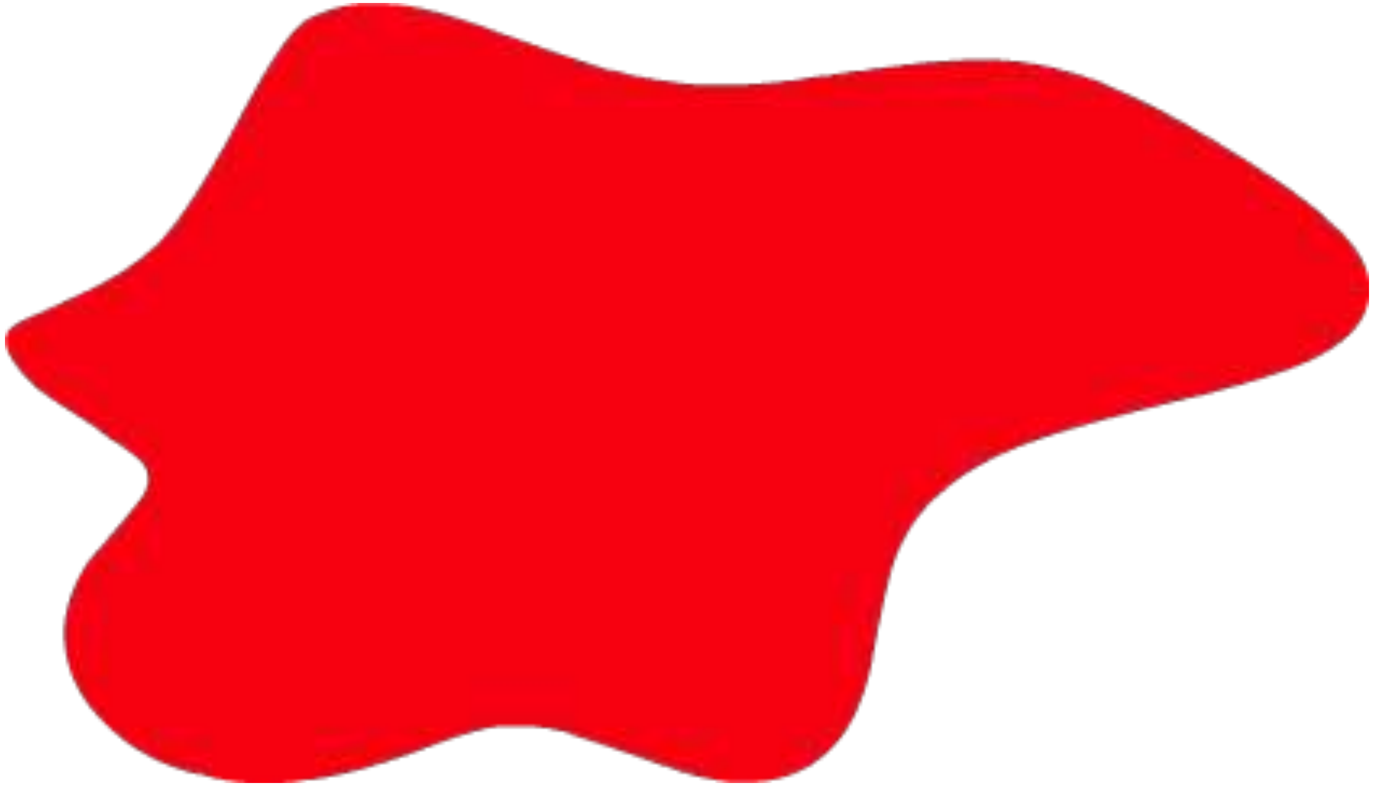
Constrained CNNs

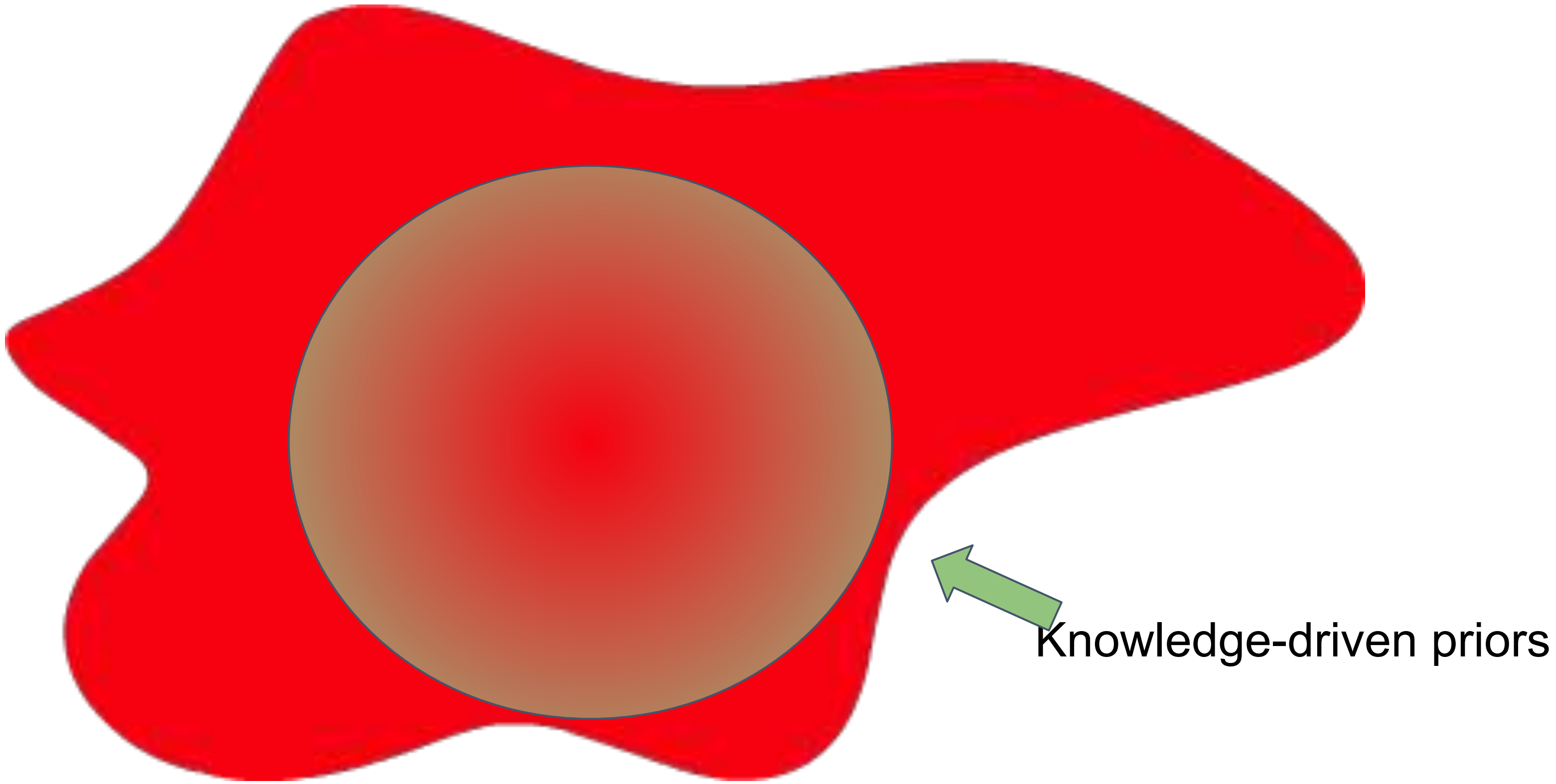
