Learning with limited supervision
(or leveraging unlabeled data)

Ismail Ben Ayed
SystemX

June 2022

Few-shot learning

Training on **base** classes

Few-shot tasks at testing time

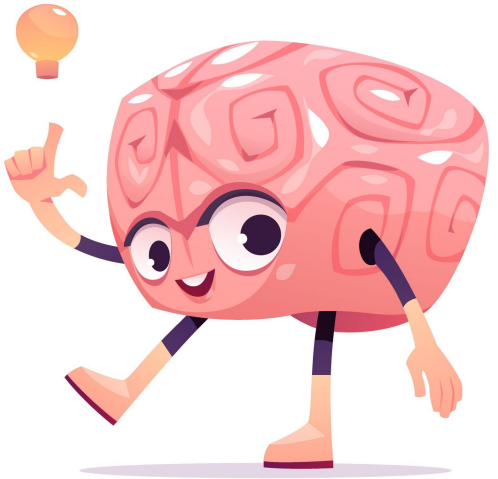
Learn from a few examples per **new** class

Classify these

Few-shot learning

-- **Humans recognize easily** with few examples

-- Modern ML generalize very poorly



Why it is interesting:

Available data sets represent small sub-domains of the world

Cityscapes (5k images; 1.5h per image):

Urban scenes, less than 30 classes

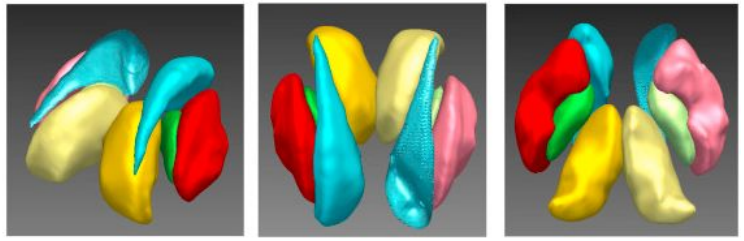
A dense prediction task: semantic segmentation

New classes, but with few examples

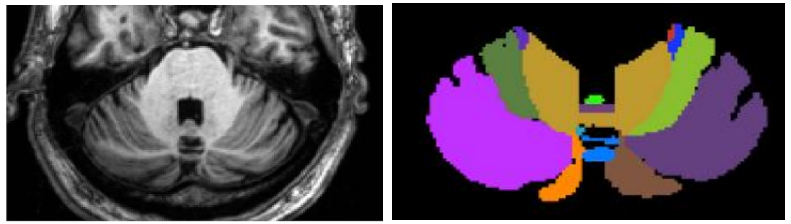
Building labels for dense prediction tasks is even worse

(e.g., semantic segmentation)

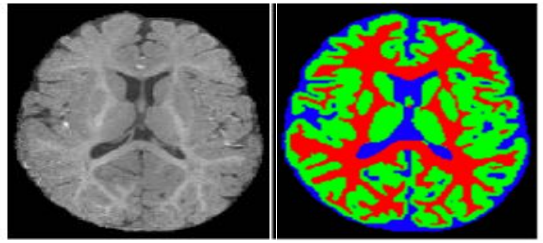
…and it gets more complex in medical imaging



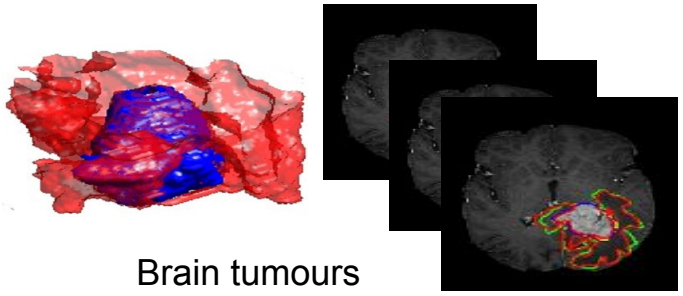
Subcortical structures
(Dolz et al., Neuroimage 2018)



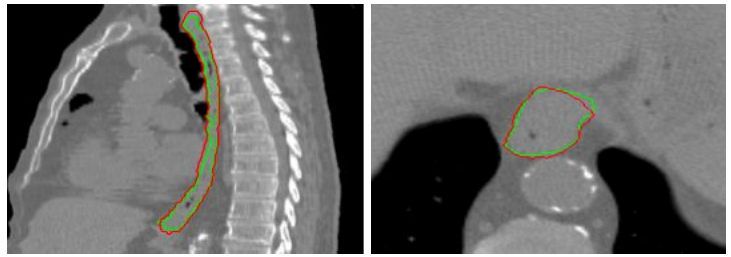
Cerebellum parcellation
(Carass et al., Neuroimage 2018)



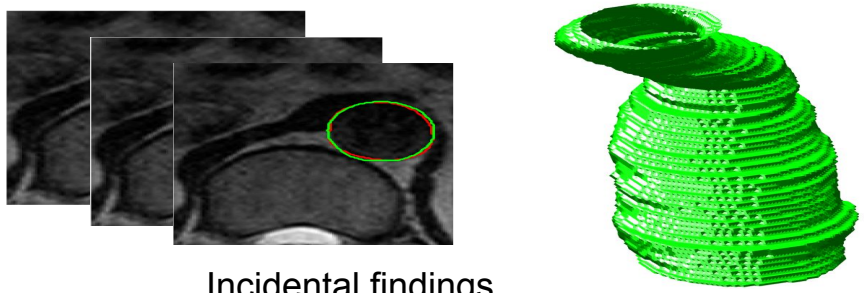
Brain tissues (6-month infant)
(Li et al., TMI 2019)



Brain tumours
(Njeh et al., CMIG 2015)



Organs at risk
(Dolz et al., Med. Phys. 2017)



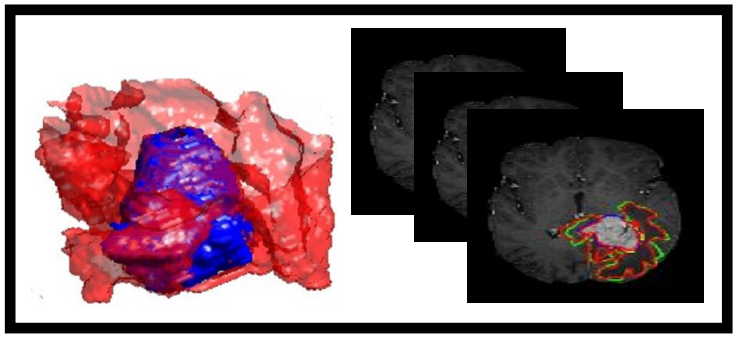
Incidental findings
(Ben Ayed et al., MICCAI 2014)

Labels: Not only expensive, but might need expert knowledge

In medical image segmentation: we are not anywhere close to the 5k of Cityscapes

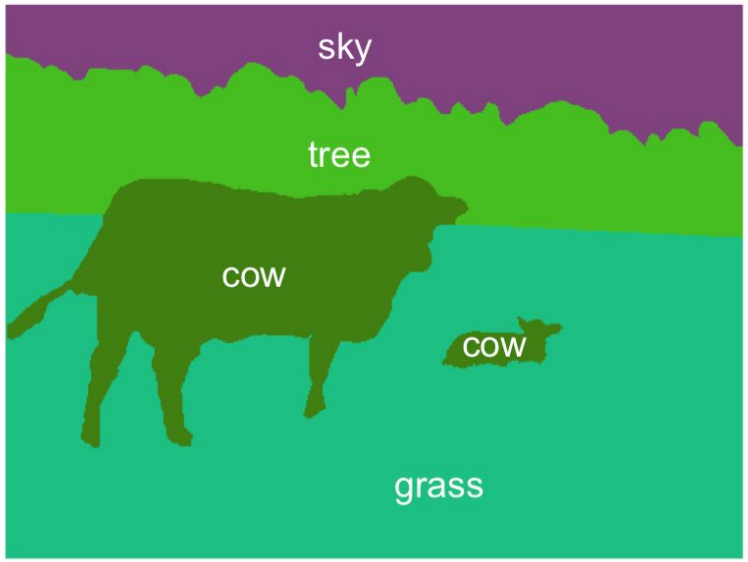
Crowdsourcing?

Dense 3D annotations: several hours
(of radiologist time)

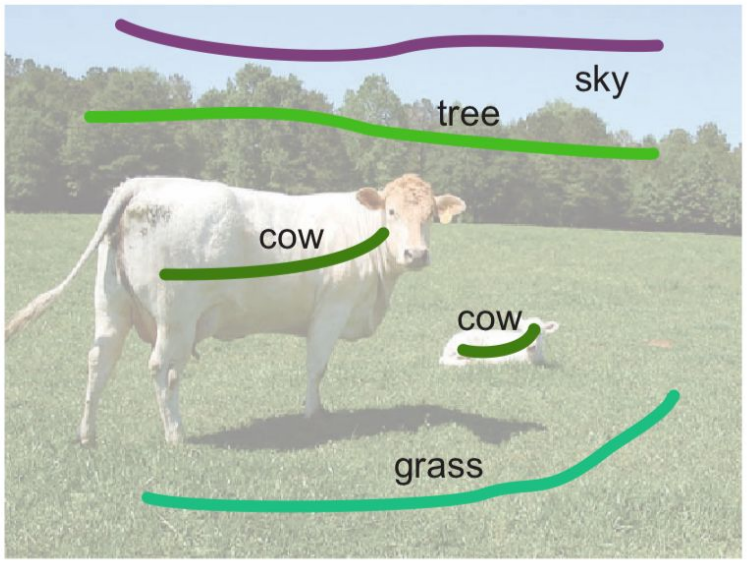


Semi-supervised learning (SSL)

A lot of non-annotated data, and a fraction of points annotated



Full annotations

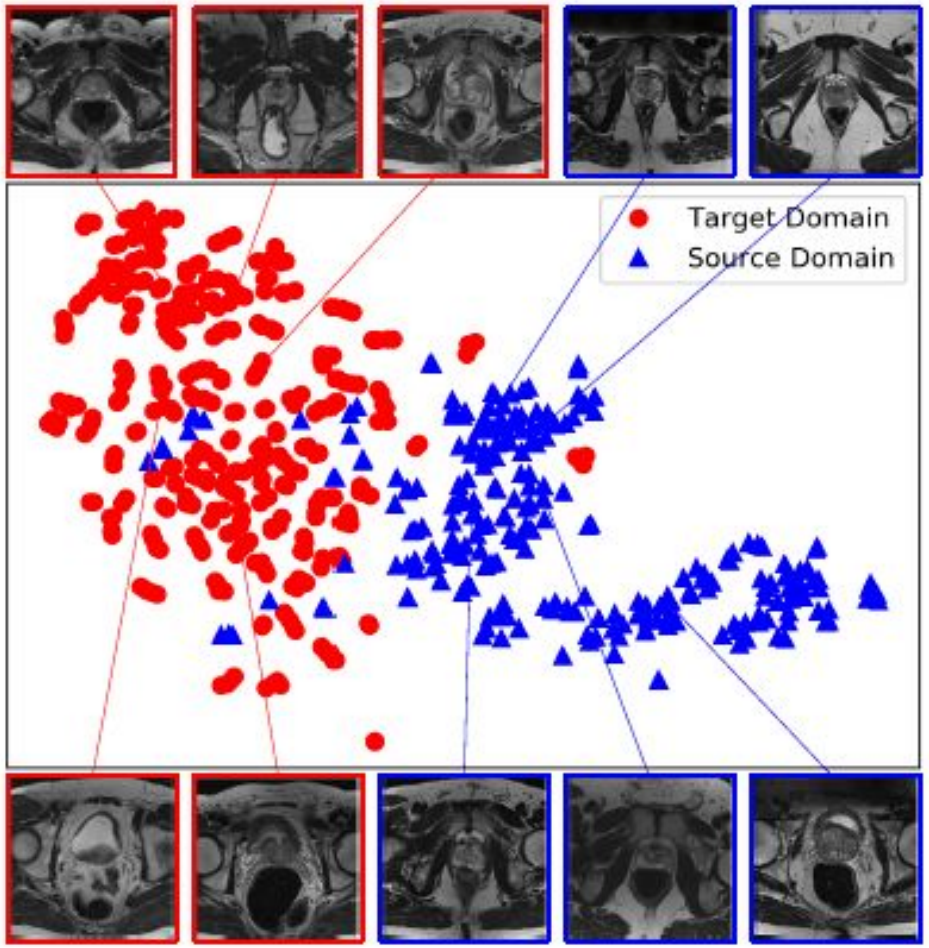


Semi-supervised

Figures from Lin et al. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation, CVPR 2016

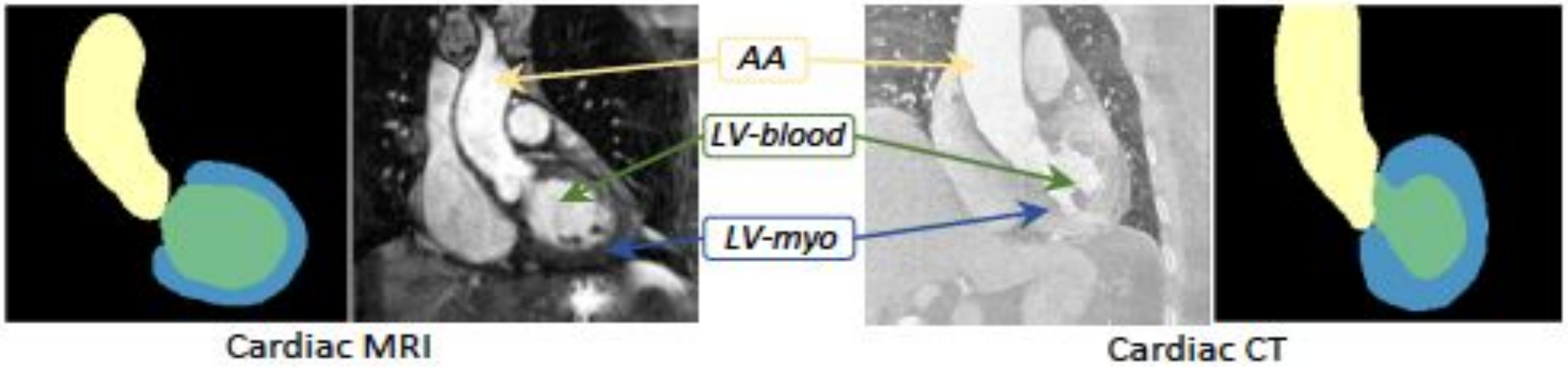
Domain shifts make things worse

(even for the same task and with full annotations in one domain)



[MRI Prostate segmentation: Figure from Zhu et al., Boundary-weighted Domain Adaptive Neural Network for Prostate MR Image Segmentation ArXiv 2019]

Domain shifts: within and across modalities

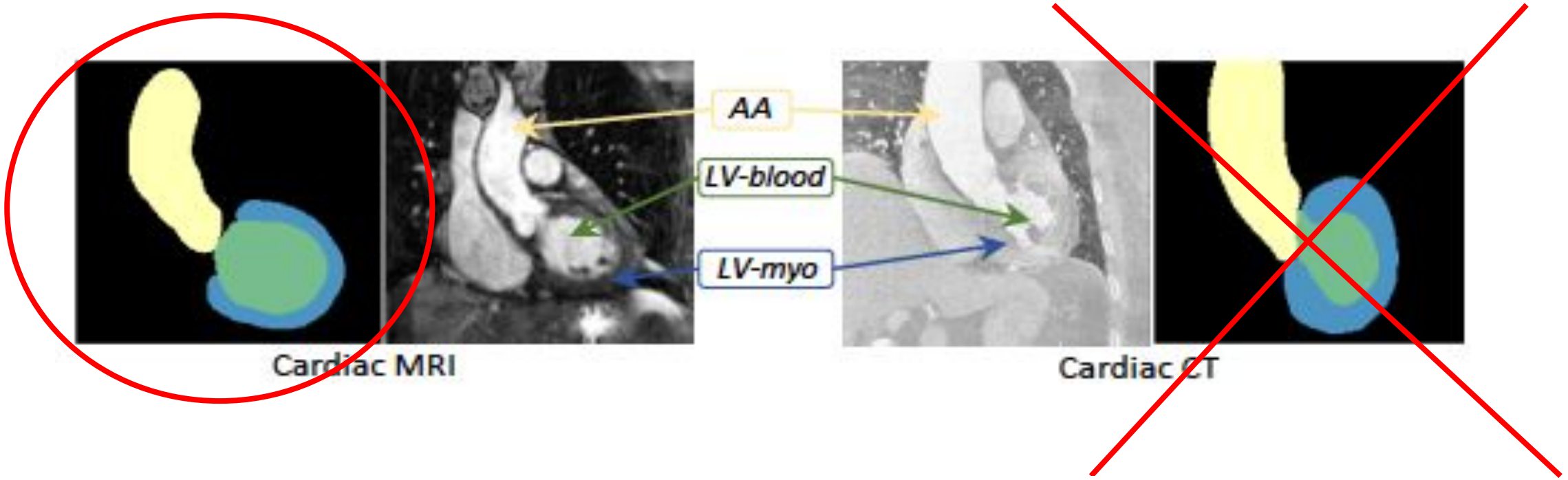


[Images from Dou et al., PnP-AdaNet: Plug-and-play adversarial domain adaptation network with a benchmark at cross-modality cardiac segmentation ArXiv 2018]

Unsupervised domain adaptation (UDA)

