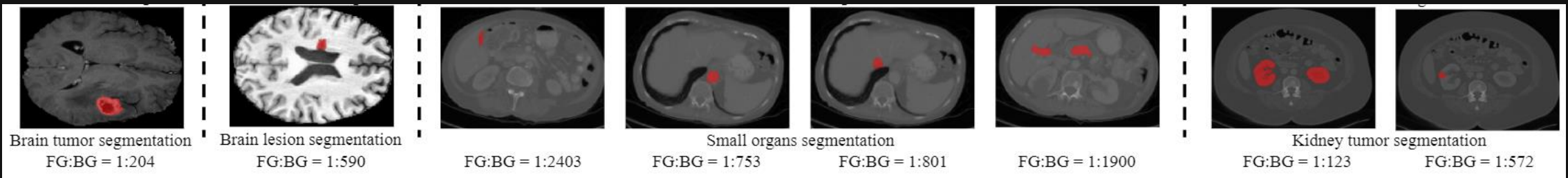
- Target tissues are always very small, leading to **class imbalance**.
- Neural networks need **more data** to generalize well.



Overfitting under Class Imbalance

Data imbalance in segmentation: problem

- Tumor and organs are relatively small in medical imaging.
- Small portions of training data lead to **overfitting of underrepresented classes**.
- The network behavior of overfitting under class imbalance is not clearly understood.



Overfitting under Class Imbalance

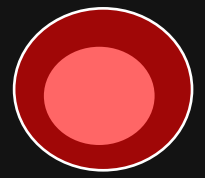
Analysis

- The segmentation is evaluated by DSC:

$$DSC = 2\frac{\text{sensitivity}\cdot\text{precision}}{\text{sensitivity}+\text{precision}}$$

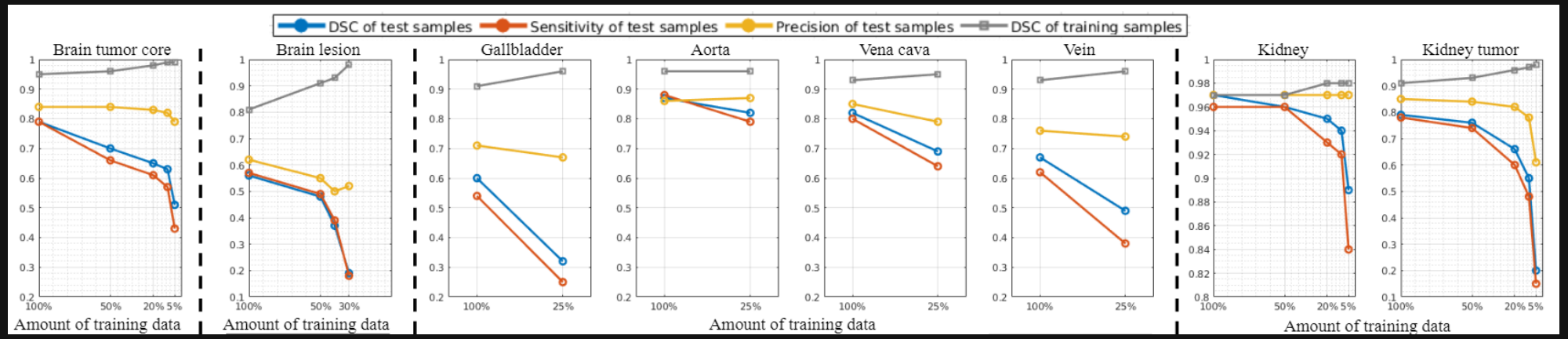
High DSC

Low precision
(over-segment)



Low Sensitivity
(under-segment)

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

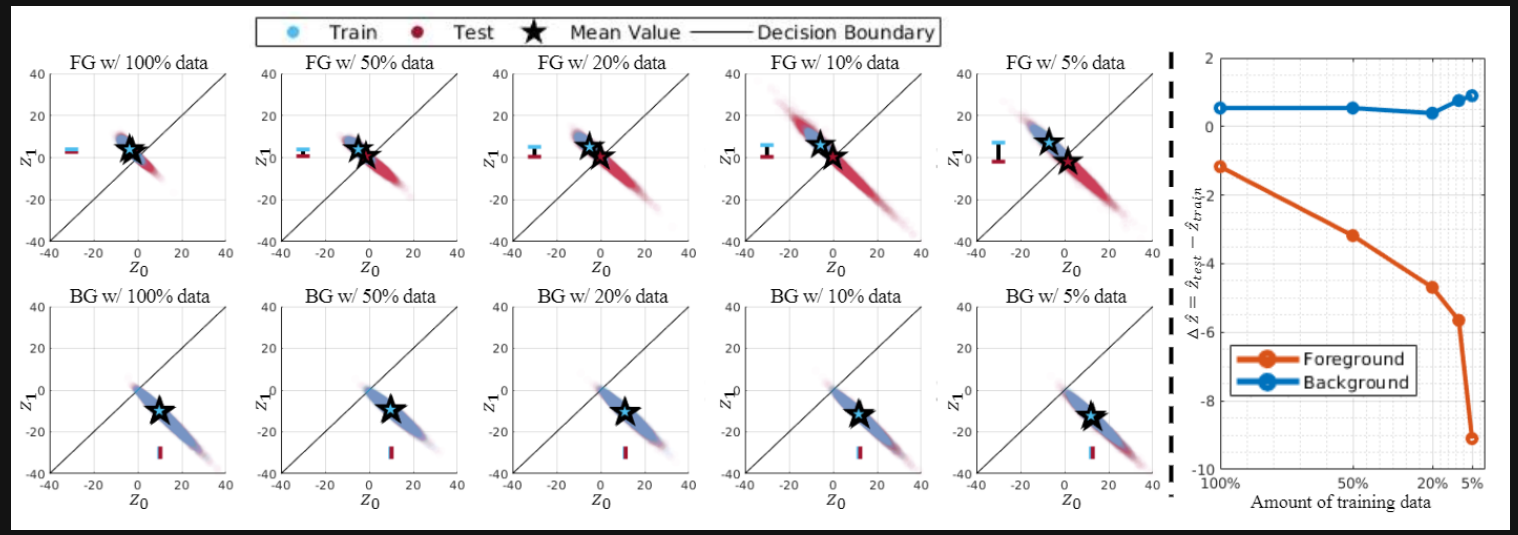


[**Z. Li**, et al. in preparation]

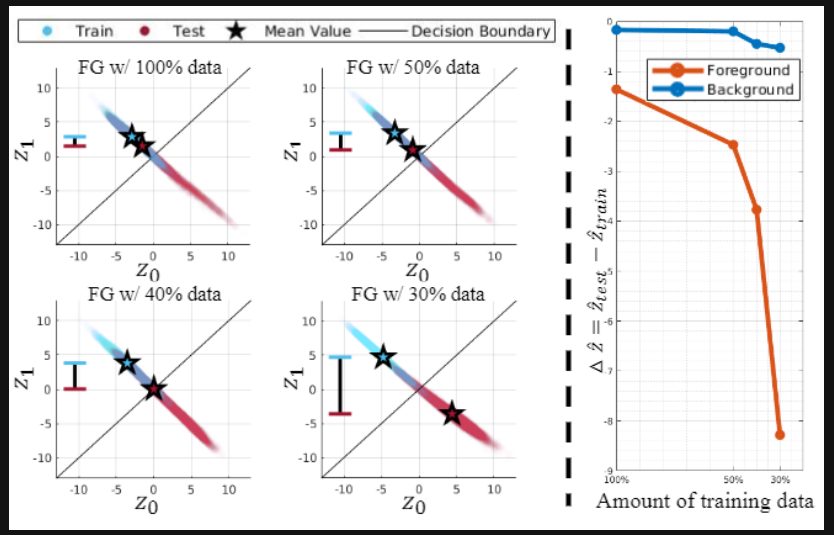
Overfitting under Class Imbalance

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.