Constrained optimization (in CNNs)

Knowledge vs data driven priors



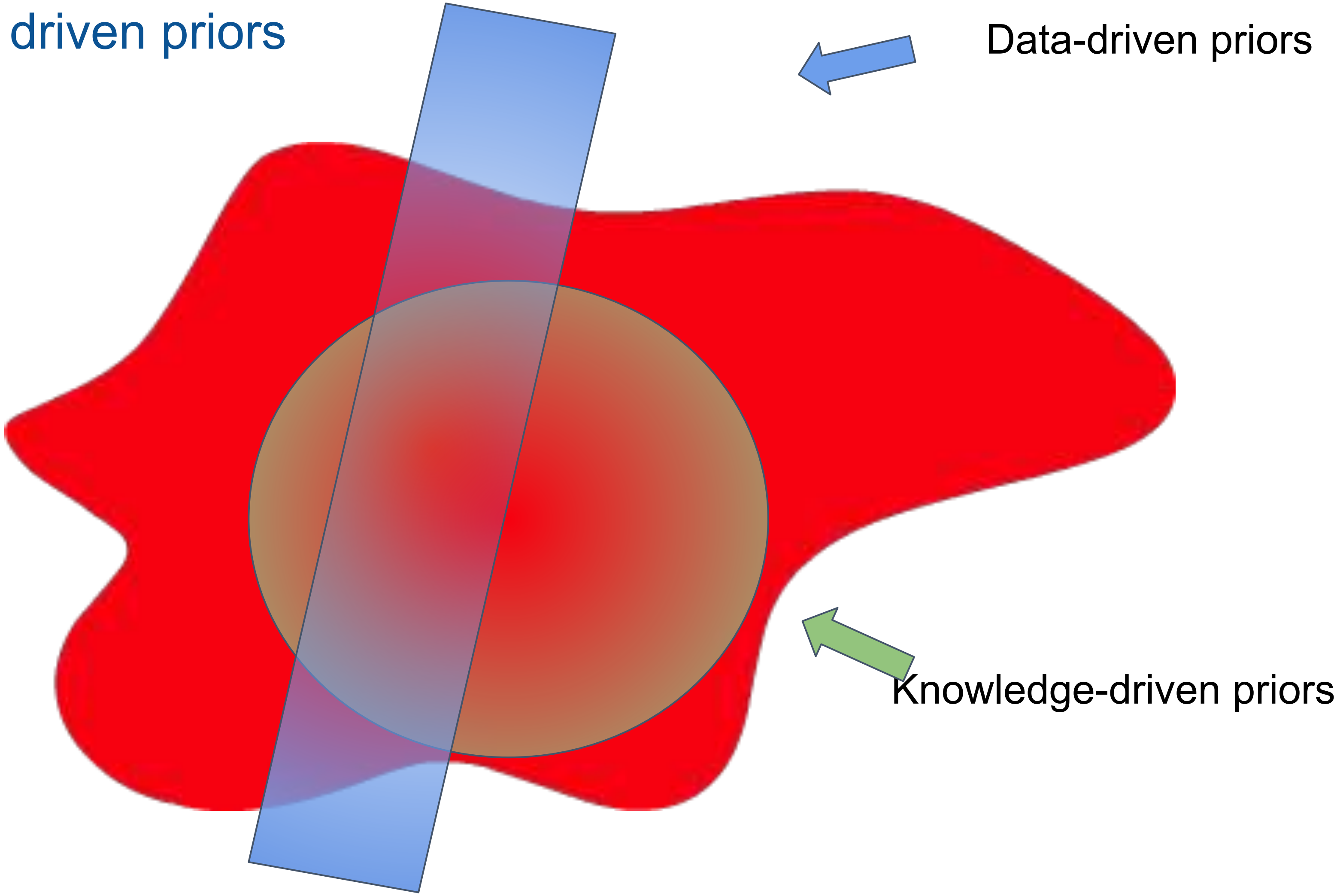