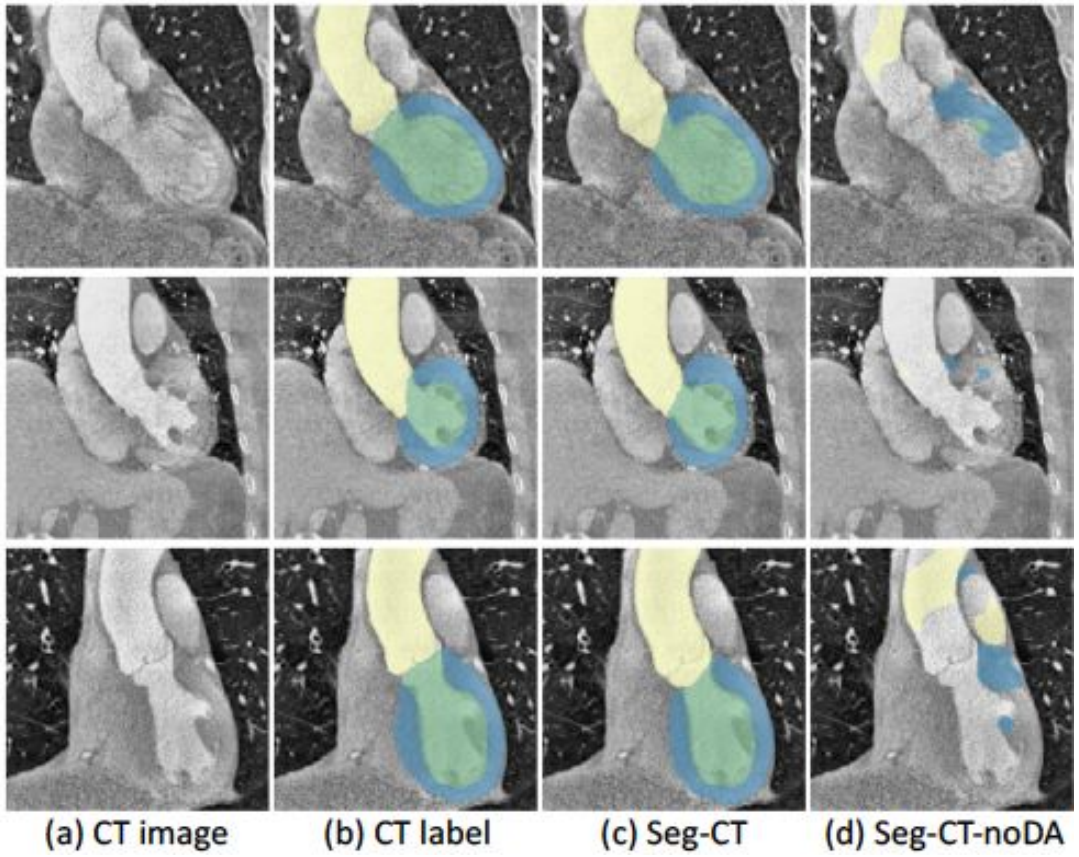
We have labels for the source domain

No labels for the target



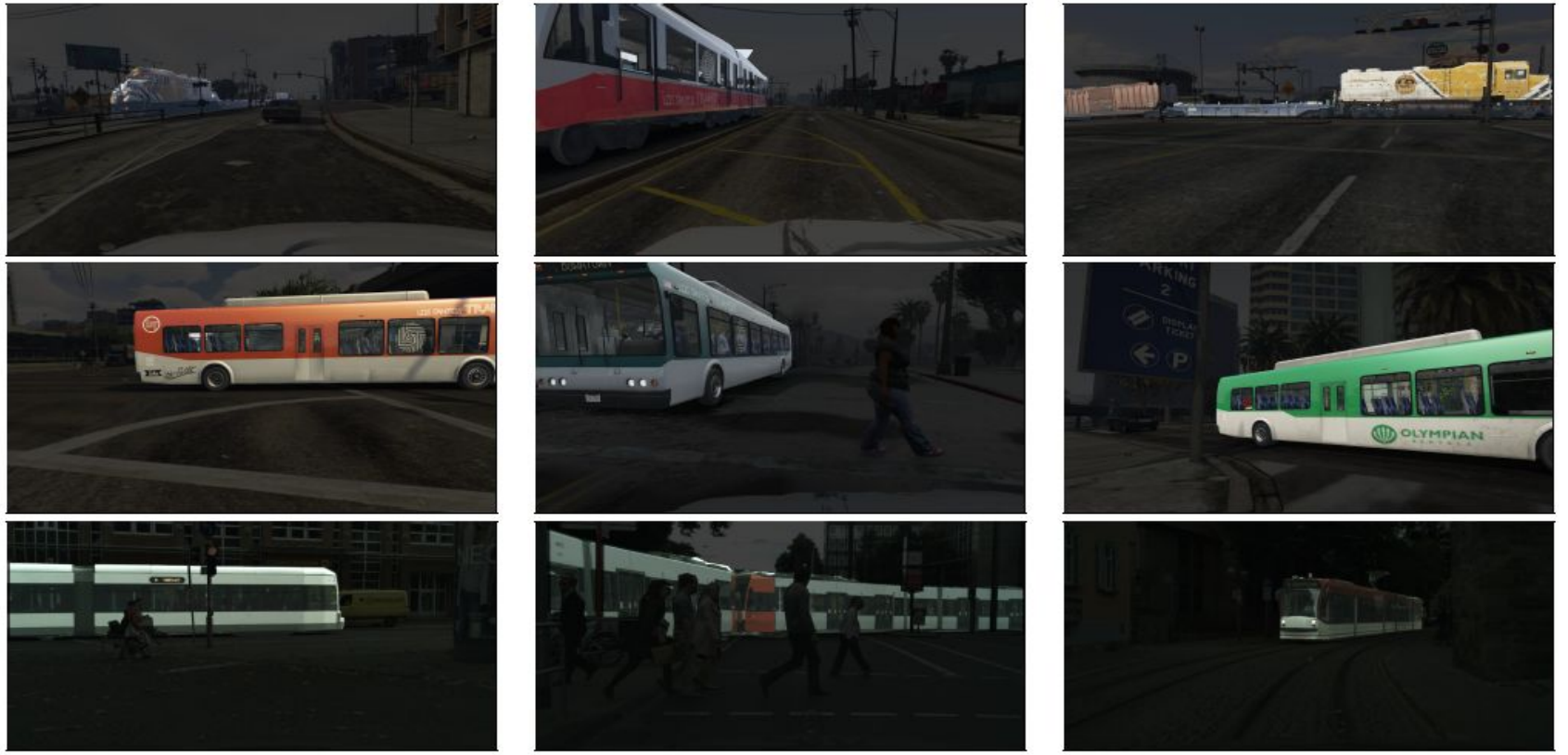
[Images from Dou et al., PnP-AdaNet: Plug-and-play adversarial domain adaptation network with a benchmark at cross-modality cardiac segmentation ArXiv 2018]

Bad generalization to the target



[Images from Dou et al., PnP-AdaNet: Plug-and-play adversarial domain adaptation network with a benchmark at cross-modality cardiac segmentation ArXiv 2018]

A lot of interest in computer vision as well:
Domain shifts are *everywhere* **BUT** we cannot label *everything*



"train"
GTA

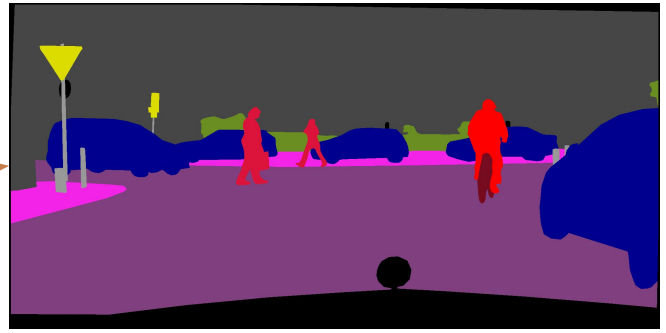
"bus"
GTA

"train"
Cityscapes

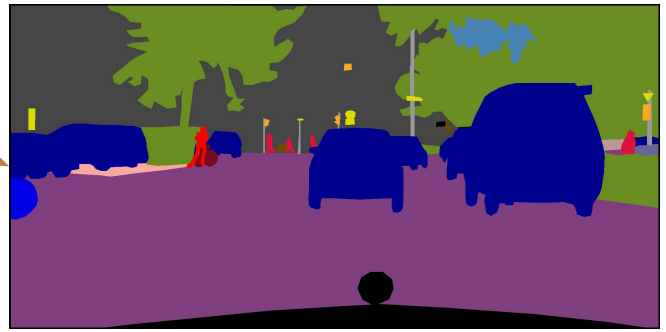
Figures from [Zhang et al., A Curriculum Domain Adaptation Approach to the Semantic Segmentation of Urban Scenes TPAMI 2019]

A lot of interest in computer vision as well:

Domain shifts are *everywhere* **BUT** we cannot label *everything*



Frankfurt

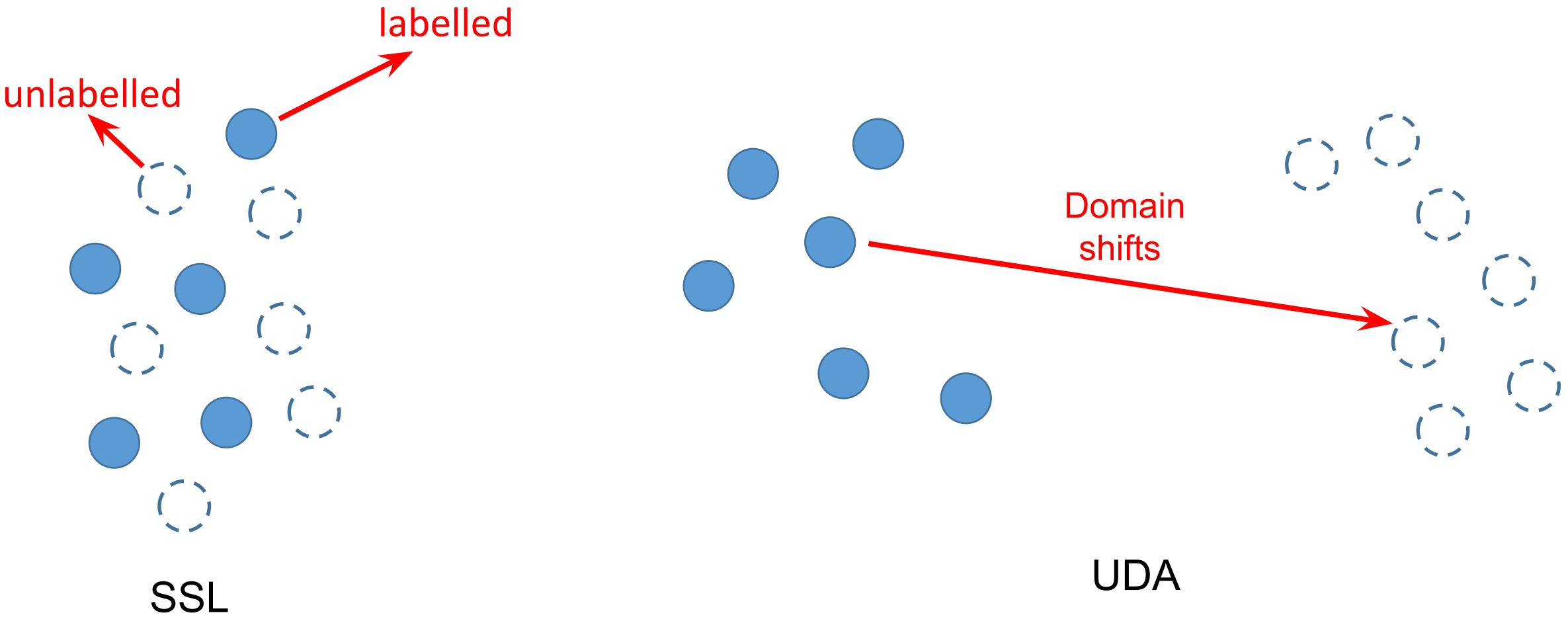


Zurich

road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrain
sky	person	rider	car	truck	bus	train	motorcycle	bicycle	unlabeled

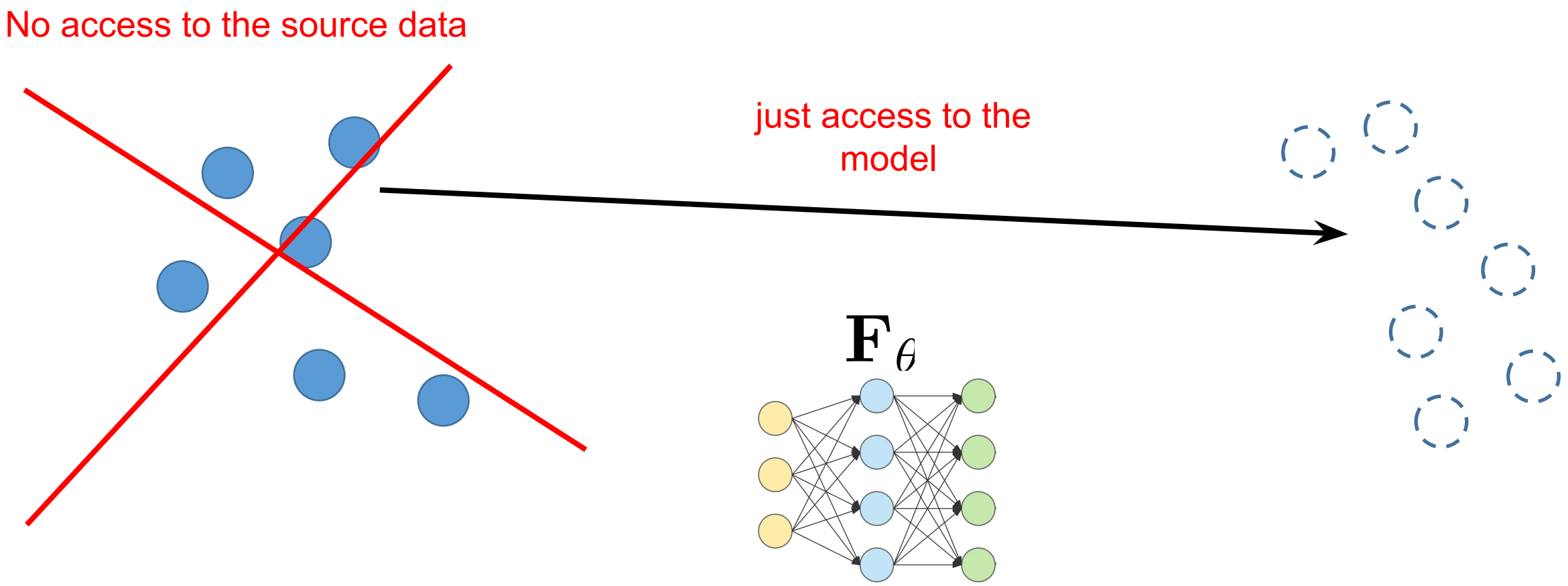
Cityscapes (5000 images): labeling of 1 image takes 90 min at average [Cordt et al., CVPR 2016]

UDA = SSL + domain shift



Test Time Adaption (TTA) or Source-Free Domain Adaptation (SFDA)

UDA *without* access to the source data



Adaption vs. Training from scratch

Shared resources **private resources**

- Efficient **re-use** of large models

- Started in NLP

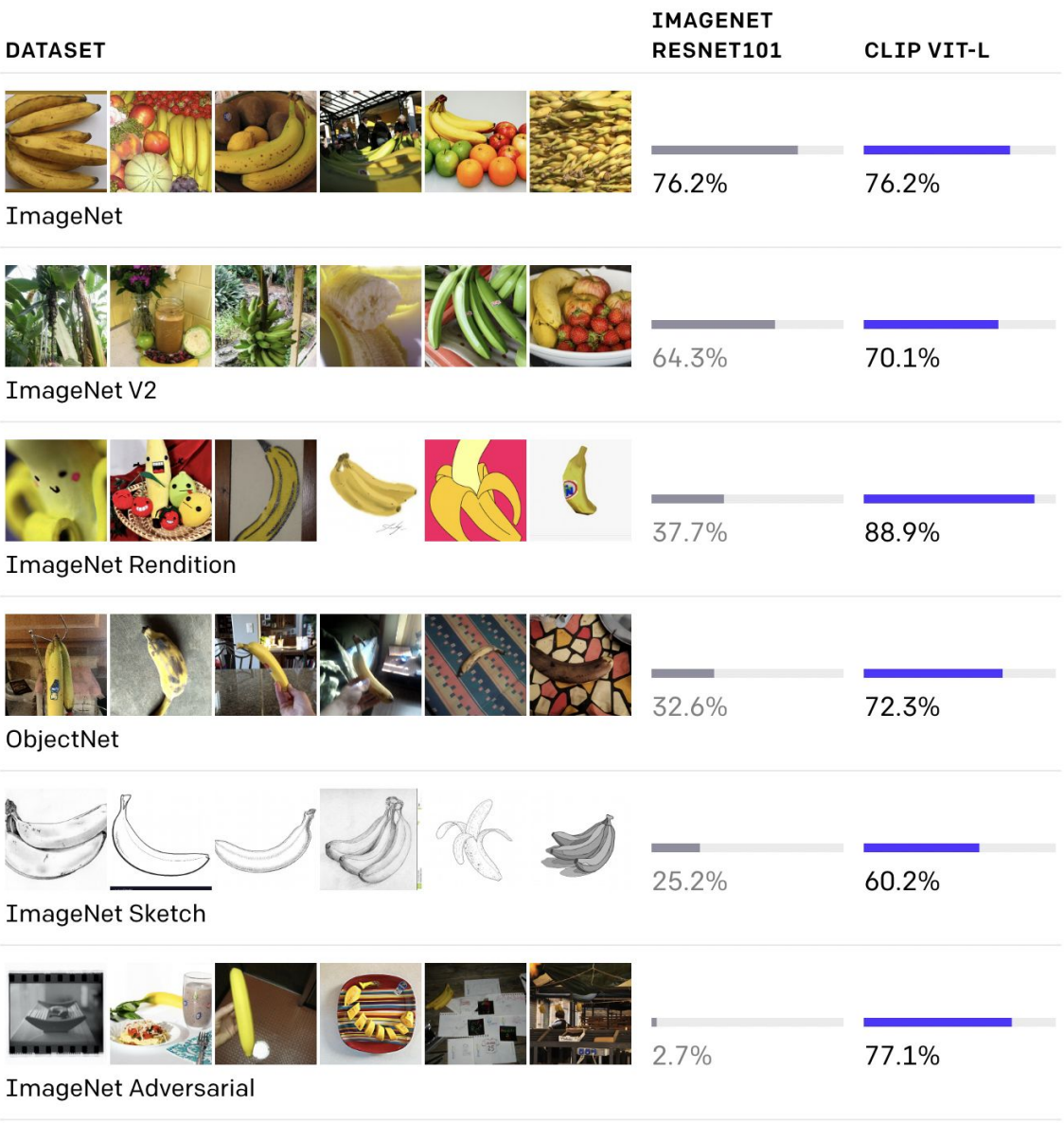
  - BERT (2018): **340M** parameters, trained on then English Wikipedia 2,5 B word
  - GPT-3 (2020): **175B** parameters, trained on hundreds of billions of words

- Gradually coming to computer vision

  - CLIP (2021): **151M** parameters, trained on 400 M (text, image) pairs
  - DALL-E 2 (2022): **12B** parameters

An example of the benefits of large models: CLIP

[Radford et al., ICML 2021]



Desirable properties of a test-time adapter

- Model-agnostic

 so that the progress in architectures could be leveraged easily

- Robust to hyper-parameters

 Ideally the same across different adaptation scenarios and tasks

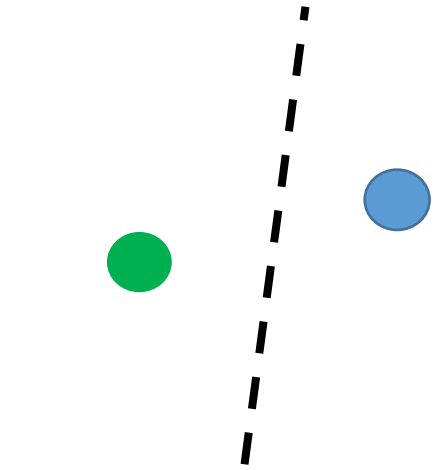
- Lightweight

Few-Shot/SSL/UDA/TTA in a nutshell: Leveraging **unlabelled** data with **priors**

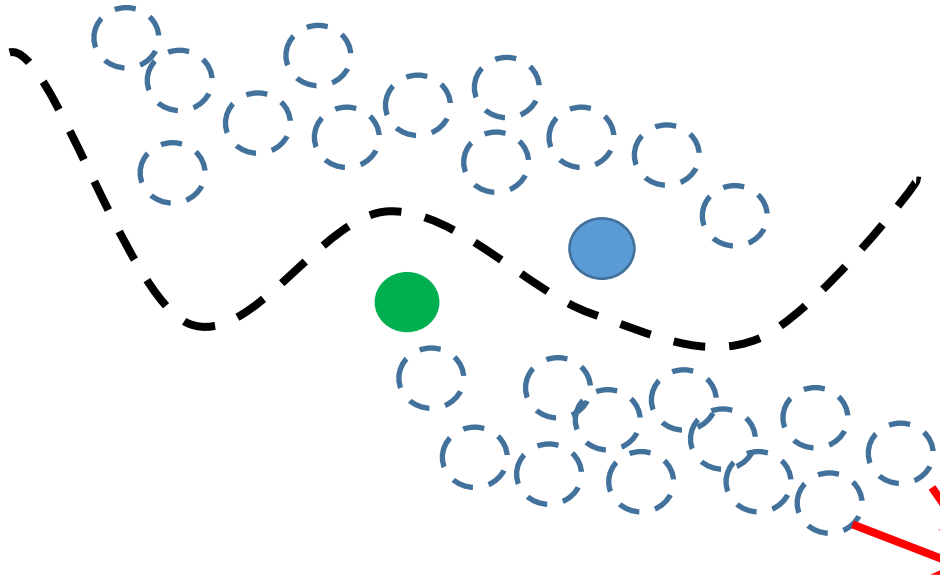
- Structure-driven priors: Regularization
- Knowledge-driven priors (e.g., anatomical constraints)
- Invariance priors (e.g., contrastive learning)
- Multi-modal priors (e.g., text info associated with the images)