DeepMedic with BRATS



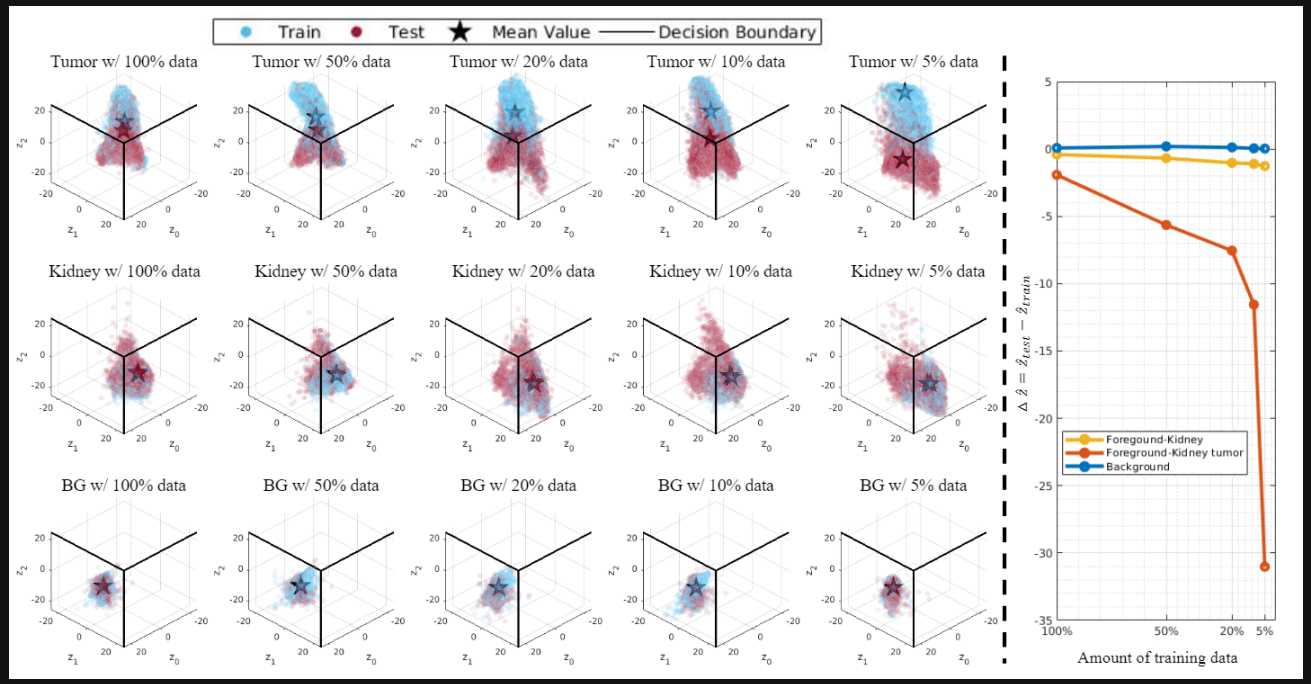
DeepMedic with ATLAS

BioMedIA

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.

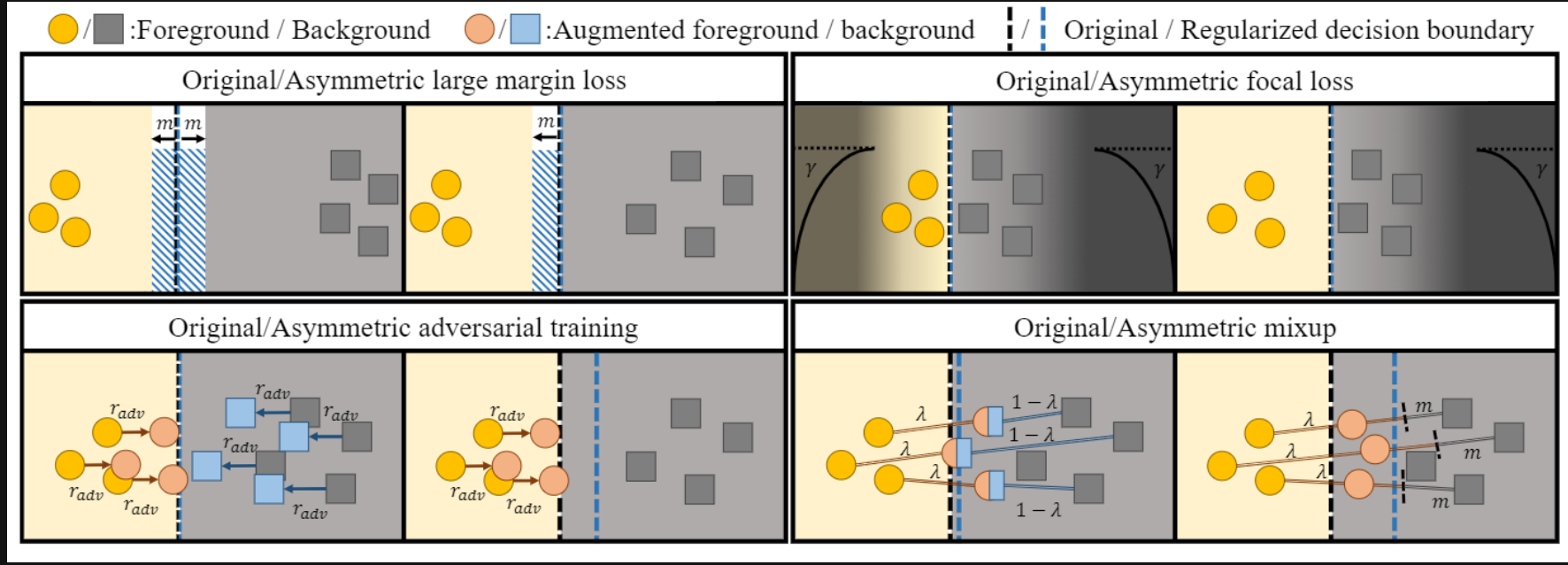


3D U-net with KiTS

Overfitting under Class Imbalance

Method

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



Cross entropy loss

$$L_{CE}(x_i, y_i) = -\frac{1}{N}\sum_c y_{i:c}\log\left(\frac{e^{z_c}}{\sum_k e^{z_k}}\right),$$

Large margin loss

$$L_{CE_M}(x_i, y_i) = -\frac{1}{N}\sum_c y_{i:c}\log\left(\frac{e^{z_c - m}}{e^{z_c - m} + \sum_{k \neq c} e^{z_k}}\right),$$

Focal loss

$$L_{CE_{focal}}(x_i, y_i) = -\frac{1}{N}\sum_c \left(1 - \frac{e^{z_c}}{\sum_k e^{z_k}}\right)^{\gamma} y_{i:c}\log\left(\frac{e^{z_c}}{\sum_k e^{z_k}}\right),$$

Adversarial training

$$L_{adv}(x_i, y_i) = L(x_i, y_i) + L\left(x_i + l \cdot \frac{d_{adv}}{\|d_{adv}\|_2}, y_i\right),$$

$$\text{with} \quad d_{adv} = \underset{d;\|d\|<\epsilon}{\operatorname{argmax}} L(x_i + d, y_i).$$

Mixup

$$\tilde{x_i} = \lambda x_i + (1-\lambda)x_j, \quad \tilde{y_i} = \lambda y_i + (1-\lambda)y_j.$$

Overfitting under Class Imbalance

Results

- The proposed variants of regularization and techniques can moderate overfitting and improve performance.