Laplacian regularization

In SSL: Learning from both labeled and unlabeled data

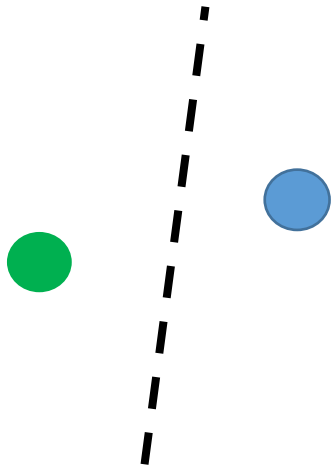


Only labeled points

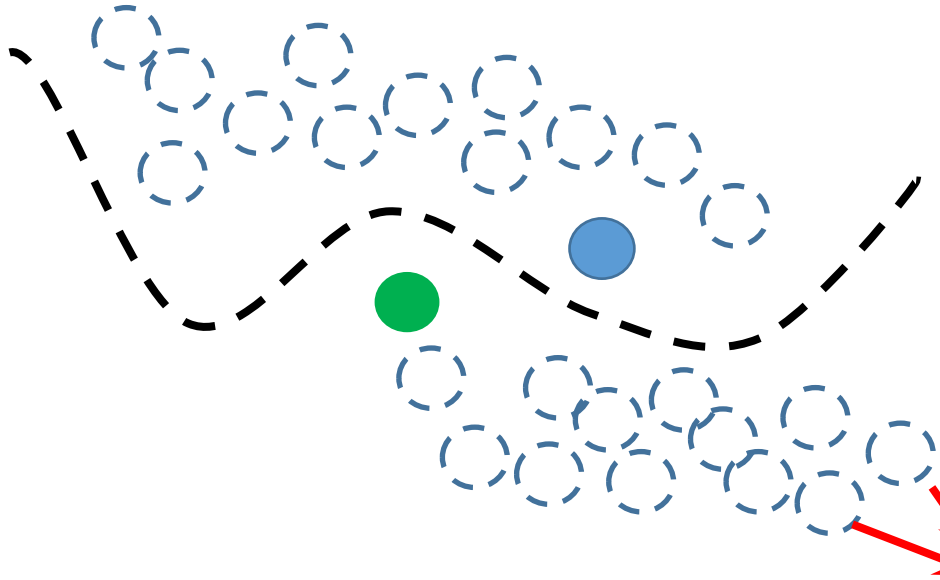


Unlabeled points

In SSL: Learning from both labeled and unlabeled data



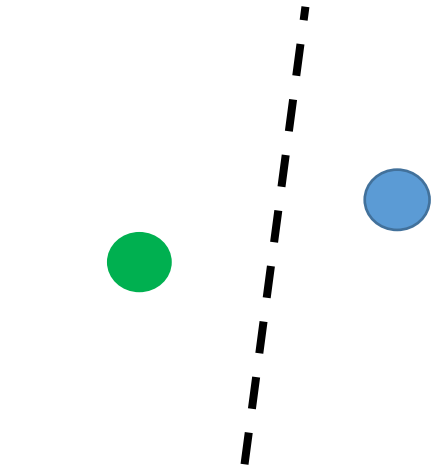
Only labeled points



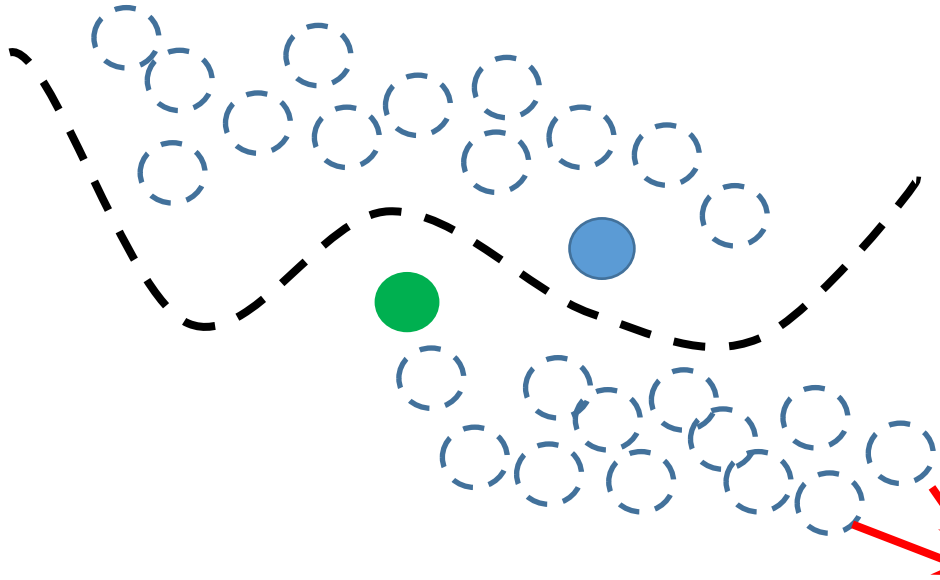
Unlabeled points

$$\sum_{p \in \mathcal{L}} l(\mathbf{f}_\theta^p, \mathbf{y}^p) + \sum_{p,q \in \mathcal{L} \cup \mathcal{U}} g(\mathbf{f}_\theta^p, \mathbf{f}_\theta^q, w_{p,q})$$

In SSL: Learning from both labeled and unlabeled data



Only labeled points



Unlabeled points

$$\sum_{p \in \mathcal{L}} l(\mathbf{f}_\theta^p, \mathbf{y}^p) + \sum_{p,q \in \mathcal{L} \cup \mathcal{U}} g(\mathbf{f}_\theta^p, \mathbf{f}_\theta^q, w_{p,q})$$

Labeled points

Unlabeled points

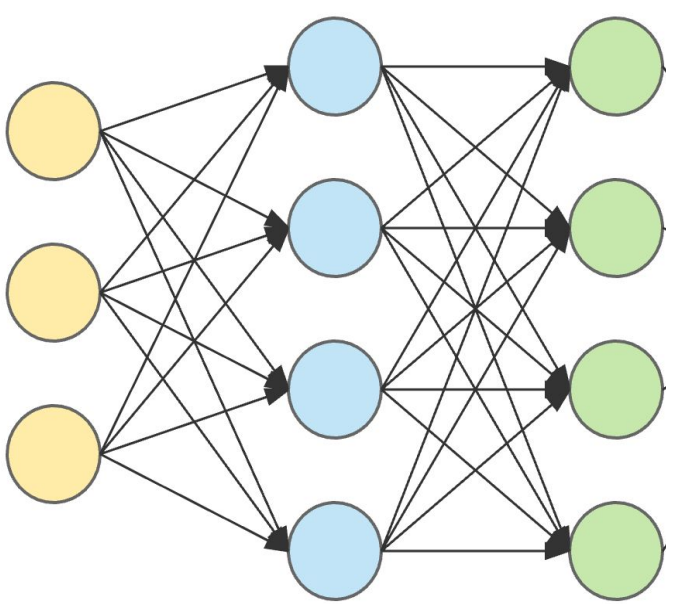
In SSL: Learning from both labeled and unlabeled data

$$\sum_{p \in \mathcal{L}} l(\mathbf{f}_\theta^p, \mathbf{y}^p) + \sum_{p,q \in \mathcal{L} \cup \mathcal{U}} g(\mathbf{f}_\theta^p, \mathbf{f}_\theta^q, w_{p,q})$$

Labels

e.g.: cross-entropy

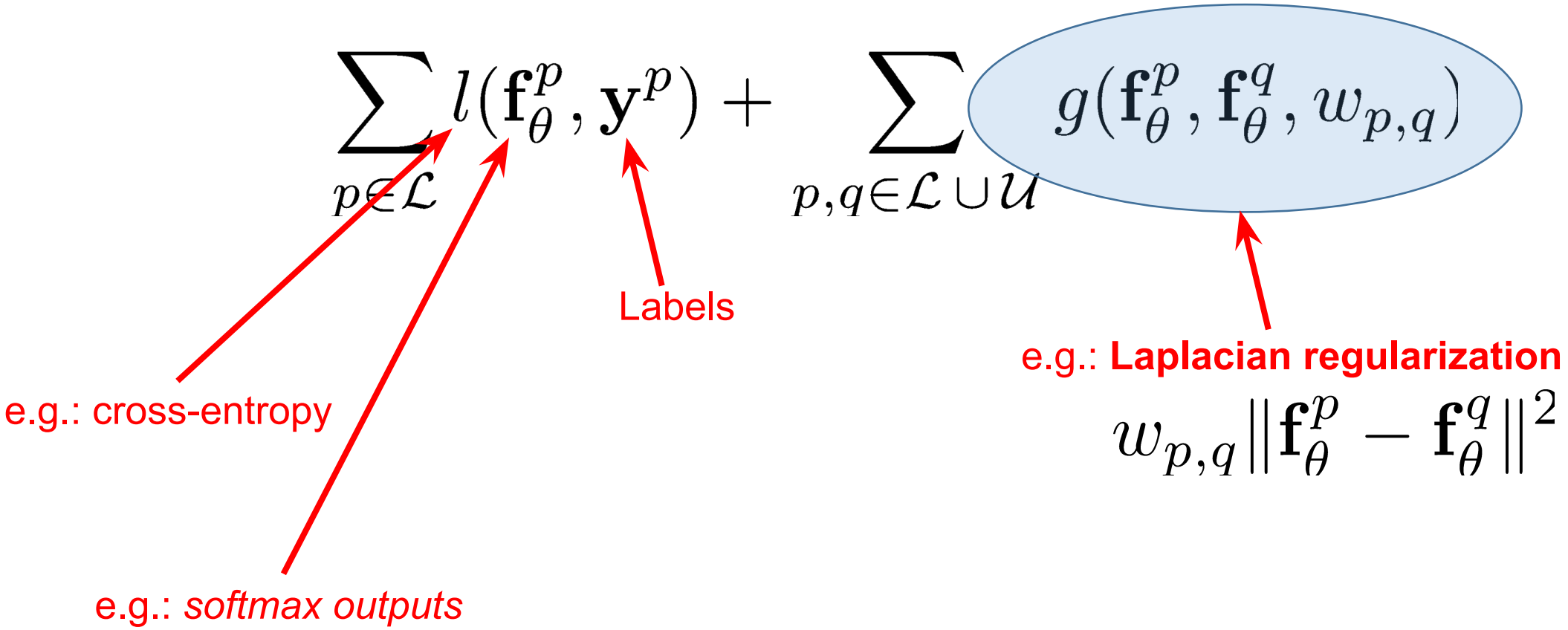
e.g.: *softmax outputs*



$$f_\theta^{p,1}$$
$$f_\theta^{p,2}$$
$$\vdots$$
$$f_\theta^{p,C}$$

$$\mathbf{f}_\theta^p = \mathbf{s}_\theta^p \in [0,1]^K$$

In SSL: Learning from both labeled and unlabeled data



Laplacian regularization: Standard in classical SSL

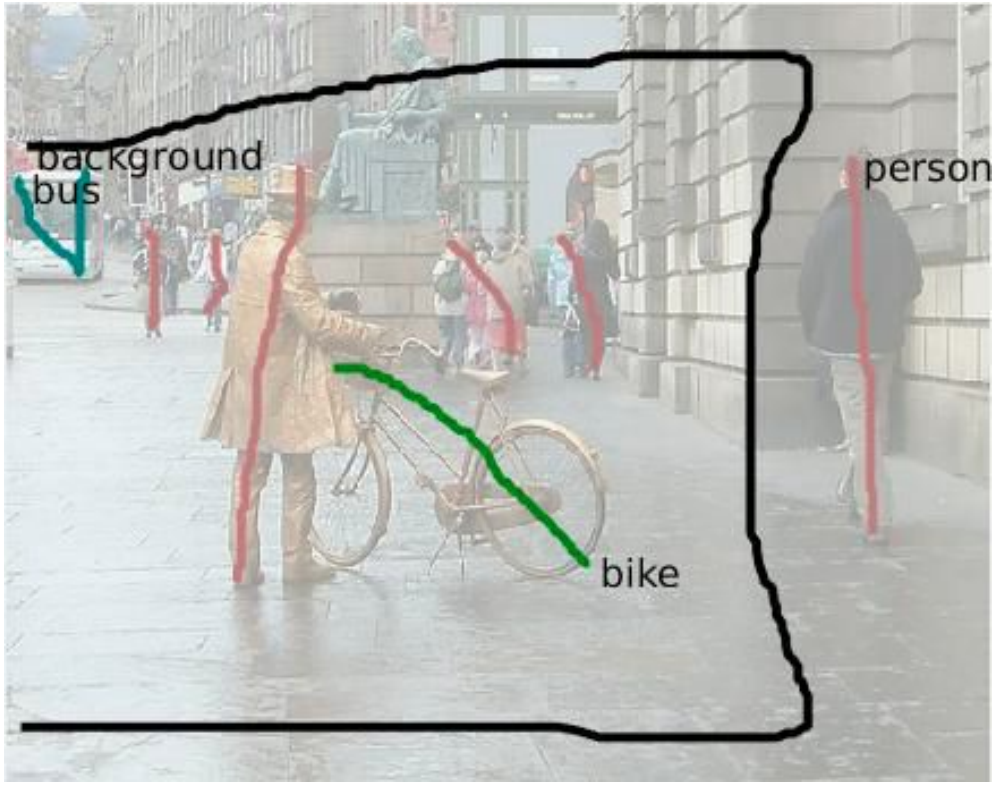
$$\sum_{p \in \mathcal{L}} l(\mathbf{f}_\theta^p, \mathbf{y}^p) + \sum_{p,q \in \mathcal{L} \cup \mathcal{U}} g(\mathbf{f}_\theta^p, \mathbf{f}_\theta^q, w_{p,q})$$

$$w_{p,q} \| \mathbf{f}_\theta^p - \mathbf{f}_\theta^q \|^2$$

- [Weston et al., Deep Learning via semi-supervised embedding, ICML 2008]
- [Belkin et al., Manifold regularization: a geometric framework for learning from Labeled and Unlabeled Examples, JMLR 2006]
- [Zhu et al., Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions, ICML 2003]

Example: Semi-supervision loss for segmentation

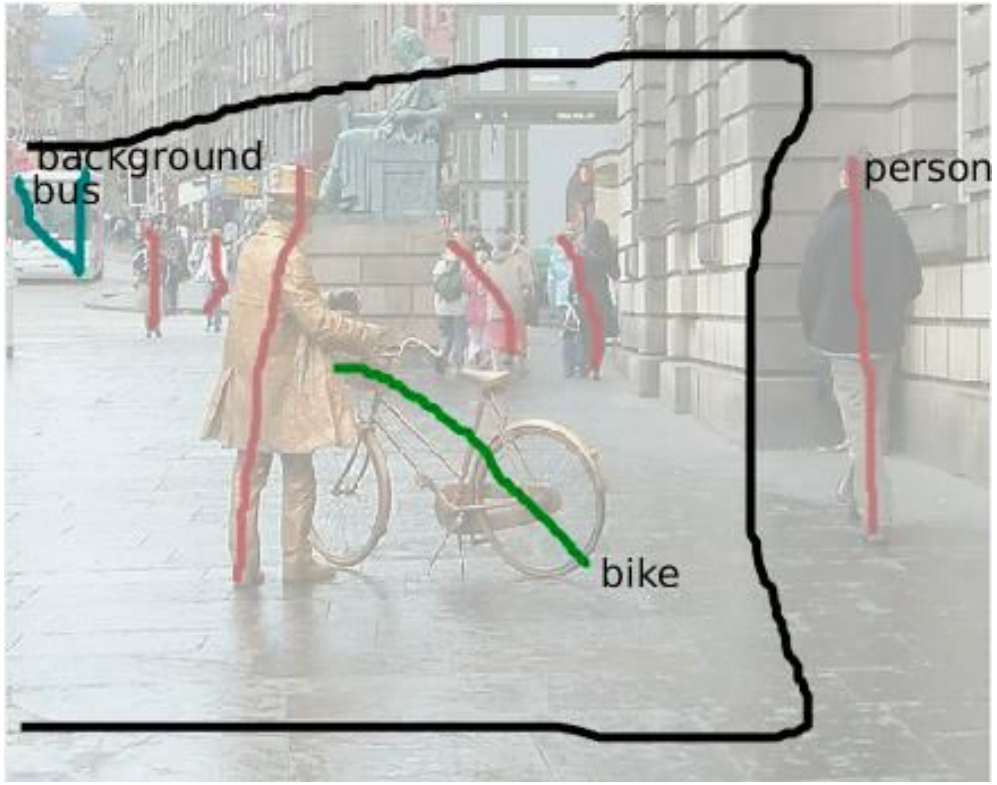
$$\min_{\theta} \sum_{p \in \mathcal{L}} l(\mathbf{y}^p, \mathbf{s}_\theta^p) + \sum_{p,q \in \mathcal{L} \cup \mathcal{U}} w_{p,q} \|\mathbf{s}_\theta^p - \mathbf{s}_\theta^q\|^2$$



[Tang et al., On regularized losses for weakly supervised segmentation, ECCV 2018]

Example: Semi-supervision loss for segmentation

$$\min_{\theta} \sum_{p \in \mathcal{L}} l(\mathbf{y}^p, \mathbf{s}_\theta^p) + \sum_{p,q \in \mathcal{L} \cup \mathcal{U}} w_{p,q} \|\mathbf{s}_\theta^p - \mathbf{s}_\theta^q\|^2$$

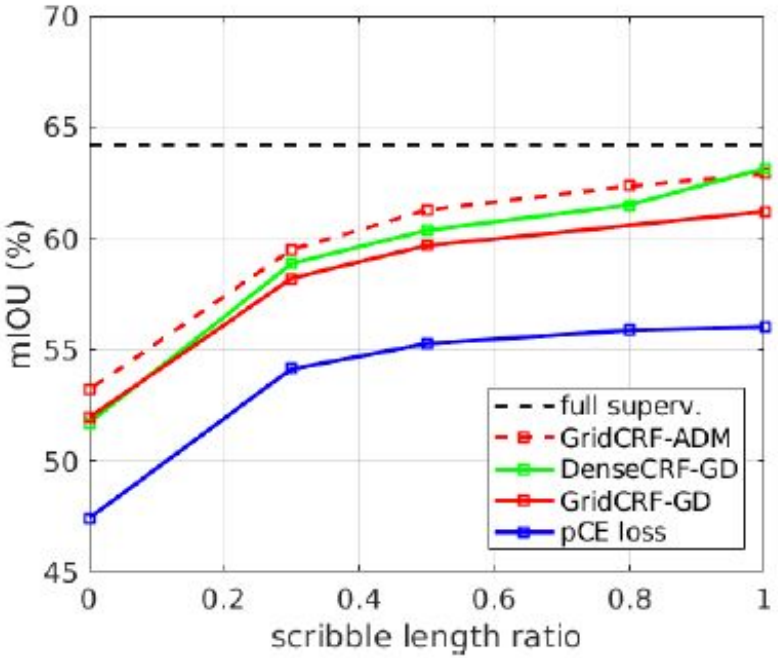


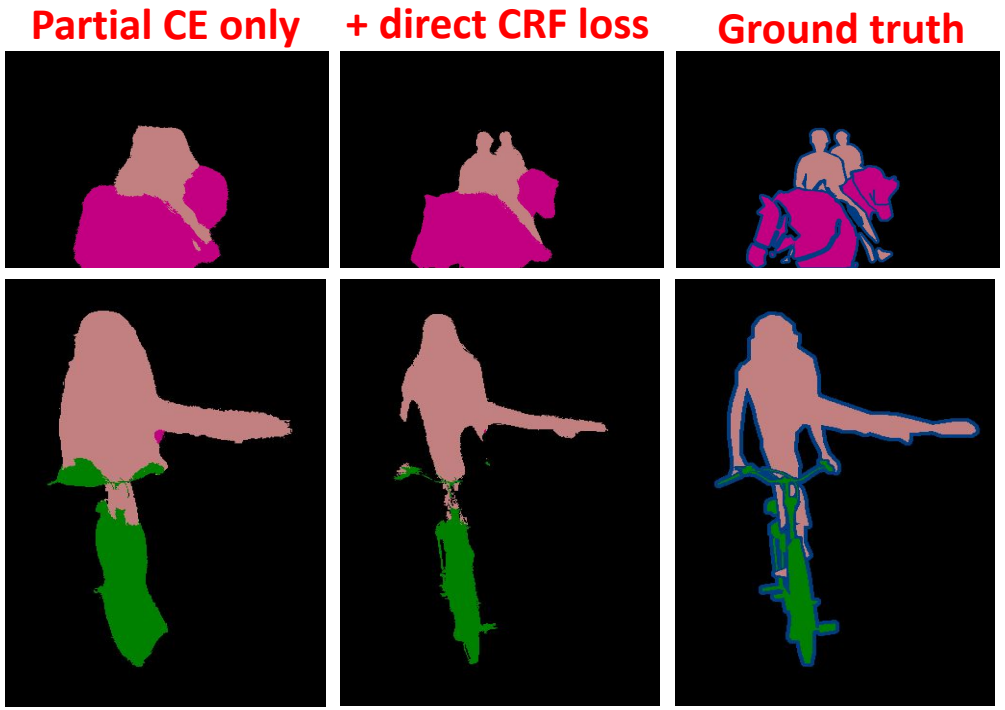
On the vertices of the simplex (binary variables), this is exactly the popular Potts model in CRFs

[Tang et al., On regularized losses for weakly supervised segmentation, ECCV 2018]

Example: Semi-supervision loss for segmentation

$$\min_{\theta} \sum_{p \in \mathcal{L}} l(\mathbf{y}^p, \mathbf{s}_\theta^p) + \sum_{p,q \in \mathcal{L} \cup \mathcal{U}} w_{p,q} \|\mathbf{s}_\theta^p - \mathbf{s}_\theta^q\|^2$$

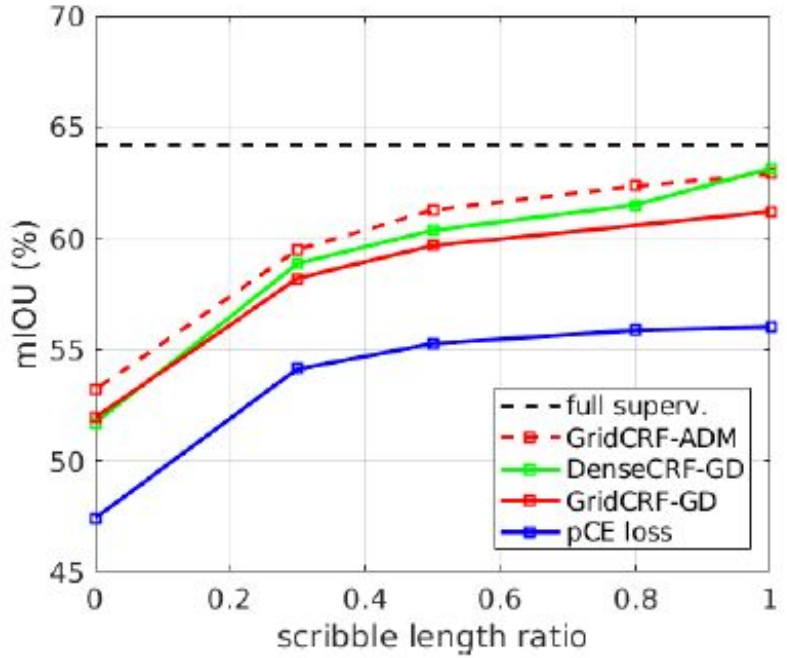




[Tang et al., On regularized losses for weakly supervised segmentation, ECCV 2018]
[Marin et al., Beyond gradient descent for regularized segmentation losses, CVPR 2019]

Example: Semi-supervision loss for segmentation

$$\min_{\theta} \sum_{p \in \mathcal{L}} l(\mathbf{y}^p, \mathbf{s}_\theta^p) + \sum_{p,q \in \mathcal{L} \cup \mathcal{U}} w_{p,q} \|\mathbf{s}_\theta^p - \mathbf{s}_\theta^q\|^2$$



97.6% of full supervision performance with **3%** of the labels!

[Tang et al., On regularized losses for weakly supervised segmentation, ECCV 2018]
[Marin et al., Beyond gradient descent for regularized segmentation losses, CVPR 2019]

Laplacian Regularized Few-Shot Learning

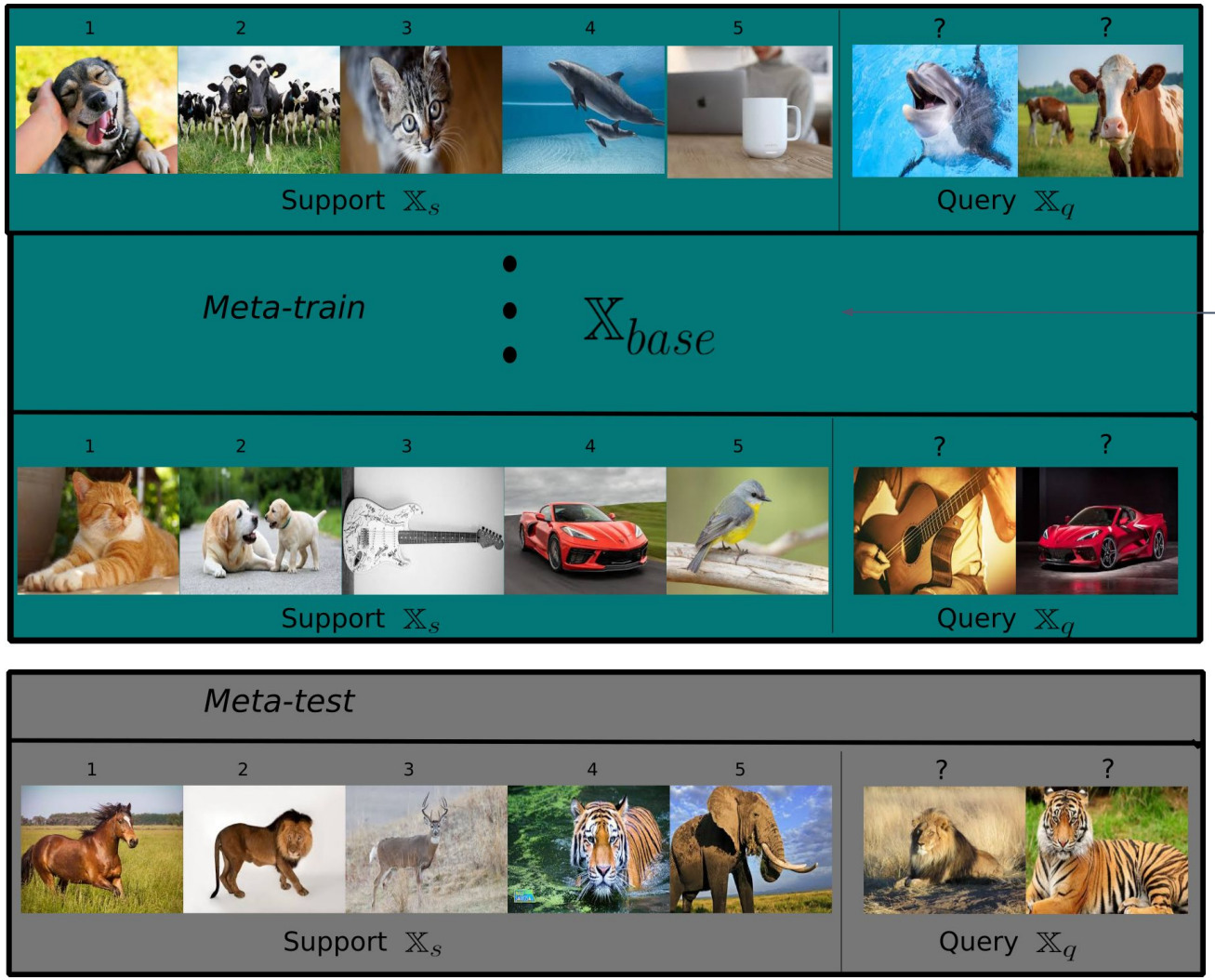
[Ziko et al., ICML 2020]

Few-shot learning

A very large body of recent works, mostly based on:

Meta-learning

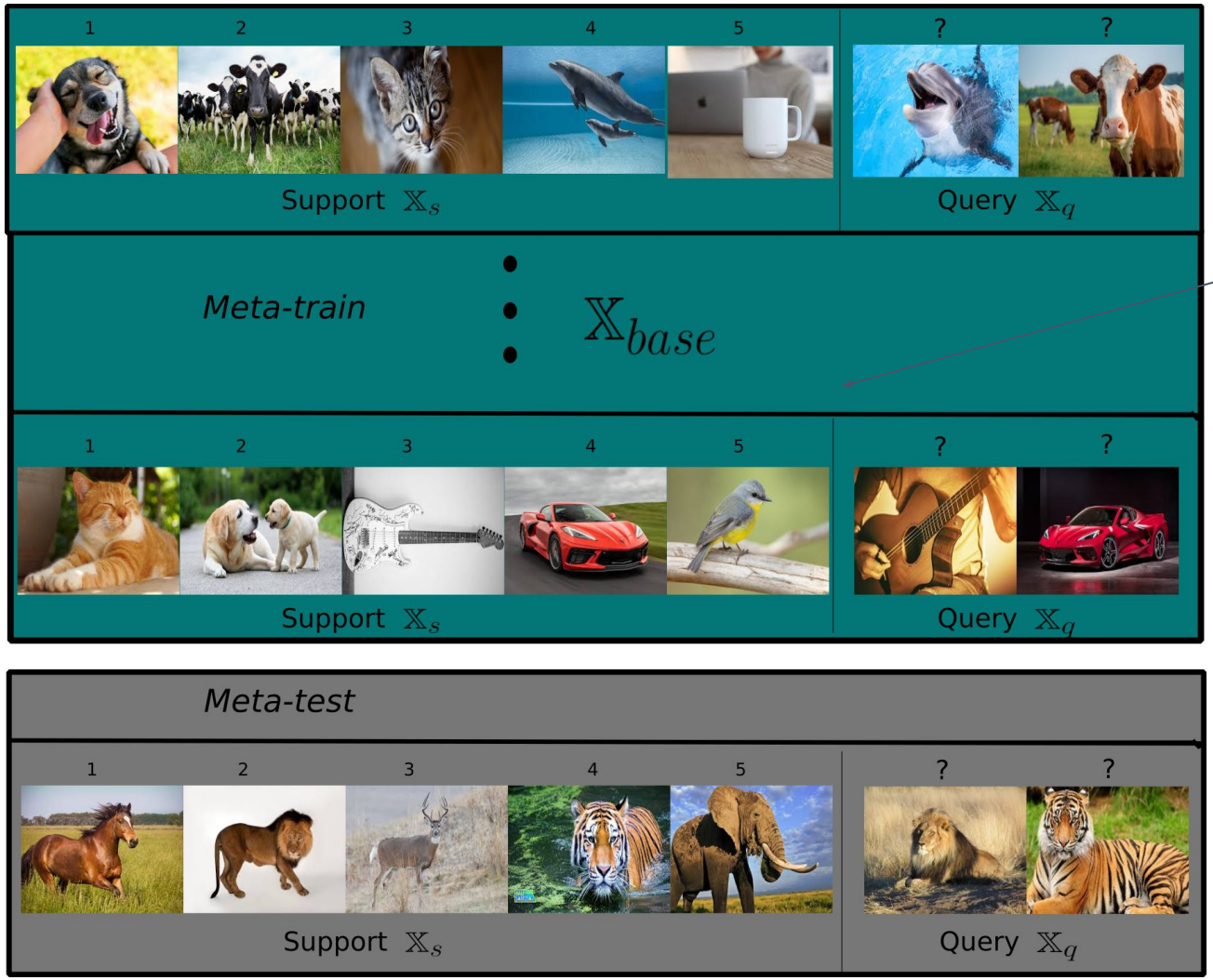
Meta-Learning or “learning-to-learn”



Base training with enough labeled data

(base classes ***different from the*** test classes)

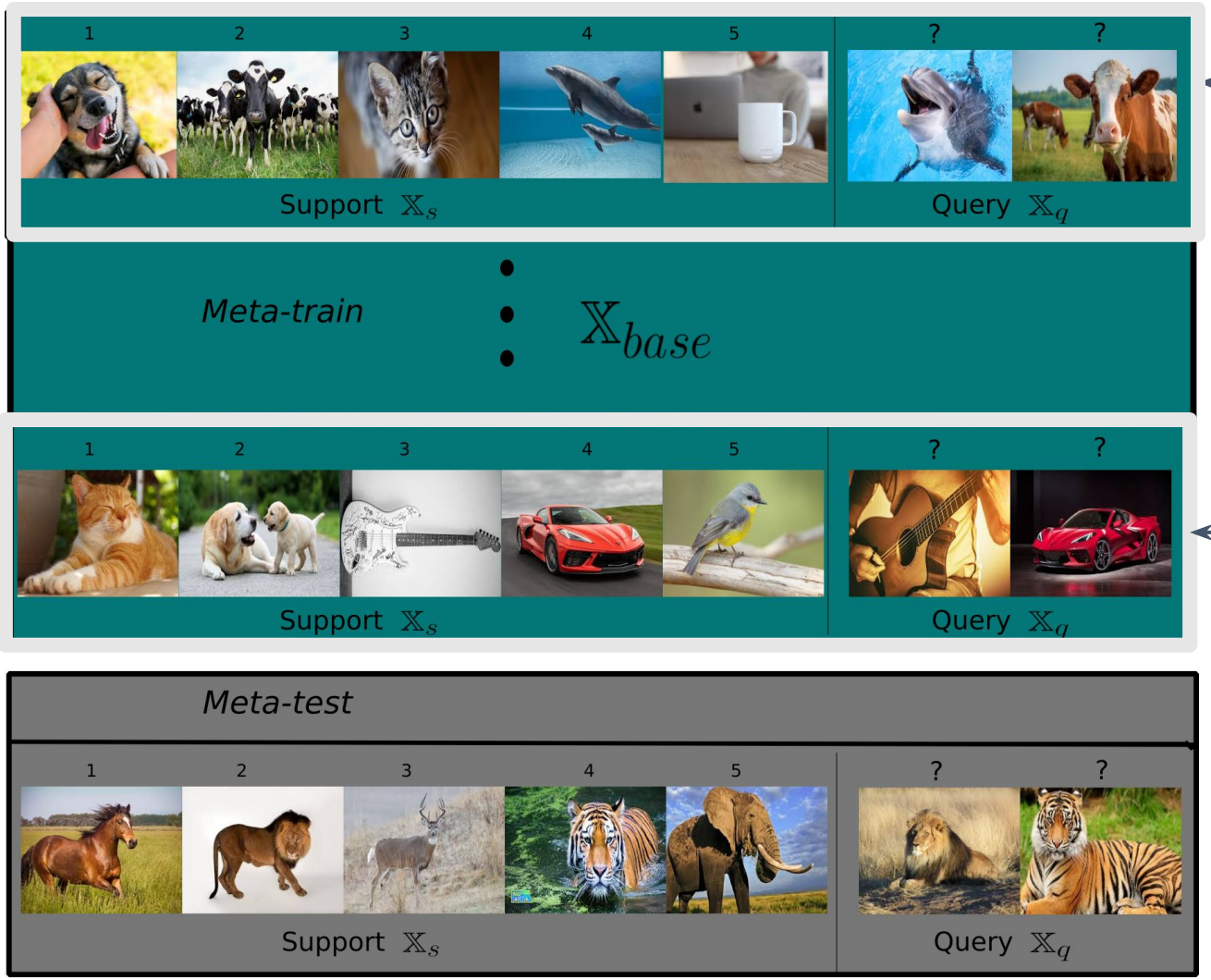
Meta-Learning or “learning-to-learn”



learn initial model

$$\mathbf{F}_\theta$$

Meta-Learning or “learning-to-learn”



Create artificial episodes for ***episodic training***

Vinyal et al, (Neurips ‘16),
Snell et al, (Neurips ‘17),
Sung et al, (CVPR ‘ 18),
Finn et al, (ICML‘ 17),
Ravi et al, (ICLR‘ 17),
Lee et al, (CVPR‘ 19),
Hu et al, (ICLR ‘20),
Ye et al, (CVPR ‘20), . . .