Method	5% training			10% training			20% training			50% training		
	DSC	SENS	PRC	DSC	SENS	PRC	DSC	SENS	PRC	DSC	SENS	PRC
Vanilla - CE [16]	0.51	0.43	0.79	0.63	0.57	0.82	0.65	0.61	0.83	0.70	0.66	0.84
Vanilla - CE - 80% tumor	0.46	0.38	0.77	0.62	0.55	0.79	0.66	0.61	0.82	0.69	0.65	0.83
Vanilla - F1 (DSC)	0.48	0.39	0.82	0.59	0.52	0.83	0.65	0.59	0.83	0.67	0.63	0.85
Vanilla - F2 [12]	0.47	0.39	0.79	0.60	0.54	0.83	0.66	0.62	0.82	0.69	0.67	0.81
Vanilla - F4 [12]	0.52	0.44	0.79	0.60	0.54	0.82	0.66	0.63	0.81	0.68	0.66	0.83
Vanilla - F8 [12]	0.49	0.40	0.80	0.60	0.54	0.83	0.65	0.61	0.82	0.68	0.66	0.80
Large margin loss [23]	0.46	0.38	0.78	0.61	0.54	0.82	0.67	0.63	0.83	0.67	0.63	0.86
Asymmetric large margin loss	0.55	0.51	0.76	0.64	0.58	0.84	0.68	0.64	0.82	0.69	0.66	0.85
Focal loss [21]	0.54	0.46	0.78	0.63	0.56	0.82	0.65	0.61	0.83	0.67	0.63	0.85
Asymmetric focal loss	0.57	0.53	0.74	0.66	0.63	0.79	0.68	0.67	0.79	0.71	0.72	0.80
Adversarial training [9]	0.53	0.46	0.79	0.62	0.56	0.83	0.65	0.60	0.83	0.66	0.62	0.85
Asymmetric adversarial training	0.57	0.52	0.75	0.64	0.59	0.80	0.68	0.64	0.83	0.71	0.69	0.82
Mixup [35]	0.50	0.42	0.78	0.61	0.55	0.81	0.65	0.60	0.82	0.67	0.63	0.86
Asymmetric mixup	0.59	0.58	0.71	0.69	0.66	0.79	0.71	0.69	0.81	0.71	0.69	0.84
Asymmetric combination	**0.62**	0.66	0.71	**0.71**	0.73	0.75	**0.72**	0.74	0.79	**0.73**	0.77	0.78

DeepMedic with BRATS

Overfitting under Class Imbalance

Results

- The proposed variants of regularization and techniques can work well with existing regularization techniques.

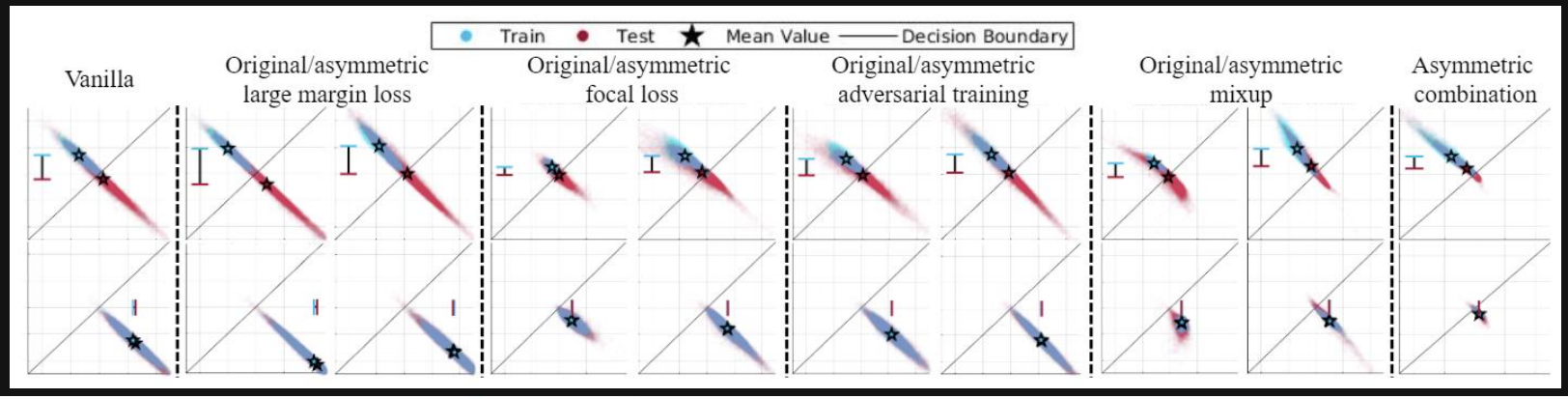
Method	5% training						10% training					
	Kidney			Kidney tumor			Kidney			Kidney tumor		
	DSC	SENS	PRC	DSC	SENS	PRC	DSC	SENS	PRC	DSC	SENS	PRC
Vanilla - w/ augmentation [15]	0.89	0.84	0.97	0.20	0.15	0.61	0.94	0.92	0.97	0.55	0.48	0.78
Vanilla - w/o augmentation	0.87	0.82	0.96	0.06	0.04	0.31	0.93	0.90	0.97	0.38	0.33	0.63
Vanilla - asymmetric augmentation	0.90	0.86	0.97	0.22	0.16	0.63	0.95	0.93	0.97	0.56	0.50	0.75
Large margin loss [23]	0.90	0.86	0.97	0.20	0.15	0.52	0.95	0.93	0.97	0.55	0.48	0.77
Asymmetric large margin loss	0.90	0.86	0.97	0.21	0.17	0.56	0.94	0.92	0.97	0.56	0.50	0.75
Focal loss [21]	0.88	0.83	0.98	0.17	0.12	0.51	0.92	0.86	0.99	0.48	0.39	0.78
Asymmetric focal loss	0.88	0.82	0.98	0.21	0.16	0.55	0.92	0.87	0.99	0.57	0.50	0.75
Adversarial training [9]	0.88	0.84	0.97	0.17	0.12	0.52	0.94	0.92	0.97	0.52	0.45	0.80
Asymmetric adversarial training	0.90	0.85	0.97	0.19	0.14	0.49	0.95	0.93	0.97	0.57	0.51	0.79
Mixup [35]	0.91	0.87	0.97	0.19	0.15	0.56	0.95	0.94	0.97	0.55	0.48	0.80
Asymmetric mixup	0.91	0.87	0.97	0.18	0.13	0.58	0.95	0.93	0.97	0.55	0.49	0.80
Asymmetric combination	0.89	0.84	0.98	**0.25**	0.19	0.60	0.94	0.90	0.98	**0.59**	0.54	0.77
Method	50% training						100% training					
	Kidney			Kidney tumor			Kidney			Kidney tumor		
	DSC	SENS	PRC	DSC	SENS	PRC	DSC	SENS	PRC	DSC	SENS	PRC
Vanilla - w/ augmentation	0.96	0.96	0.97	0.76	0.74	0.84	0.97	0.96	0.97	0.79	0.78	0.85
Vanilla - w/o augmentation	0.96	0.96	0.97	0.64	0.62	0.75	0.97	0.96	0.97	0.71	0.70	0.79
Vanilla - asymmetric augmentation	0.96	0.96	0.96	0.75	0.74	0.84	0.96	0.97	0.96	0.78	0.79	0.83
Large margin loss [23]	0.96	0.96	0.97	0.77	0.75	0.84	0.96	0.97	0.96	0.81	0.82	0.83
Asymmetric large margin loss	0.96	0.96	0.97	0.78	0.76	0.85	0.97	0.97	0.97	**0.82**	0.82	0.84
Focal loss [21]	0.94	0.90	0.99	0.73	0.67	0.86	0.94	0.90	0.99	0.79	0.74	0.88
Asymmetric focal loss	0.95	0.91	0.99	0.78	0.77	0.85	0.95	0.91	0.99	0.81	0.81	0.84
Adversarial training [9]	0.97	0.96	0.97	0.74	0.72	0.83	0.97	0.96	0.97	0.81	0.82	0.84
Asymmetric adversarial training	0.97	0.96	0.97	0.77	0.75	0.86	0.97	0.97	0.97	**0.82**	0.81	0.86
Mixup [35]	0.97	0.96	0.97	0.77	0.74	0.86	0.97	0.97	0.97	0.81	0.79	0.85
Asymmetric mixup	0.96	0.96	0.97	0.77	0.75	0.85	0.97	0.96	0.97	0.80	0.79	0.86
Asymmetric combination	0.94	0.91	0.98	**0.79**	0.79	0.85	0.97	0.96	0.98	**0.82**	0.83	0.85