Taking a few steps backwards

Simple baselines and good regularization
outperform convoluted meta-learning approaches

[Chen et al., ICLR'19]; [Tian et al., ECCV'20]; [Veilleux et al., NeurIPS'21];
[Dhillon et al., ICLR'20]; [Ziko et al., ICML'20]; [Boudiaf et al., NeurIPS'20]

Taking a few steps backwards

Simple baselines and good regularization
outperform convoluted meta-learning approaches

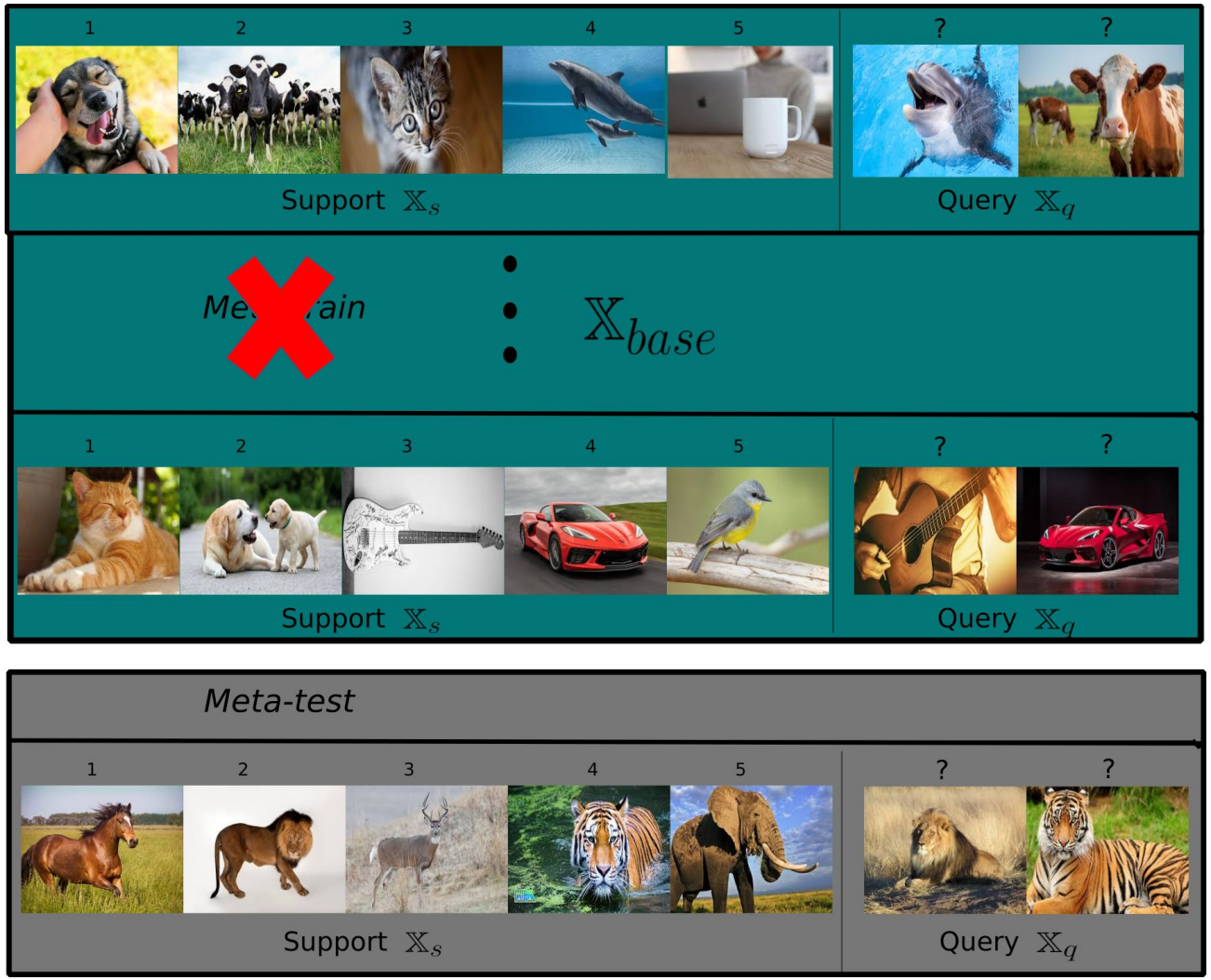
[Chen et al., ICLR'19]; [Tian et al., ECCV'20]; [Veilleux et al., NeurIPS'21];
[Dhillon et al., ICLR'20]; [Ziko et al., ICML'20]; [Boudiaf et al., NeurIPS'20]

Entropy regularization

Laplacian regularization

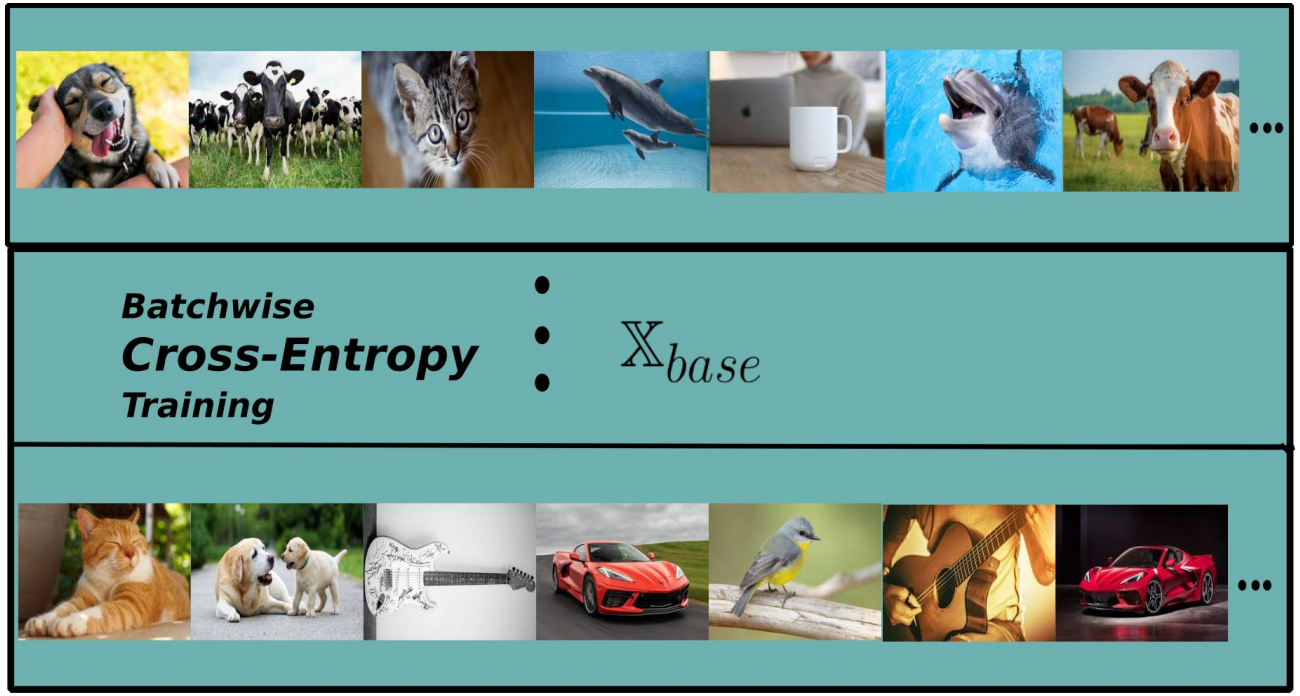
Mutual-information regularization

Baseline Framework



No need to
meta-train

Baseline Framework

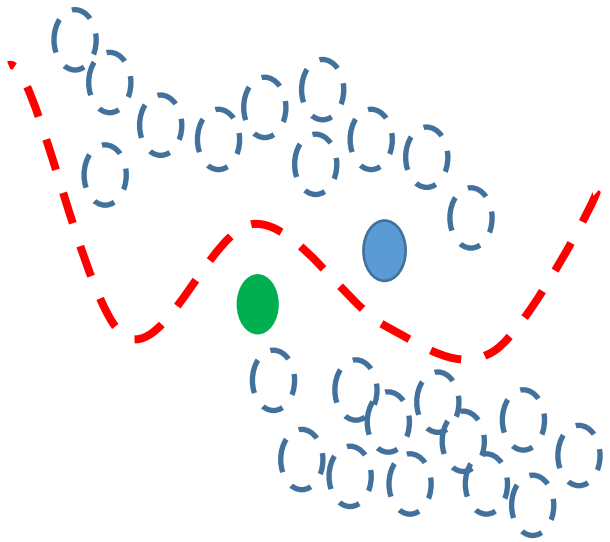


Conventional training

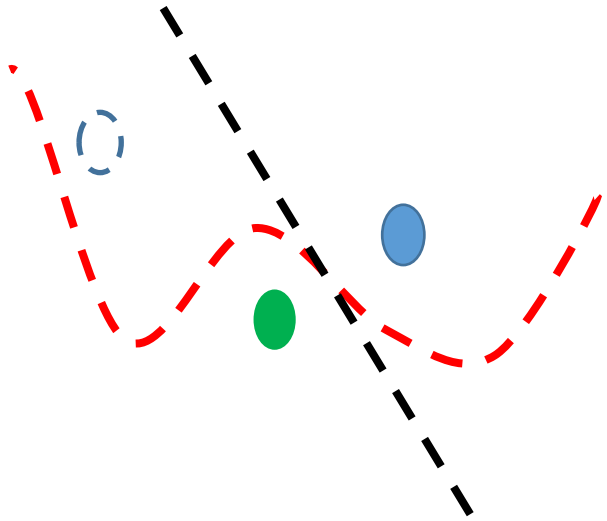
Types of test-time prediction:

Induction vs. Transduction

Transductive vs. inductive test-time prediction

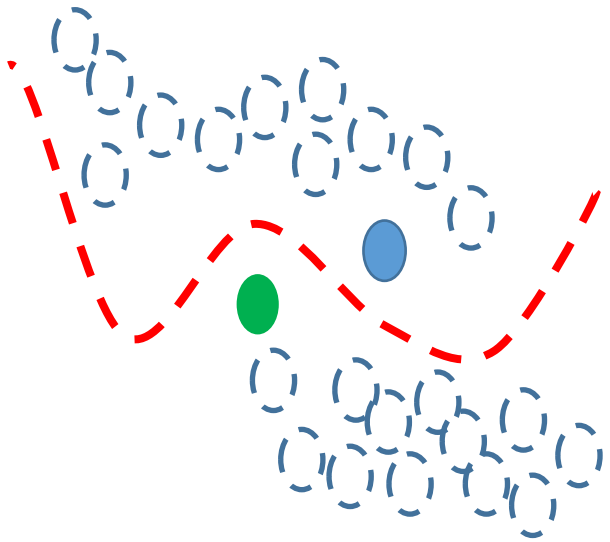


Transduction
(joint test-time prediction)



Induction
(one test sample at a time)

Transductive vs. inductive test-time prediction



Transduction
(joint test-time prediction)

Transduction is popular in few-shot learning, e.g.,

[Finn et al., ICML'17]
(MAML: Trans. batchnorm)

[Dhillon et al., ICLR'20]
(Entropy Min)

[Ziko et al., ICML'20]
(Laplacian Regularization)

[Boudiaf et al., NeurIPS'20]
(Information Maximization)

Why transduction is a relevant and promising avenue for adaptation?

- Tons of practical scenarios that allow transduction

 image segmentation, videos streams, smart-device photos, autonomous vehicles, medical imaging, demographic or financial records, etc.

Boudiaf et al., Few-shot segmentation without meta-learning: A good transductive inference is all you need? CVPR 2021

Why transduction is a relevant and promising avenue for adaptation?

- Transduction benefits from the statistics of unlabeled test data

- Widely used in classical machine learning: [Vapnik 99]

 e.g., graph regularization & label propagation techniques

Laplacian regularization for test-time prediction

[Ziko et al., ICML'20]

$$E(\mathbf{Z}) = G(\mathbf{Z}) + \frac{\lambda}{2} L(\mathbf{Z})$$

$$G(\mathbf{Z}) = \sum_{p \in \mathcal{U}} \sum_{c=1}^{C} z_{p,c} \|\mathbf{f}_\theta^p - \mu_c\|^2$$

Unknown label assignments for query points

(equal to 1 if sample *p* belongs to class *c* and 0 otherwise)

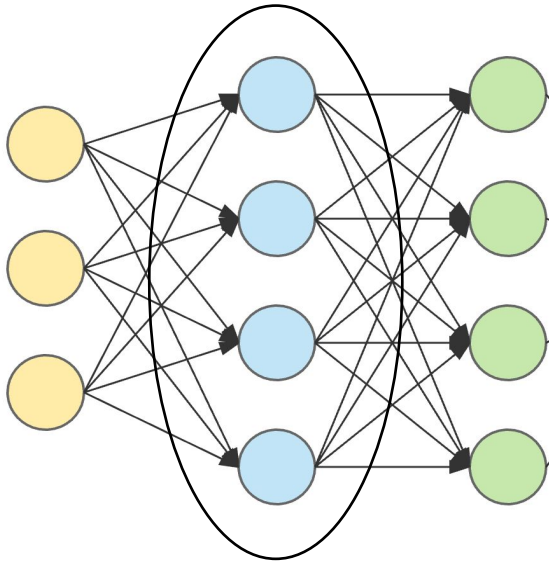
Laplacian regularization for test-time prediction

[Ziko et al., ICML'20]

$$E(\mathbf{Z}) = G(\mathbf{Z}) + \frac{\lambda}{2} L(\mathbf{Z})$$

$$G(\mathbf{Z}) = \sum_{p \in \mathcal{U}} \sum_{c=1}^{C} z_{p,c} \| \mathbf{f}_\theta^p - \mu_c \|^2$$

Feature embedding
(fixed, no re-training)



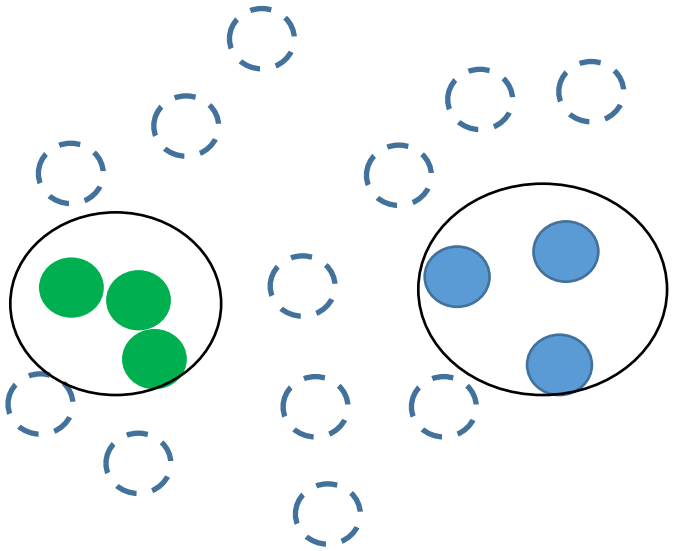
Laplacian regularization for test-time prediction

[Ziko et al., ICML'20]

$$E(\mathbf{Z}) = G(\mathbf{Z}) + \frac{\lambda}{2} L(\mathbf{Z})$$

$$G(\mathbf{Z}) = \sum_{p \in \mathcal{U}} \sum_{c=1}^{C} z_{p,c} \| \mathbf{f}_\theta^p - \mu_c \|^2$$

mean of the support samples in class *c*



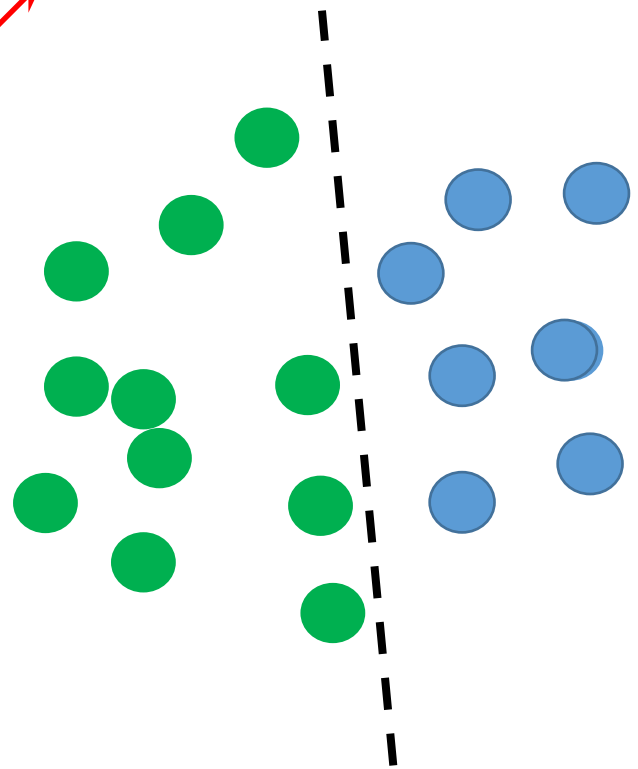
Laplacian regularization for test-time prediction

[Ziko et al., ICML'20]

$$E(\mathbf{Z}) = G(\mathbf{Z}) + \frac{\lambda}{2} L(\mathbf{Z})$$

$$G(\mathbf{Z}) = \sum_{p \in \mathcal{U}} \sum_{c=1}^{C} z_{p,c} \| \mathbf{f}_\theta^p - \mu_c \|^2$$

Minimizing this term is trivial
(assign a sample to the nearest mean)



Laplacian regularization for test-time prediction

[Ziko et al., ICML'20]

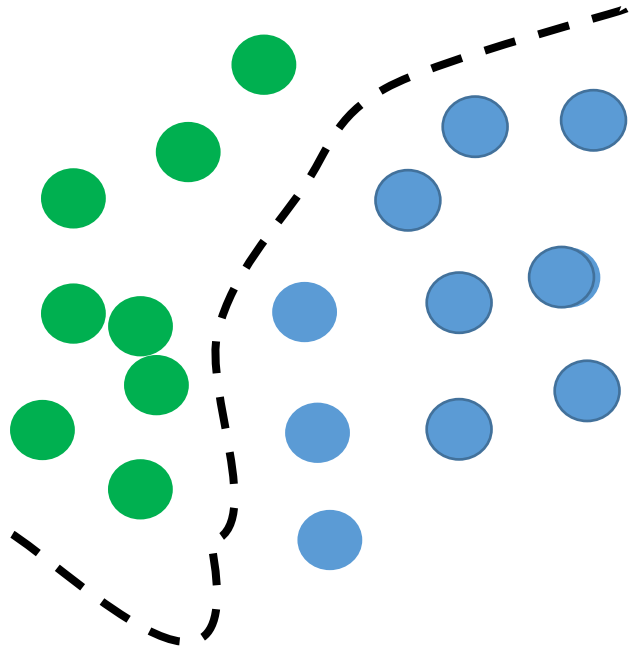
$$E(\mathbf{Z}) = G(\mathbf{Z}) + \frac{\lambda}{2} L(\mathbf{Z})$$

$$L(\mathbf{Z}) = \sum_{p,q \in \mathcal{U}^2} w_{p,q} \left\| \mathbf{z}_p - \mathbf{z}_q \right\|^2$$

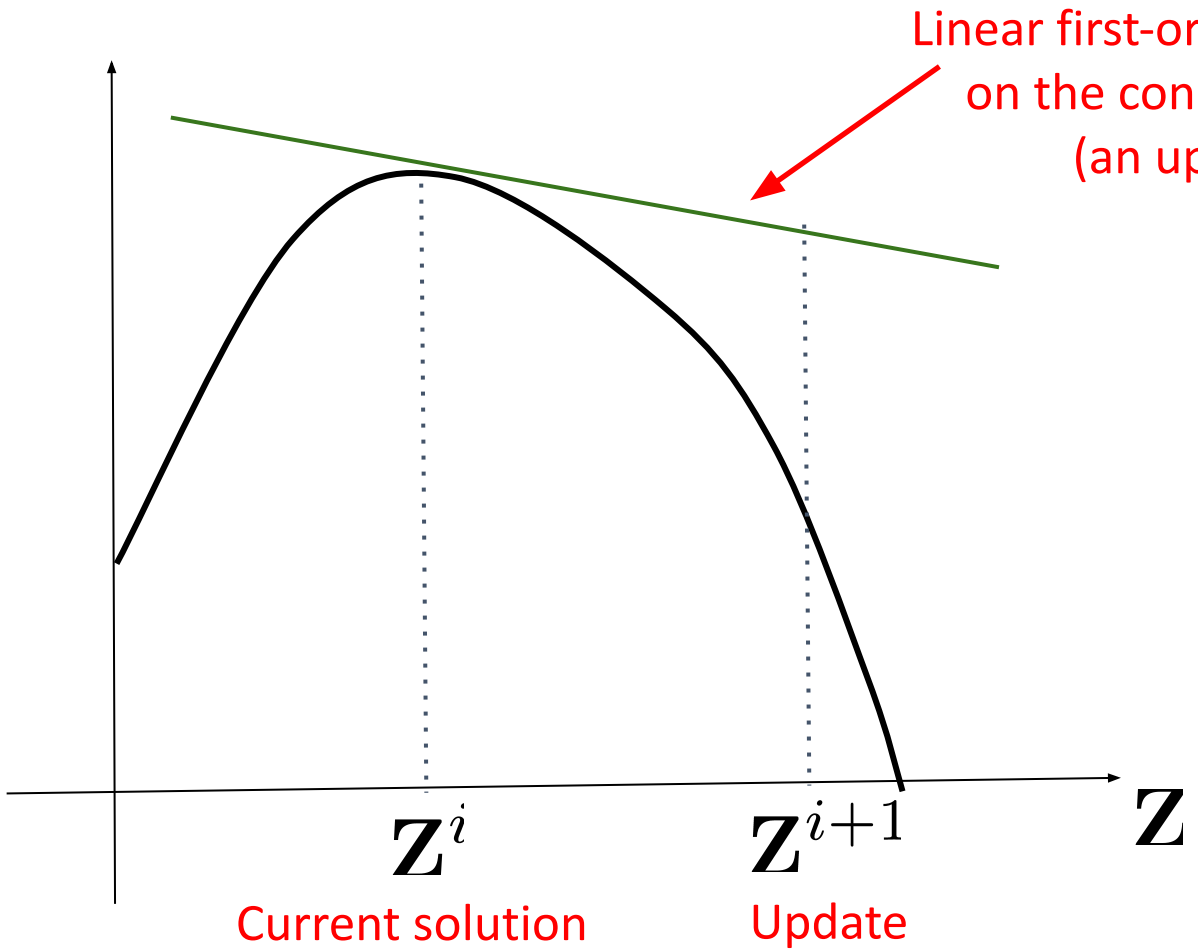
Affinity between feature vectors

Laplacian regularization

(encourages nearby samples to have the same assignments)



Optimization: Concave-Convex procedure



Yuille & Rangarajan, The concave-convex procedure (CCCP), NIPS 2001

The results question an abundant meta-learning literature

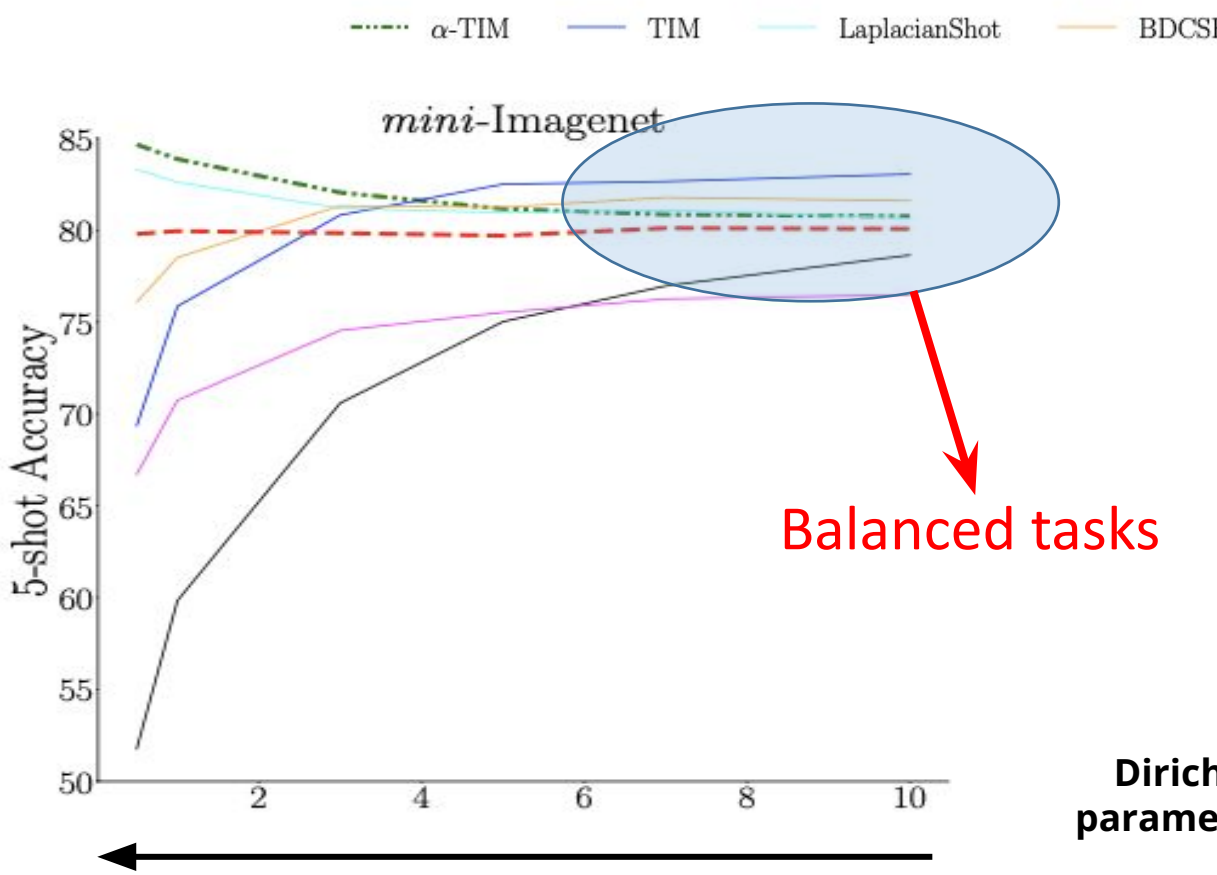
Methods	Network	1-shot	5-shot
MAML [Finn et al., 2017]	ResNet-18	49.61 ± 0.92	65.72 ± 0.77
Chen [Chen et al., 2019]	ResNet-18	51.87 ± 0.77	75.68 ± 0.63
RelationNet [Sung et al., 2018]	ResNet-18	52.48 ± 0.86	69.83 ± 0.68
MatchingNet [Vinyals et al., 2016]	ResNet-18	52.91 ± 0.88	68.88 ± 0.69
ProtoNet [Snell et al., 2017]	ResNet-18	54.16 ± 0.82	73.68 ± 0.65
Gidaris [Gidaris and Komodakis, 2018]	ResNet-15	55.45 ± 0.89	70.13 ± 0.68
SNAIL[Mishra et al., 2018]	ResNet-15	55.71 ± 0.99	68.88 ± 0.92
AdaCNN [Munkhdalai et al., 2018]	ResNet-15	56.88 ± 0.62	71.94 ± 0.57
TADAM [Oreshkin et al., 2018]	ResNet-15	58.50 ± 0.30	76.70 ± 0.30
CAML [Jiang et al., 2019]	ResNet-12	59.23 ± 0.99	72.35 ± 0.71
TPN [Yanbin et al., 2019]	ResNet-12	59.46	75.64
TEAM [Qiao et al., 2019]	ResNet-18	60.07	75.90
MTL [Sun et al., 2019]	ResNet-18	61.20 ± 1.80	75.50 ± 0.80
VariationalFSL [Zhang et al., 2019]	ResNet-18	61.23 ± 0.26	77.69 ± 0.17
Transductive tuning [Dhillon et al., 2020]	ResNet-12	62.35 ± 0.66	74.53 ± 0.54
MetaoptNet[Lee et al., 2019]	ResNet-18	62.64 ± 0.61	78.63 ± 0.46
SimpleShot [Wang et al., 2019]	ResNet-18	63.10 ± 0.20	79.92 ± 0.14
CAN+T [Hou et al., 2019]	ResNet-12	67.19 ± 0.55	80.64 ± 0.35
LaplacianShot (ours)	ResNet-18	**72.11** ± 0.19	**82.31** ± 0.14

Several recent baselines made similar observations:

[Chen et al., ICLR'19]; [Tian et al., ECCV'20]; [Dhillon et al., ICLR'20]; [Boudiaf et al., NeurIPS'20]

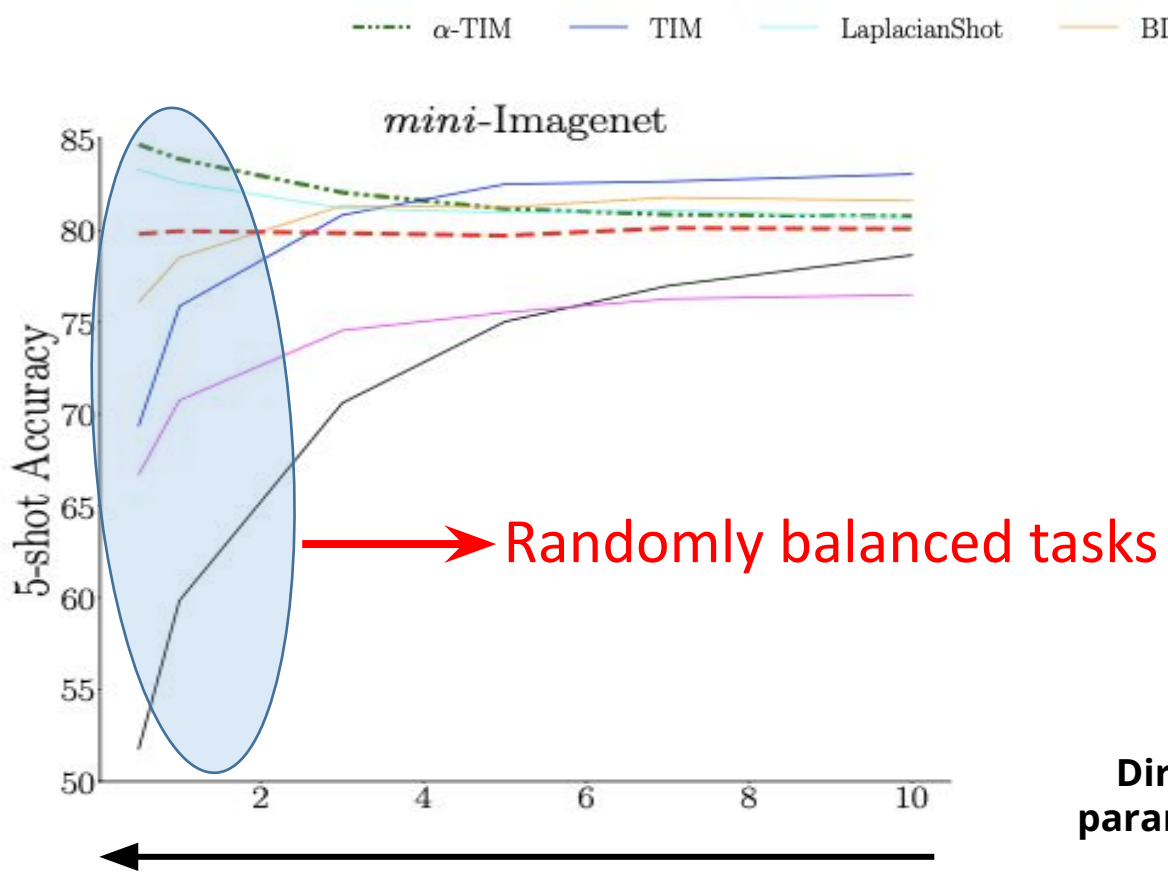
More realistic benchmark: Further surprises

[Veilleux et al., NeurIPS'21]



More realistic benchmark: Further surprises

[Veilleux et al., NeurIPS'21]



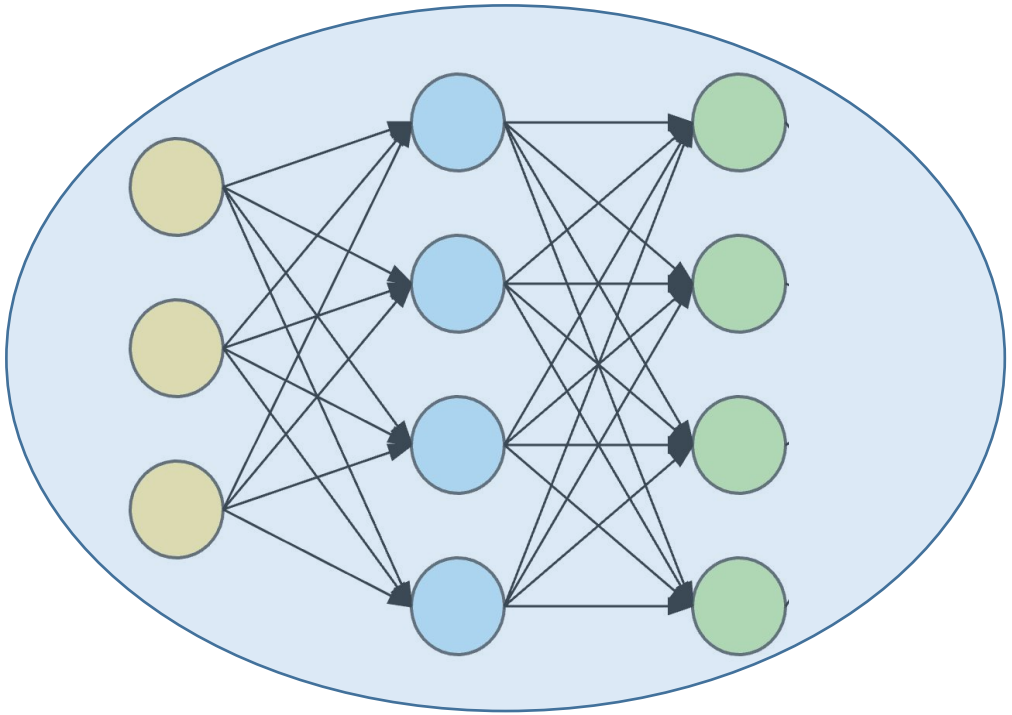
Parameter-free online test-time adaptation

[Boudiaf et al., CVPR'22]

Test-time adaptation re-visited

State-of-the-art TTA methods, e.g., TENT: [Wang et al., ICLR'21]

Network adaptation



Needs different hyper-parameters for each target scenario

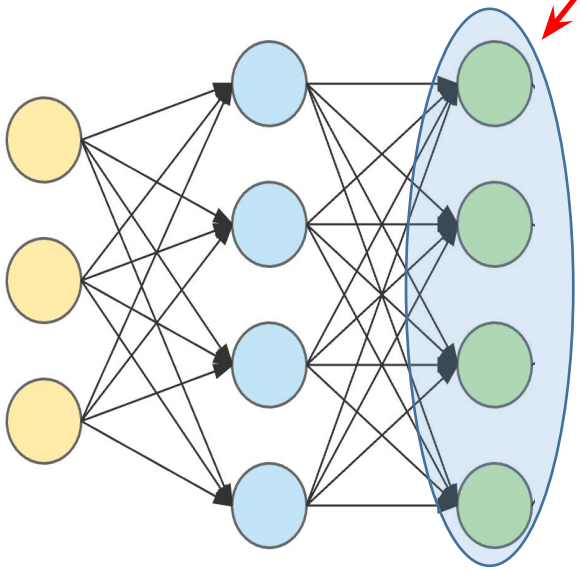
		A	B	C	D	A	B	C	D	A	B	C	D
ImageNet-Val	A	0.3	-0.5	2.8	-1.1	0.1	2.2	4.6	10	2	-2.9	3.7	-2.9
	B	0.2	0.2	0.3	0.3	0.3	2.3	1.8	7.1	0.4	0.4	0.9	0.2
	C	-0.4	-37	3	-66	-3.4	11	3.8	20	4.5	-65	5.1	-64
	D	0.2	0.2	0.3	0.3	0.3	2.3	1.8	7.1	0.4	0.4	0.9	0.2
ImageNet-C	A	0	-14	-0.7	-40	3.9	13	5.6	18	2.3	-52	-0.3	-52
	B	0	-14	-0.7	-40	3.9	13	5.6	18	2.3	-52	-0.3	-52
	C	0	-14	-0.7	-40	3.9	13	5.6	18	2.3	-52	-0.3	-52
	D	-0.4	-37	3	-66	-3.4	11	3.8	20	4.5	-65	5.1	-64
ImageNet-C_{16}	A	-3.1	-15	-0.7	-32	-37	-35	-35	-30	4.6	-61	2	-58
	B	-0.2	-0.6	0.6	-1	-11	-7.9	-8.6	2	1.8	0.8	2.8	0.5
	C	-0.4	-37	2.9	-65	-3.2	10	3.8	20	4.5	-65	5.2	-64
	D	-0.2	-0.6	0.6	-1	-11	-7.9	-8.6	2	1.8	0.8	2.8	0.5
		ImageNet-Val				ImageNet-C				ImageNet-C_{16}			

Evaluation in [Boudiaf et al., CVPR'22]

Regularizing the network outputs behave better:

Laplacian in action, again!

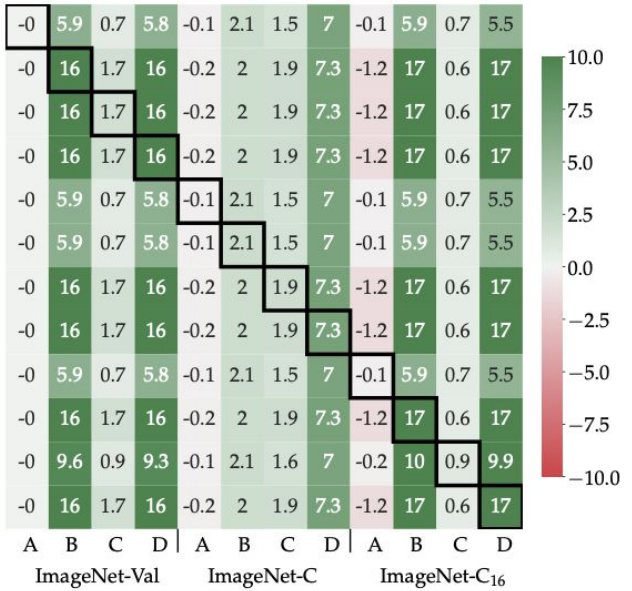
$$\min_{\mathbf{z}} \left\{ -\sum_{p \in \mathcal{U}} \sum_{c=1}^{C} z_{p,c} \log s_\theta^{p,c} + \sum_{p,q \in \mathcal{U}^2} w_{p,q} \|\mathbf{z}_p - \mathbf{z}_q\|^2 \right\}$$



Correcting **only** the network probability outputs

LAME

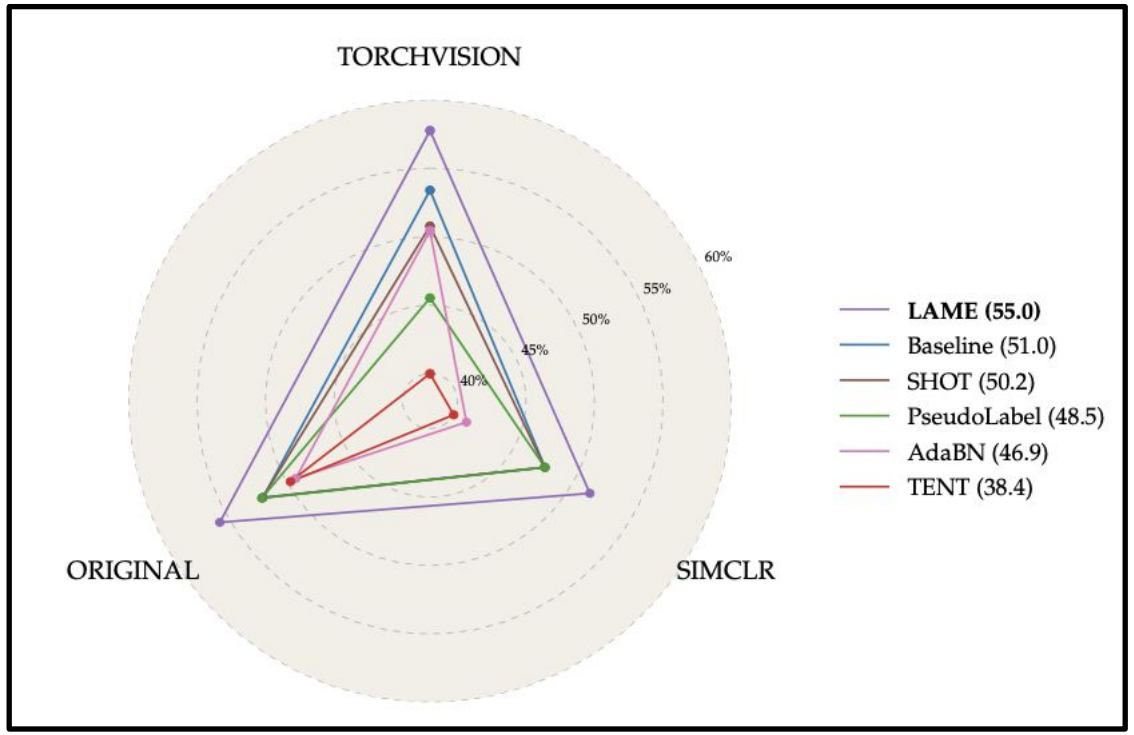
(Laplacian-Adjusted Maximum Likelihood Estimate)