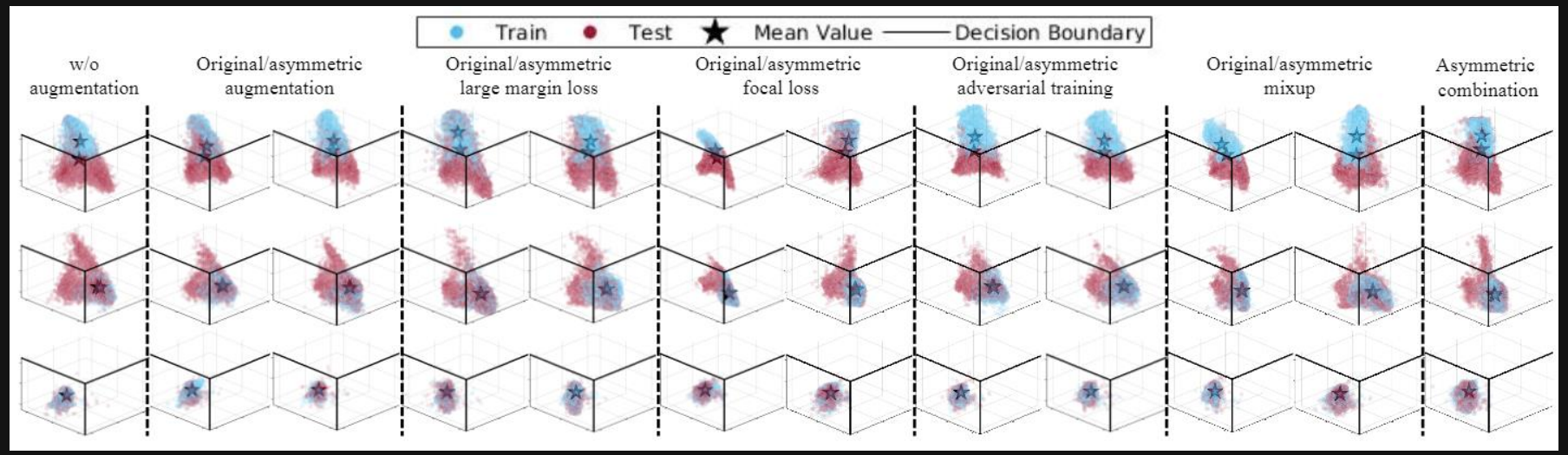
3D U-net with KiTS

Overfitting under Class Imbalance

Results

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

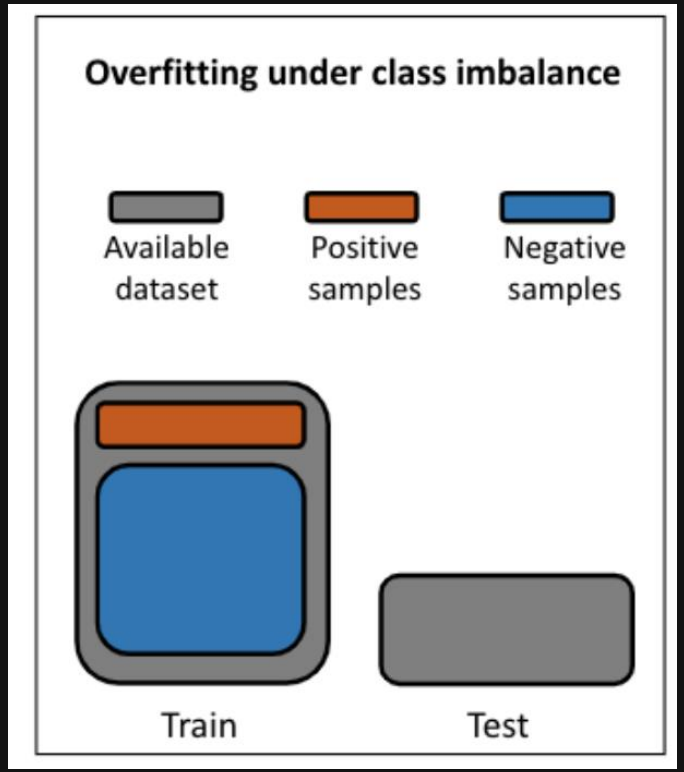




Overfitting under Class Imbalance

Conclusion

- Overfitting under class imbalance leads to **loss of sensitivity**.
- The distribution of logit activations when processing unseen test samples of an **under-represented class tends to shift** towards and even across the decision boundary.
- We propose several asymmetric techniques based on our observations of logit distribution.
- **Logit distribution plots** can be a valuable tool for practitioners to study overfitting and other behaviour of different models.



Learning the Sampling Distribution for Data Augmentation

Data augmentation is useful, but..

- It needs **prior domain knowledge** to design.
- Optimal strategies are **application specific** and difficult to hand engineer.

	BraTS	Liver lowres	Liver fullres	Hippocampus	Prostate	Lung nodule	Pancreas
Vanilla nnU-Net	0.72	0.79	0.78	0.89	0.77	0.65	0.65
Batch norm instead of Inst. norm	1.0%	-0.1%	2.9%	-0.1%	-1.3%	-14.2%	-3.7%
No feature map normalization	1.1%	-4.6%	-22.8%	-0.2%	-4.2%	3.0%	-100.0%
ReLU instead of LeakyReLU	0.6%	0.0%	1.0%	-0.1%	-0.2%	-0.4%	0.5%
No data augmentation	-0.8%	-4.9%	1.5%	-1.5%	-0.4%	4.2%	-11.3%
Only cross-entropy loss	-0.6%	-12.0%	-6.3%	0.0%	-1.4%	-25.4%	-8.8%
Only dice loss	0.9%	-2.5%	-10.1%	-0.3%	-3.0%	-11.5%	1.6%

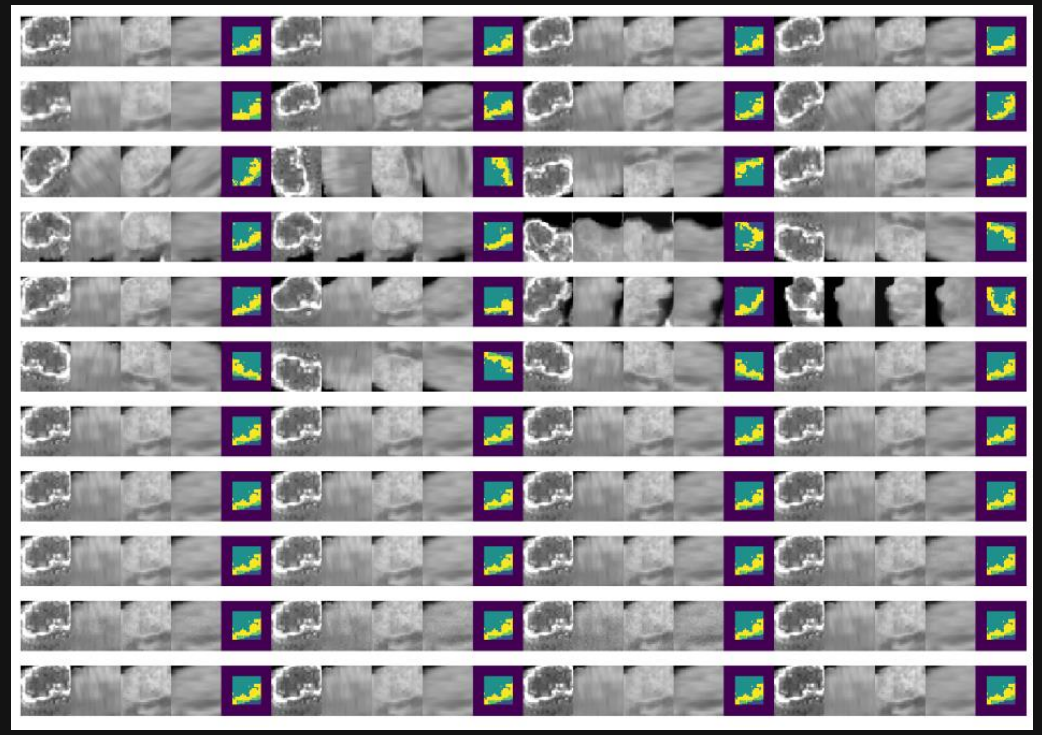
[F. Isensee et al. arxiv:1904.08128]

Learning the Sampling Distribution for Data Augmentation

Method

- We aim to learn the data augmentation strategy by drawing transformations from **a learnable probability distribution**.

ID	Operation Name	Description	Range of magnitudes
0	Identity	No augmentation, $\mathcal{A}(I) = I$	None
1	Scaling	Scale up/down the image	[0.05, 0.10, 0.15, 0.20, 0.25]
2	RotateLongitudinal	Rotate the image along longitudinal axis anticlockwise/clockwise	[10, 20, 30, 90, 180]
3	RotateFrontal	Rotate the image along frontal axis anticlockwise/clockwise	[10, 20, 30, 90, 180]
4	RotateSagittal	Rotate the image along sagittal axis anticlockwise/clockwise	[10, 20, 30, 90, 180]
5	Mirroring	Flip the sample in different planes	[Sagittal, Frontal, Axial]
6	Contrast	$\mathcal{A}(I) = I * (1 \pm \text{scale})$, add/reduce image contrast	[0.01, 0.02, 0.03, 0.04, 0.05]
7	Brightness	$\mathcal{A}(I) = I \pm \text{shift}$, add/reduce image intensity	[0.05, 0.10, 0.15, 0.20, 0.25]
8	Gaussian noise	Add Gaussian noise with different variances	[0.05, 0.10, 0.15, 0.20, 0.25]
9	Gamma correction	Scale I to [0,1], then $\mathcal{A}(I) = I^{1\pm\gamma}$, and scale it back	[0.05, 0.10, 0.15, 0.20, 0.25]



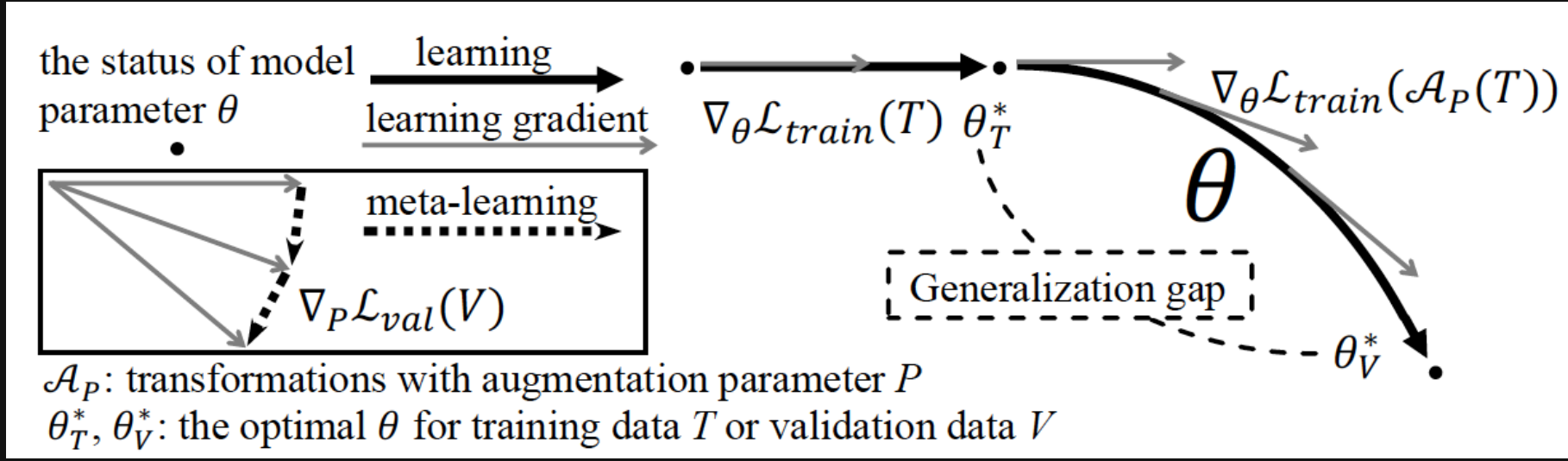
[**Z. Li**, et al. in preparation]

BioMedIA

Learning the Sampling Distribution for Data Augmentation

Method

- We aim to close this generalization gap explicitly by learning a probability distribution of data augmentations *P* based on **meta-gradients**.