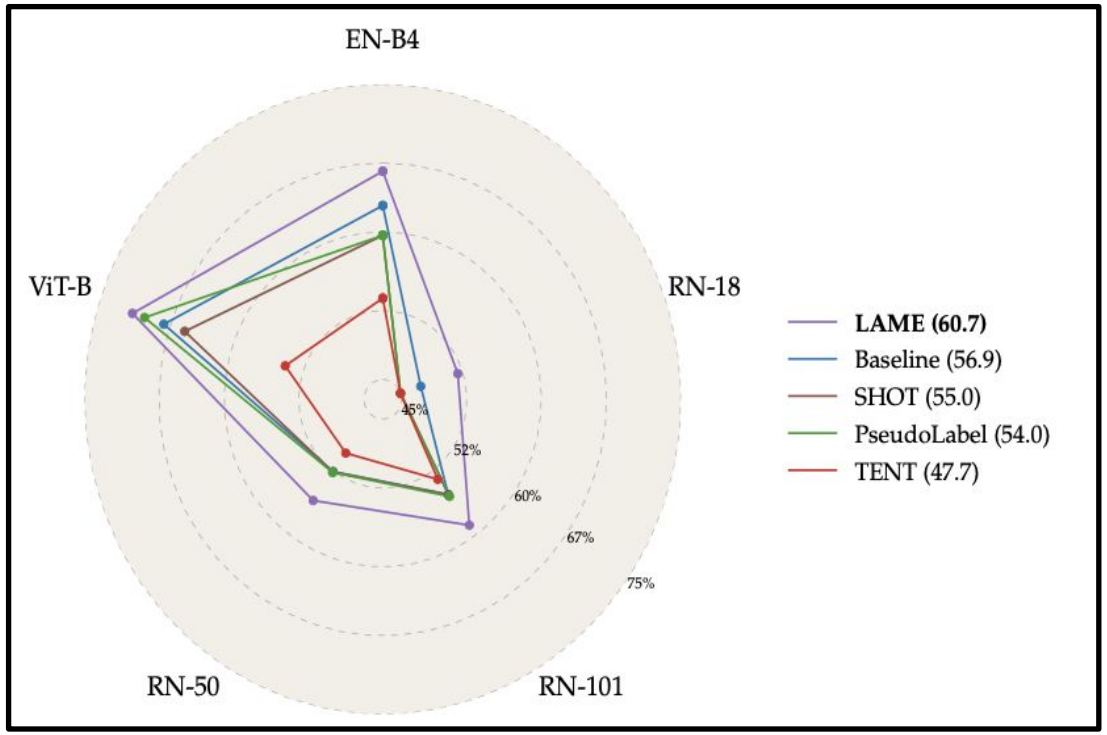


[Boudiaf et al., CVPR'22]

Transferability of hyperparameters across models

[Boudiaf et al., CVPR'22]

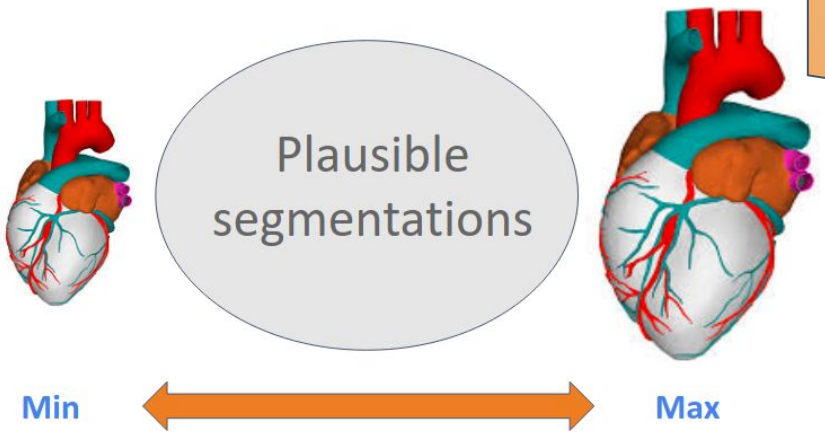




Constrained deep networks

Constrained optimization (in deep networks)

Data meet domain knowledge

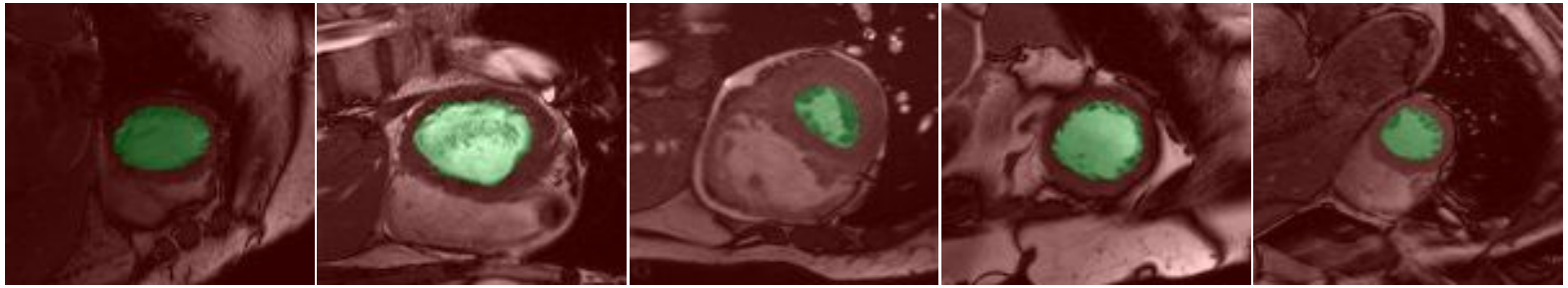


Anatomical priors (e.g., shapes)

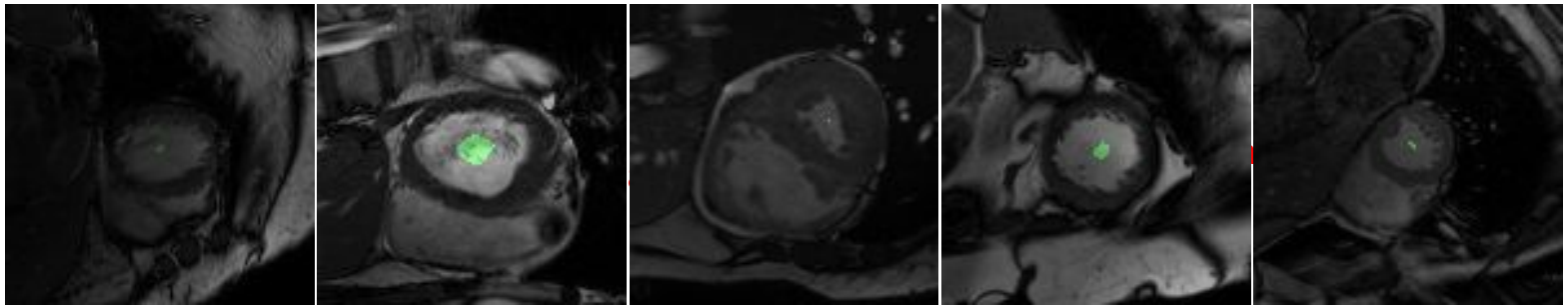
A lot of priors in medical imaging

Partially labeled data
(e.g., exploiting organ relationships)

Example: Left ventricle segmentation in cardiac MRI with volumetric constraints



Full annotations

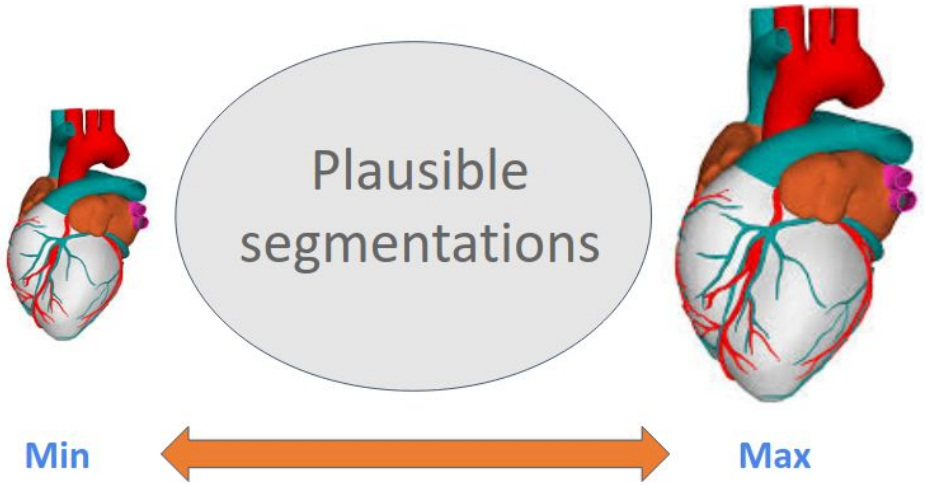


Partial annotations for cross-entropy

Example: Left ventricle segmentation in cardiac MRI with volumetric constraints

[Kervadek et al., MedIA'19]

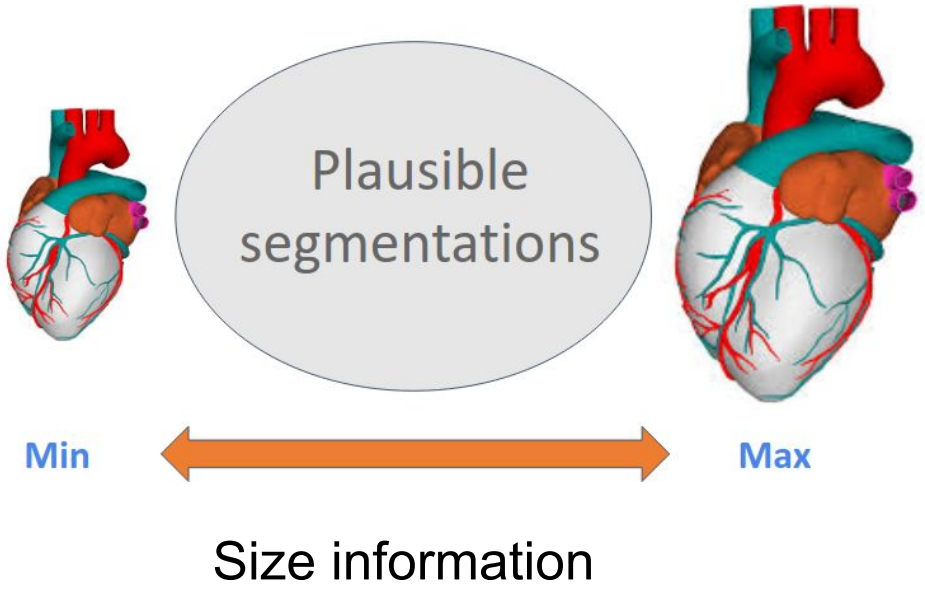
$$\min \leq \sum_{p \in \Omega} s_\theta^{p,c} \leq \max$$

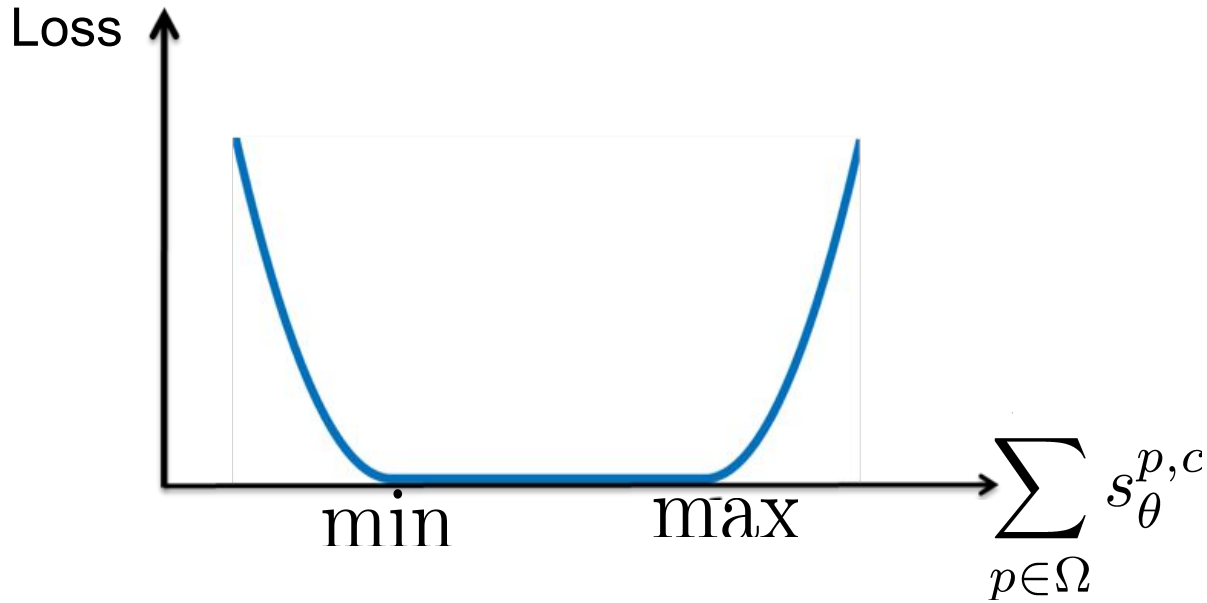


Example: Left ventricle segmentation in cardiac MRI with volumetric constraints

[Kervadek et al., MedIA'19]

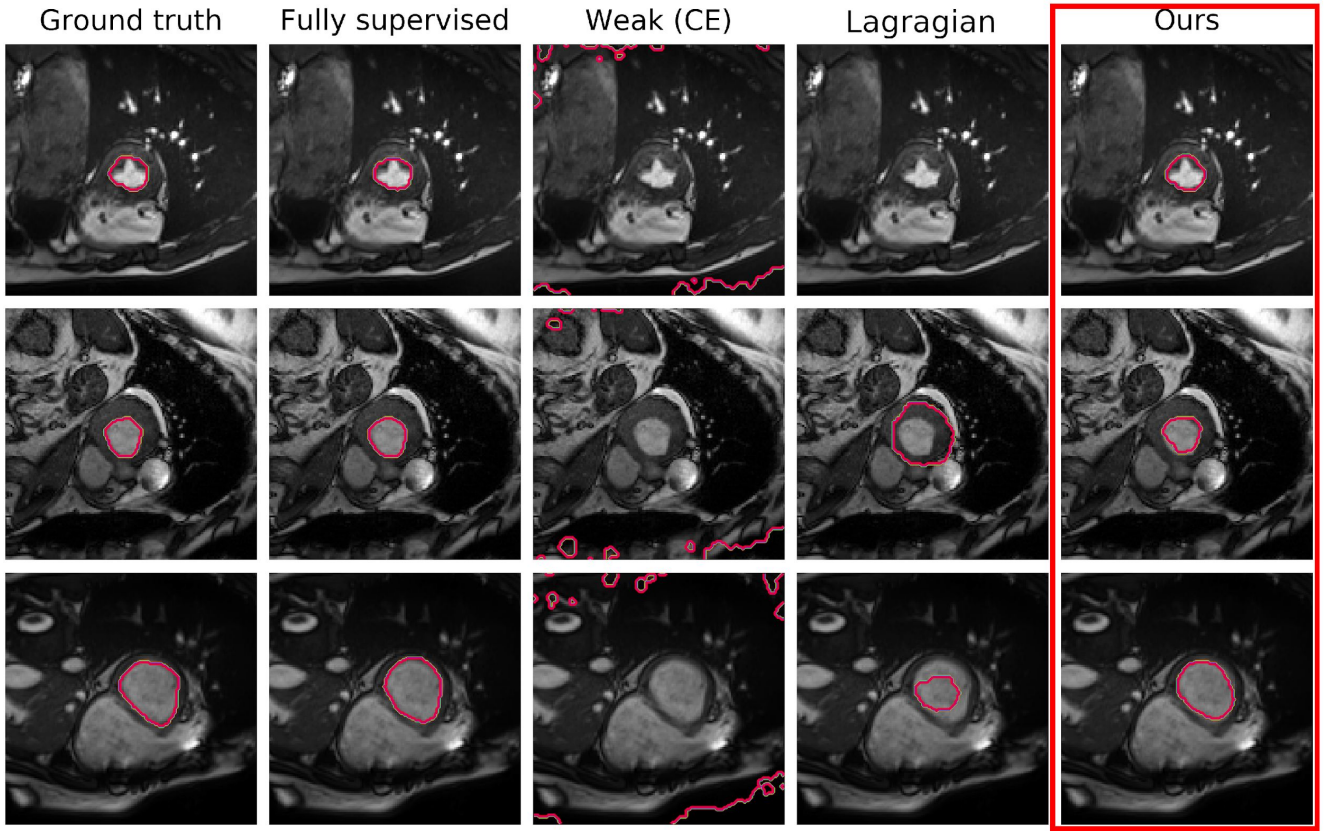
$$\min \leq \sum_{p \in \Omega} s_\theta^{p,c} \leq \max$$





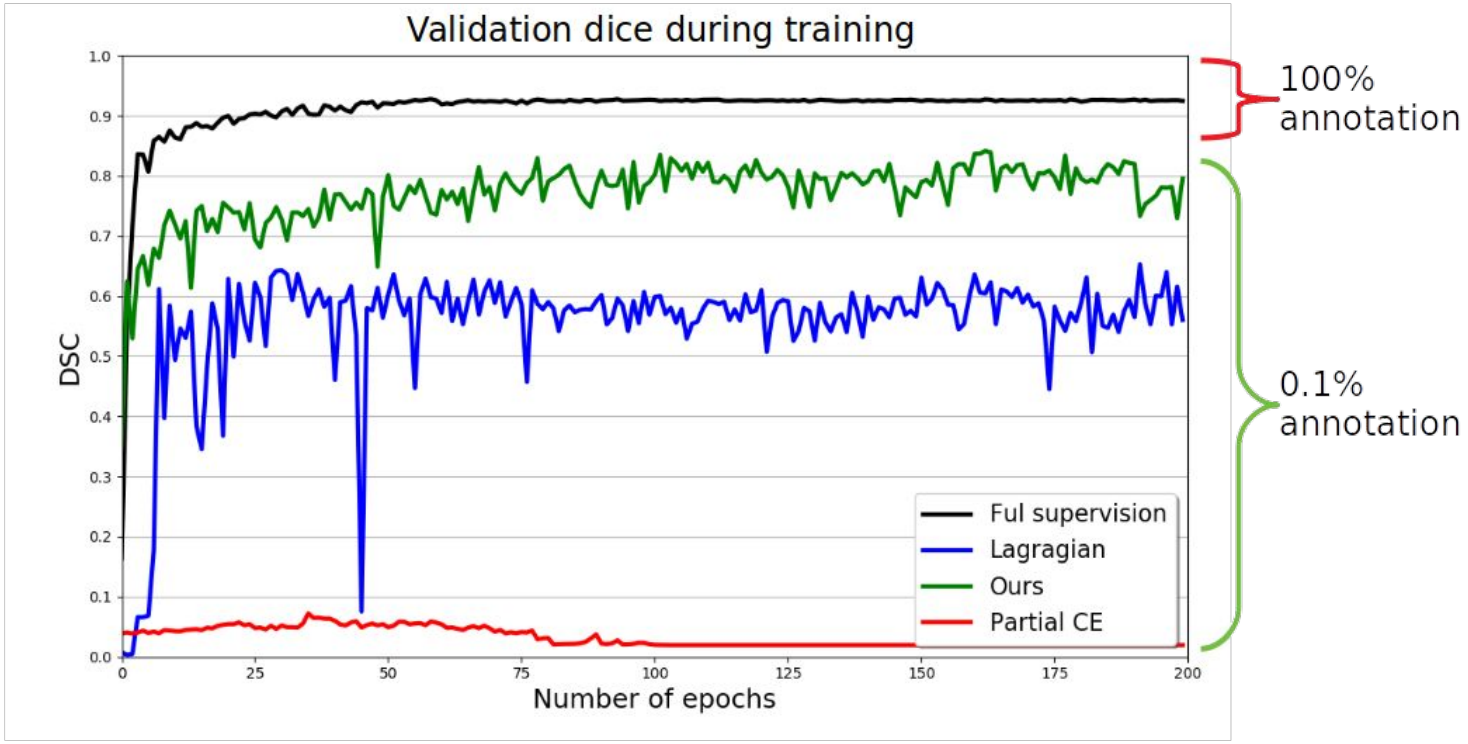
Example: Left ventricle segmentation in cardiac MRI with volumetric constraints

[Kervadek et al., MedIA'19]



Example: Left ventricle segmentation in cardiac MRI with volumetric constraints

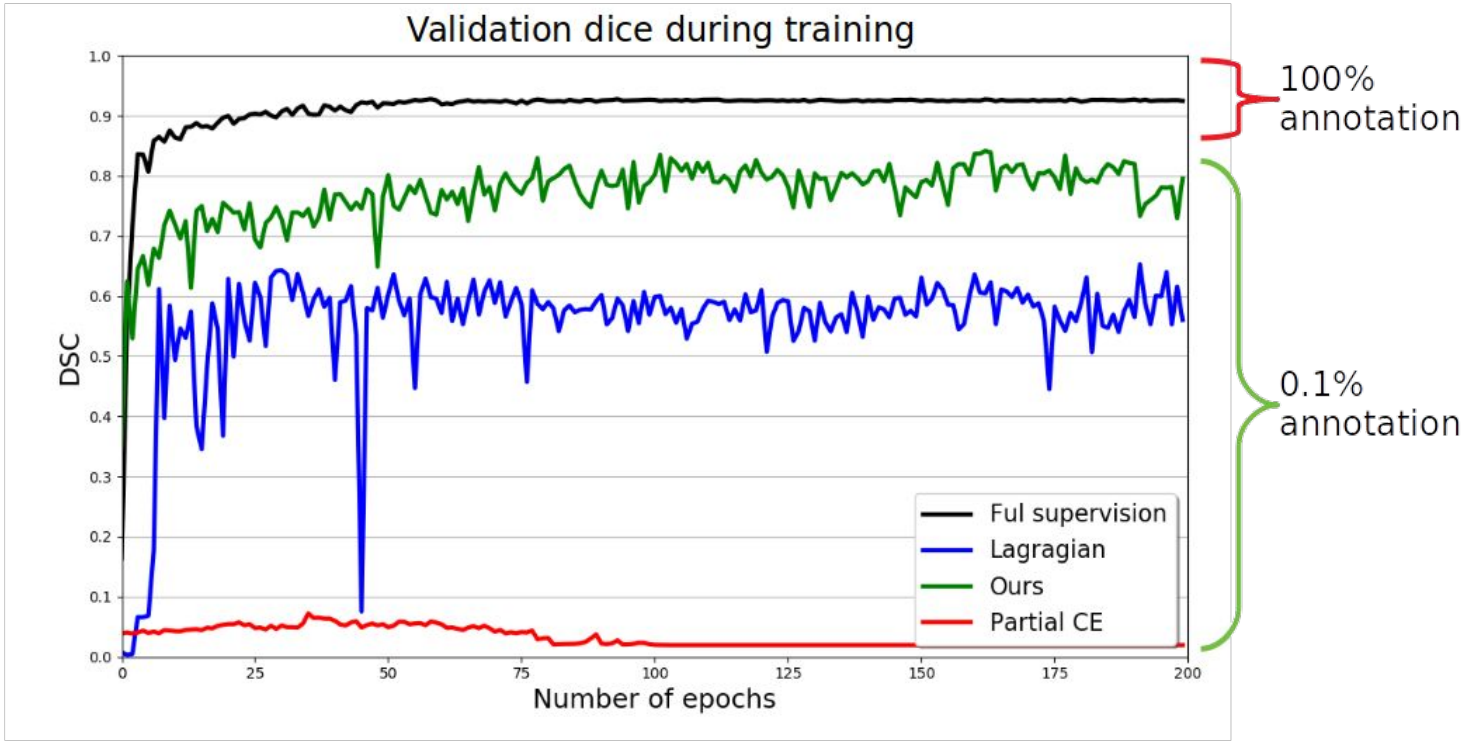
[Kervadek et al., MedIA'19]



Example: Left ventricle segmentation in cardiac MRI with volumetric constraints

[Kervadek et al., MedIA'19]

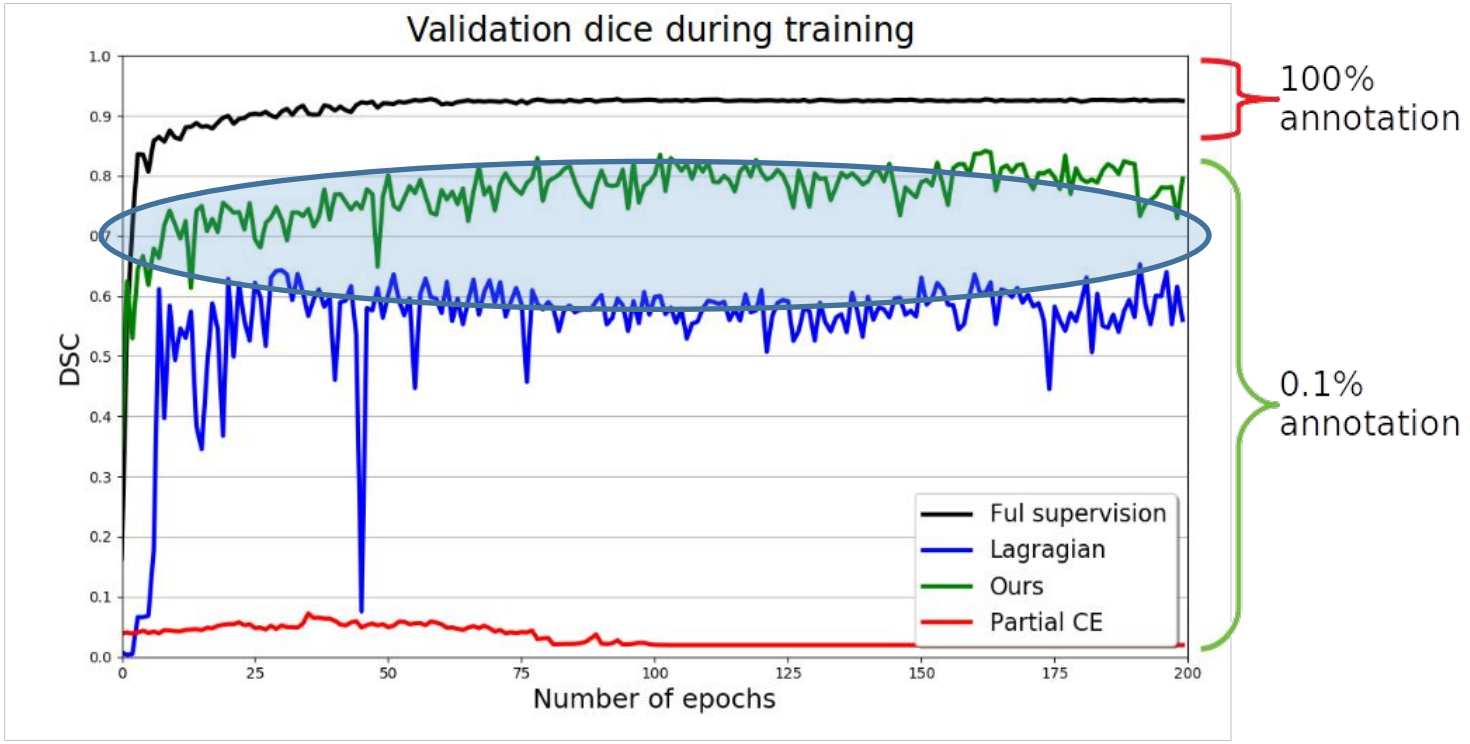
The exciting part: 90% of full supervision Dice with 0.1% of labels



Example: Left ventricle segmentation in cardiac MRI with volumetric constraints

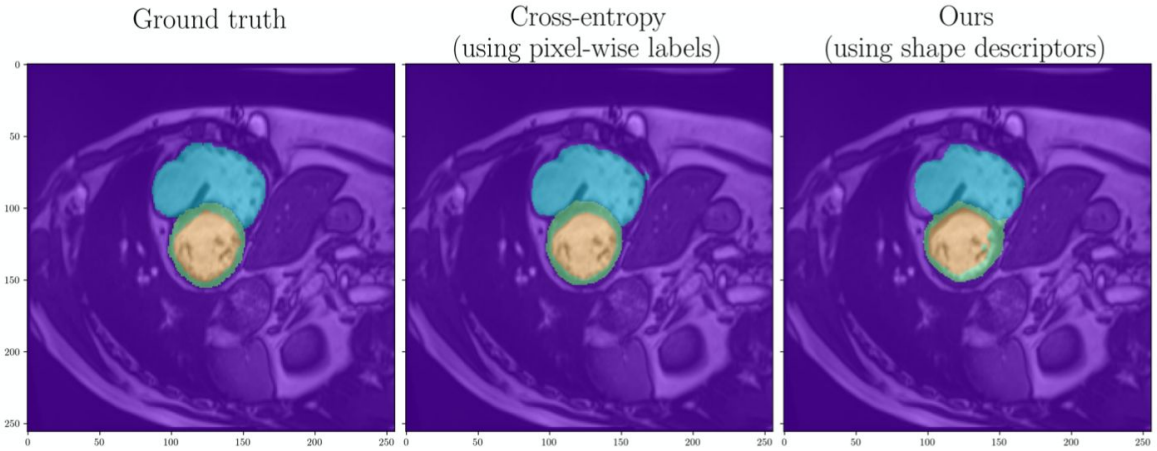
[Kervadek et al., MedIA'19]

The surprising part: Lagrangian optimization is much worse than a simple penalty



Beyond size: Exploring shape priors as functions of network outputs

[Kervadek et al., MIDL'21]



(a) A visual comparison of the different supervision methods on the ACDC dataset.

Pixel	Label
0	RV
1	Background
2	LV
⋮	
65536	Background

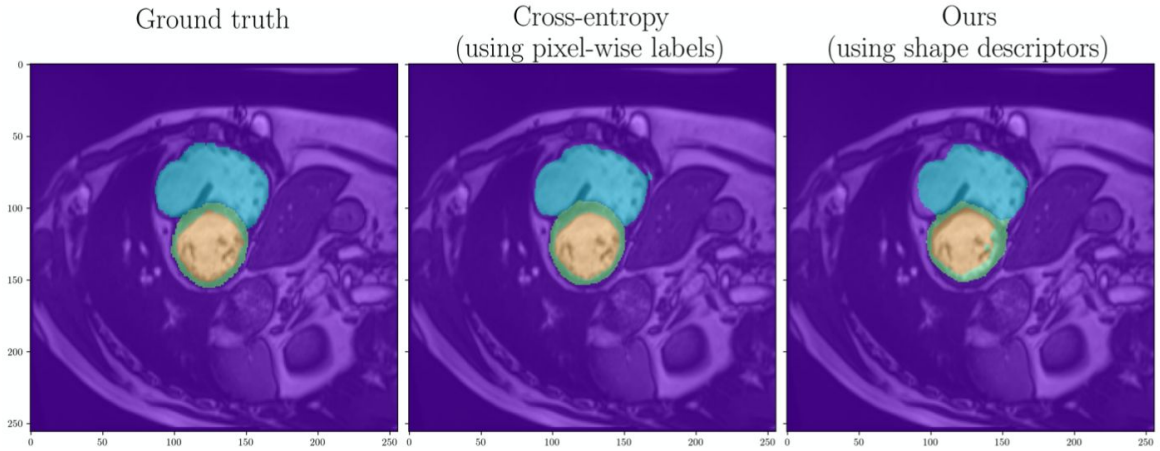
(b) Pixel-wise labels (**65k** discrete values)

Shape descriptor (in pixels)	Class		
	RV	Myo	LV
Object volume $\mathfrak{V}$	3100	800	1600
Centroid location $\mathfrak{C}$	(125, 80)	(125, 125)	
Avg. dist. to centroid $\mathfrak{D}$	(20, 15)	(15, 20)	(10, 10)
Object length $\mathfrak{L}$	750	1000	500

(c) Shape descriptors (**16** continuous values)

Beyond size: Exploring shape priors as functions of network outputs

[Kervadek et al., MIDL'21]



(a) A visual comparison of the different supervision methods on the ACDC dataset.

Pixel	Label
0	RV
1	Background
2	LV
⋮	
65536	Background

(b) Pixel-wise labels (**65k** discrete values)

Shape descriptor (in pixels)	Class		
	RV	Myo	LV
Object volume $\mathfrak{V}$	3100	800	1600
Centroid location $\mathfrak{C}$	(125, 80)	(125, 125)	
Avg. dist. to centroid $\mathfrak{D}$	(20, 15)	(15, 20)	(10, 10)
Object length $\mathfrak{L}$	750	1000	500

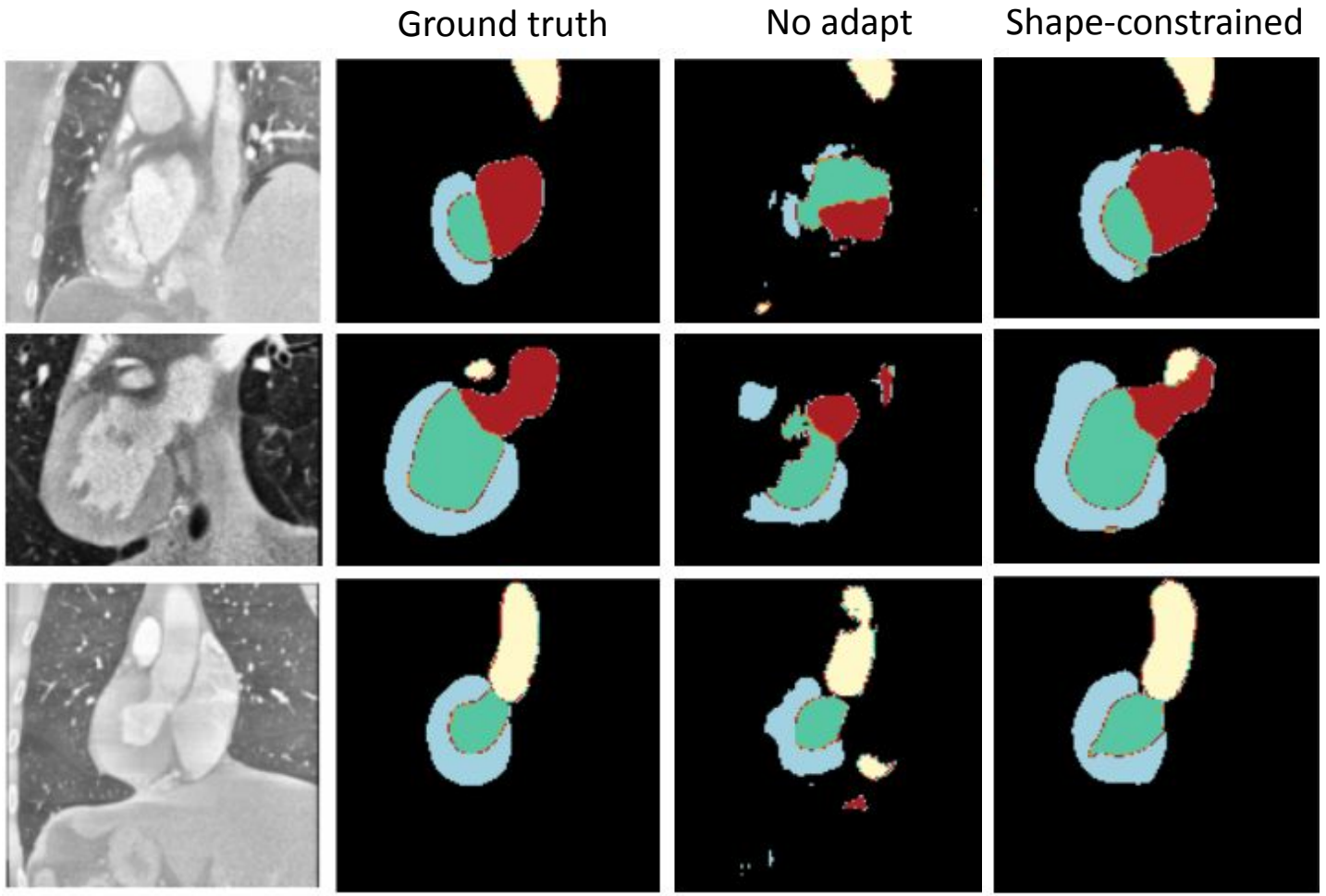
(c) Shape descriptors (**16** continuous values)

$$\mu_{i,j}^{c} = \sum_{p} s_{\theta}^{p,c} x_{p}^{i} y_{p}^{j}$$

Spatial coordinates

A few shape descriptors are surprisingly powerful in Test-Time Adaptation for Segmentation

[Bateson et al., MICCAI 2022]



References & acknowledgments of co-authors (1)

I. M. Ziko, J. Dolz, E. Granger and I. Ben Ayed, Laplacian regularized few-shot learning, **ICML 2020**
https://github.com/imtiazziko/LaplacianShot

M. Boudiaf, R. Mueller, I. Ben Ayed and L. Bertinetto, Parameter-free online test time adaptation, **CVPR 2022**
https://github.com/fiveai/lame

M. Tang, F. Perazzi, A. Djelouah, I. Ben Ayed, C. Schroers, Y. Boykov, On regularized losses for weakly supervised CNN segmentation, **ECCV 2018**
https://github.com/meng-tang/rloss

O. Veilleux, M. Boudiaf, P. Piantanida and I. Ben Ayed, Realistic Evaluation of Transductive Few-Shot Learning, **NeurIPS 2021**
https://github.com/oveilleux/realistic_transductive_few_shot

M. Bateson, H. Lombaert, I. Ben Ayed, Test-time adaptation with shape moments for image segmentation, **MICCAI 2022**
https://github.com/mathilde-b/TTA

References & acknowledgments of co-authors (2)

H. Kervadec, J. Dolz, M. Tang, E. Granger, Y. Boykov, I. Ben Ayed, Constrained-CNN losses for weakly supervised segmentation, **MedIA 2019**
https://github.com/LIVIAETS/SizeLoss_WSS

H. Kervadec, H. Bahig, L. Letourneau-Guillon, J. Dolz, I. Ben Ayed, Pixel-wise supervision for segmentation: A few global shape descriptors might be surprisingly good!, **MIDL 2021**
https://github.com/hkervadec/shape_descriptors

M. Boudiaf, I. M. Ziko, J. Rony, J. Dolz, P. Piantanida, I. Ben Ayed, Transductive information maximization for few-shot learning, **NeurIPS 2020**
https://github.com/mboudiaf/TIM

M. Boudiaf, H. Kervadec, I. M. Ziko, P. Piantanida, I. Ben Ayed, J. Dolz , Few-shot segmentation without meta-learning: A good transductive inference is all you need? **CVPR 2021**
https://github.com/mboudiaf/RePRI-for-Few-Shot-Segmentation

Thank you…