Learning the Sampling Distribution for Data Augmentation

Method

- The method mainly contains three steps in one iteration:
  - Sample a data augmentation strategy;

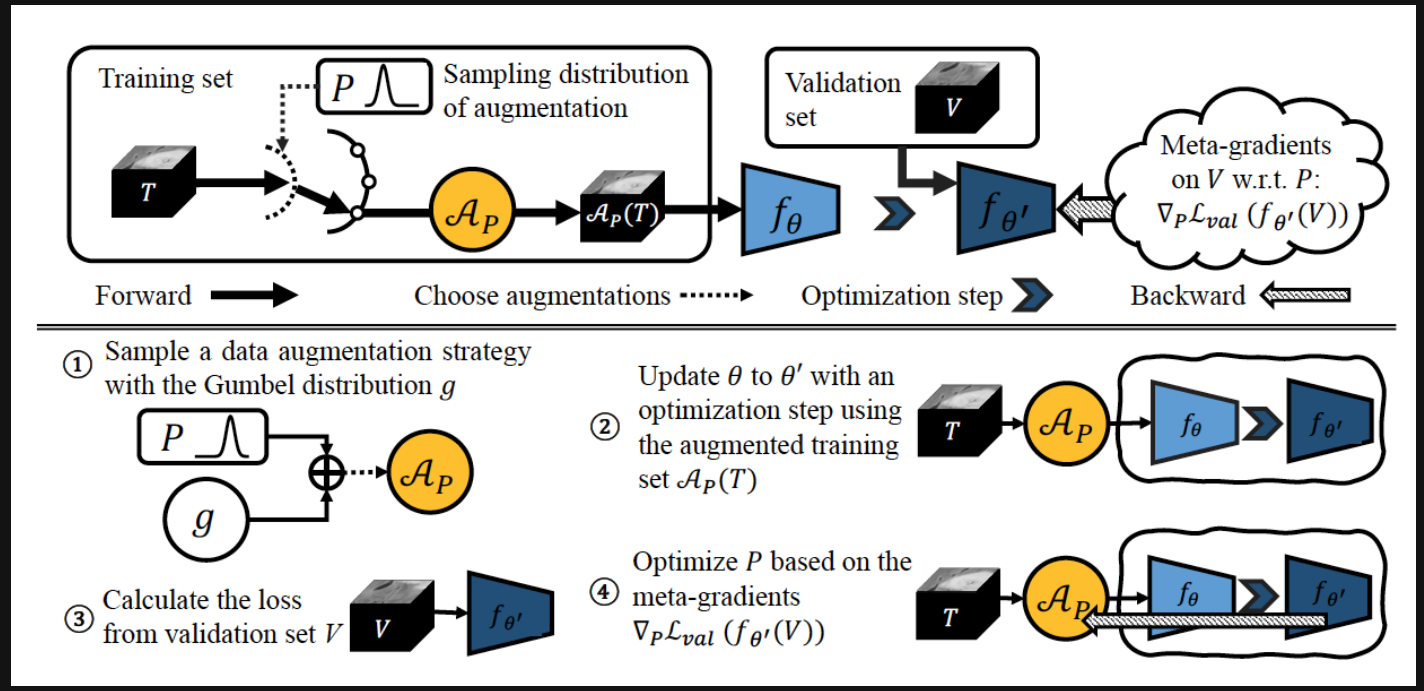
$$\mathcal{A}_P = \mathcal{A}[\underset{i}{\text{argmax}}(\frac{e^{\log(p_1)+g_1}, ..., e^{\log(p_i)+g_i}, ..., e^{\log(p_N)+g_N}}{\sum_k e^{\log(p_k)+g_k}})],$$

  - Update the segmentation model to $f_{\theta'}$;

$$\theta' = \theta - \alpha \nabla_\theta \mathcal{L}_{train}(f_\theta(\mathcal{A}_P(T))).$$

  - Optimize P with meta-gradients by evaluating V.

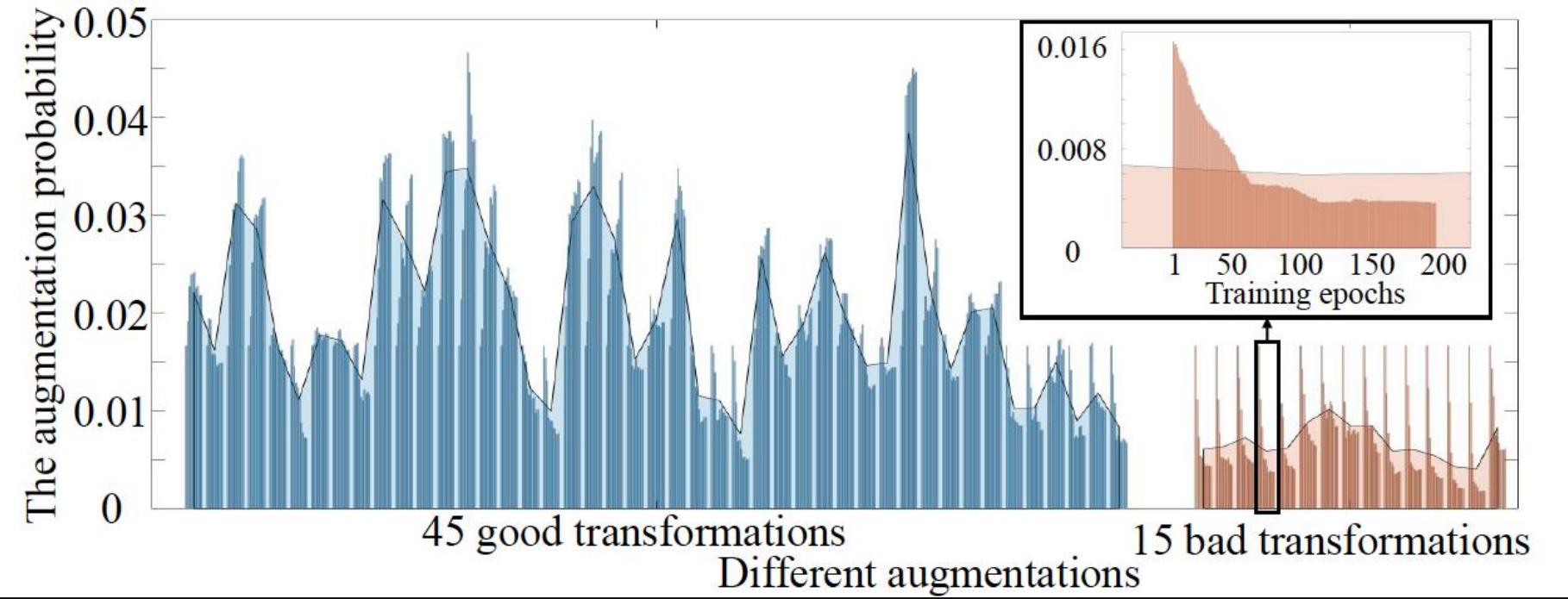
$$P^{t+1} = P^t - \beta \nabla_{P^t} \mathcal{L}_{val}(f_{\theta'}(V))$$



Learning the Sampling Distribution for Data Augmentation

Results: Proof of Concept

- We **manually add 15 bad transformations** to the augmentation set for CIFAR10 based on wide residual net. Our method learns to decrease their probabilities during training.



Learning the Sampling Distribution for Data Augmentation

Results: Medical Image Segmentation

- Our learned augmentation policy consistently improves the performance on all three segmentation tasks.

DeepMedic with ATLAS

training data		Lesion DSC	Lesion SENS	Lesion PRC
40%	w/o augment	31.4	33.0	45.7
	heuristic Kamnitsas et al. (2017)	**43.3**	53.2	49.8
	random Cubuk et al. (2019)	42.8	47.5	54.4
	learned	**43.3**	49.0	54.1
50%	w/o augment	39.5	35.7	57.9
	heuristic Kamnitsas et al. (2017)	51.2	58.0	56.2
	random Cubuk et al. (2019)	51.9	55.6	58.0
	learned	**53.7**	57.8	59.8
100%	w/o augment	50.5	47.9	64.2
	heuristic Kamnitsas et al. (2017)	58.2	60.5	64.4
	random Cubuk et al. (2019)	58.8	62.1	63.4
	learned	**61.1**	61.8	68.9

DeepMedic with KiTS

training data		Kidney DSC	Kidney SENS	Kidney PRC	Tumor DSC	Tumor SENS	Tumor PRC
20%	w/o augment	87.3	82.5	96.4	15.9	11.7	49.3
	heuristic Kamnitsas et al. (2017)	92.3	90.0	95.8	51.4	50.3	63.6
	random Cubuk et al. (2019)	94.0	91.9	96.8	52.4	45.1	75.5
	learned	**94.7**	93.4	96.3	**55.5**	48.7	74.9
50%	w/o augment	93.3	90.8	96.8	35.4	29.9	59.6
	heuristic Kamnitsas et al. (2017)	95.1	95.0	95.4	59.9	56.6	71.3
	random Cubuk et al. (2019)	95.0	95.9	94.7	66.3	65.8	74.1
	learned	**95.5**	96.0	95.5	**67.4**	64.8	75.7
100%	w/o augment	94.6	93.4	96.4	51.8	47.0	67.7
	heuristic Kamnitsas et al. (2017)	95.5	96.7	94.5	66.7	72.2	68.5
	random Cubuk et al. (2019)	**96.4**	96.7	96.2	70.4	70.0	76.2
	learned	96.3	97.3	95.3	**72.9**	74.4	76.3

DeepMedic with BRATS

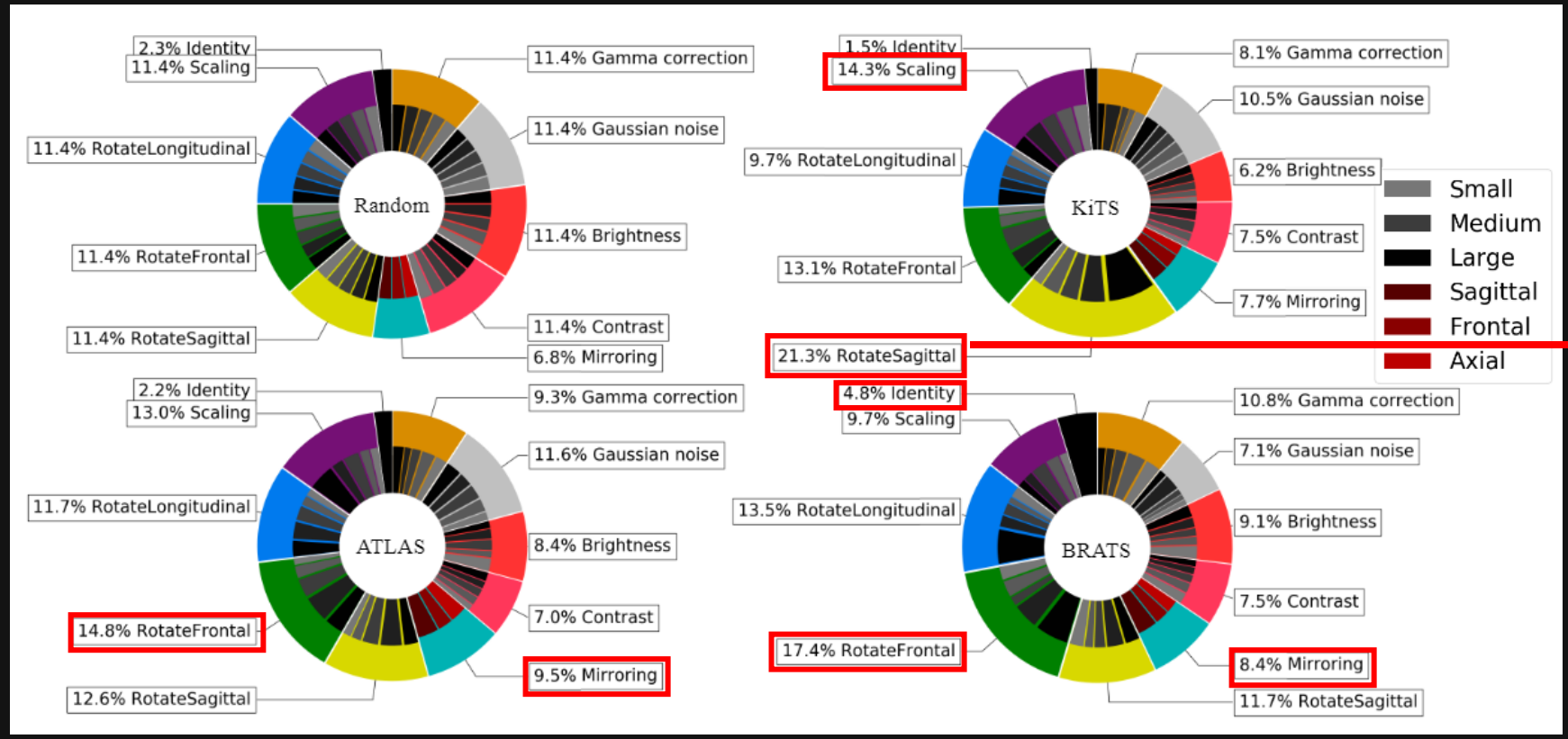
training data		Whole DSC	Whole SENS	Whole PRC	Core DSC	Core SENS	Core PRC	Enhancing DSC	Enhancing SENS	Enhancing PRC
10%	w/o augment	78.9	71.2	94.0	65.1	61.9	80.4	62.9	57.9	77.5
	heuristic Kamnitsas et al. (2017)	85.7	80.7	93.7	71.9	67.5	83.7	67.9	62.8	81.3
	random Cubuk et al. (2019)	85.8	80.9	94.2	74.3	71.0	84.4	69.5	66.3	80.6
	learned	**87.2**	83.0	94.0	**75.0**	70.9	86.1	**69.9**	65.5	81.6
50%	w/o augment	88.2	85.0	93.0	75.9	73.7	86.8	73.4	73.0	78.7
	heuristic Kamnitsas et al. (2017)	89.0	87.0	92.5	77.8	75.8	88.4	74.5	73.6	82.0
	random Cubuk et al. (2019)	89.4	88.3	91.7	77.4	75.1	88.8	74.0	73.5	81.7
	learned	**89.6**	87.9	92.6	**78.1**	77.1	85.2	**74.6**	75.2	79.2
100%	w/o augment	89.1	87.5	91.7	78.9	77.1	87.2	75.2	76.0	79.0
	heuristic Kamnitsas et al. (2017)	89.8	88.3	92.3	79.4	78.8	87.5	75.2	76.4	80.2
	random Cubuk et al. (2019)	90.0	89.5	91.7	80.4	79.0	89.4	75.2	75.6	81.1
	learned	**90.2**	88.4	92.8	**81.0**	80.0	87.3	**75.8**	77.2	79.5

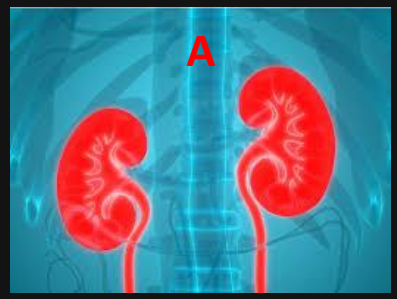
BioMedIA

Learning the Sampling Distribution for Data Augmentation

Results: Medical Image Segmentation

- The learned probability distributions seem to reflect well the type of data harmonization that has been carried out by the providers of the datasets.



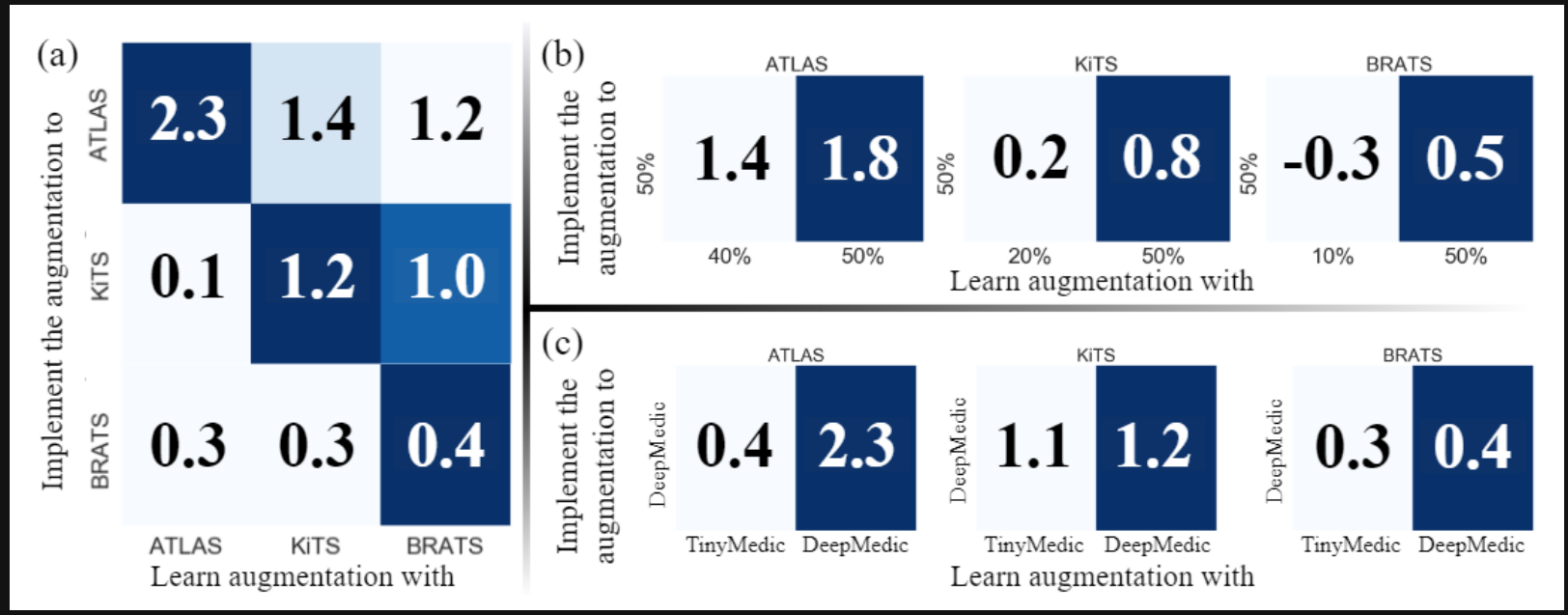


Rotation large (180°) in the sagittal axis

Learning the Sampling Distribution for Data Augmentation

Results: Medical Image Segmentation

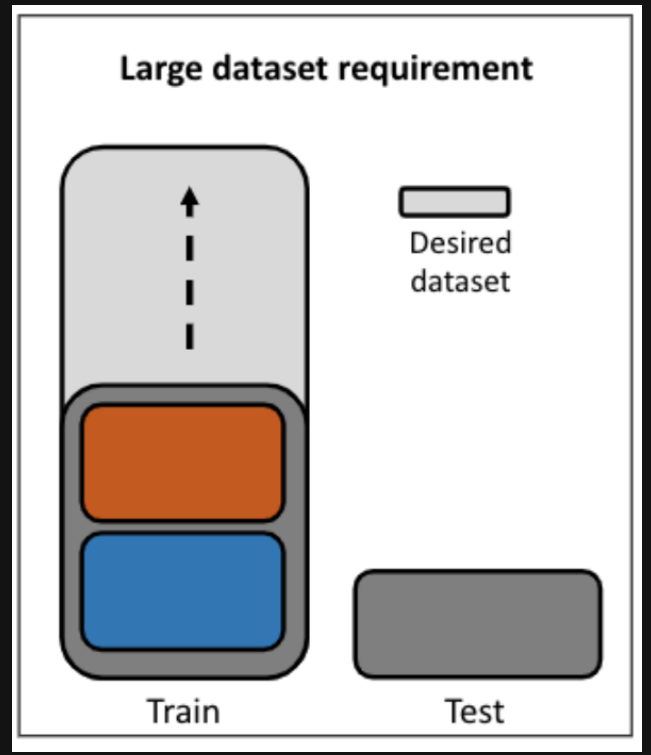
- The optimal augmentation strategies varied between **datasets**, **models** and **the size of training datasets**.



Learning the Sampling Distribution for Data Augmentation

Conclusion

- We propose to automate the process of data augmentation by **meta-gradients** with **high efficiency in both time and data**. (just twice the time of normal training!)
- We can provide optimal augmentation strategies for different application scenarios.
- In our case, **geometry transformations** are more likely to be chosen than intensity transformations.
- Random augmentation is a very strong baseline.
- Different **dataset**, **model**, **size of training dataset** favour different data augmentation strategies.



BioMedIA

Conclusion

Take home message

- Medical image segmentation is a well-studied research area, but there are still many open problems (opportunities).
- We are focusing on improving the generalization capacity of neural networks, which is **a practical and fundamental problem** for medical image segmentation, from the **data perspective**.
  - We observe the logit distribution of image segmentation and propose asymmetric techniques to counter **overfitting under class imbalance**.
  - We propose to learn the sampling distribution of **data augmentation** and provide optimal augmentation strategies.

Useful tools

Model

- nnU-net
  - https://github.com/MIC-DKFZ/nnUNet
- DeepMedic
  - https://github.com/deepmedic/deepmedic

Dataset

- Grand challenge
  - https://grand-challenge.org/
- BRATS
  - https://www.med.upenn.edu/sbia/brats2018/data.html
- KiTS
  - https://kits19.grand-challenge.org/data/
- ATLAS
  - http://fcon_1000.projects.nitrc.org/indi/retro/atlas.html

Thank you!

Contact: zeju.li18@imperial.ac.uk