Constrained optimization (in CNNs)

Knowledge vs data driven priors



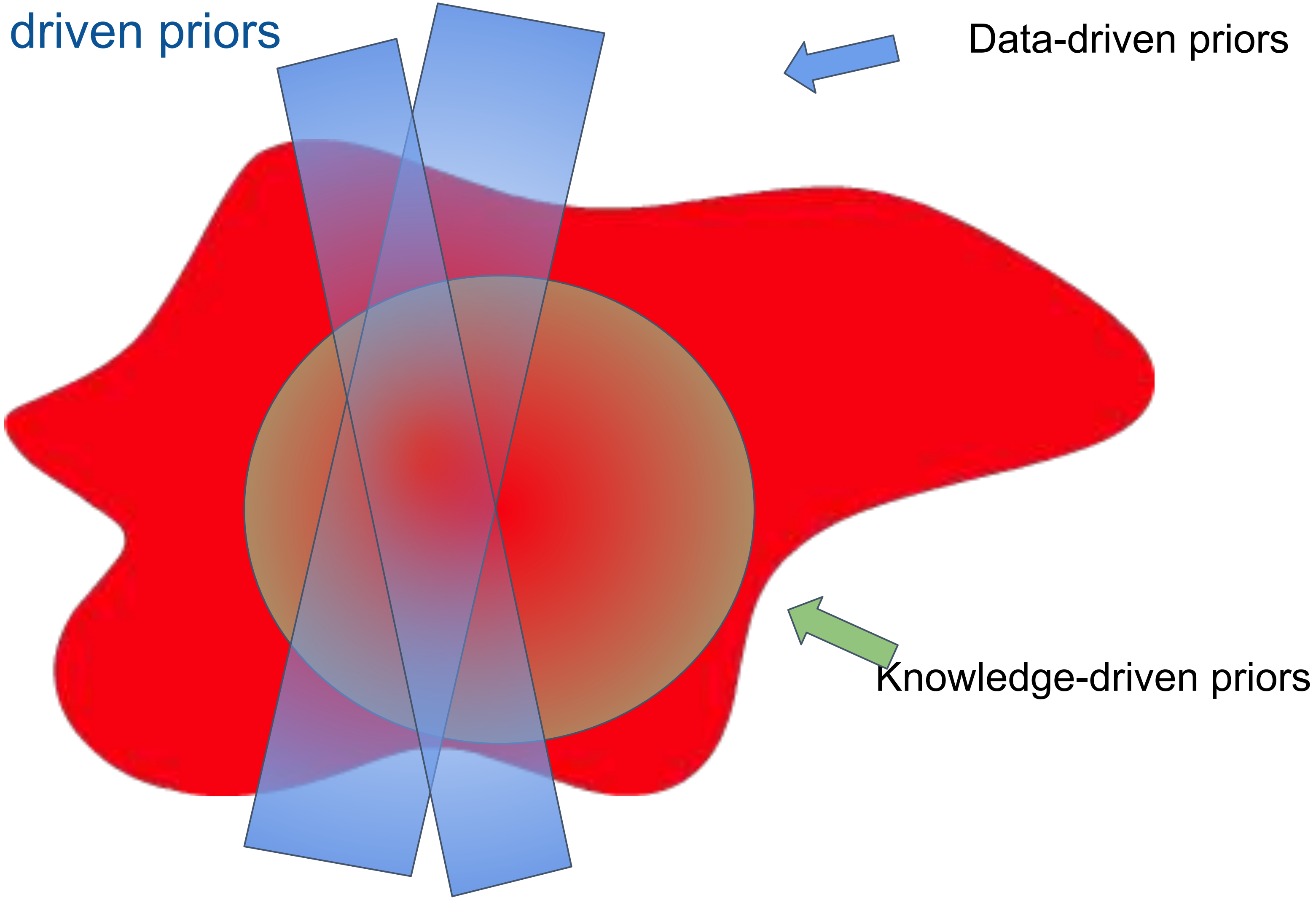
Constrained optimization (in CNNs)

Knowledge vs data driven priors



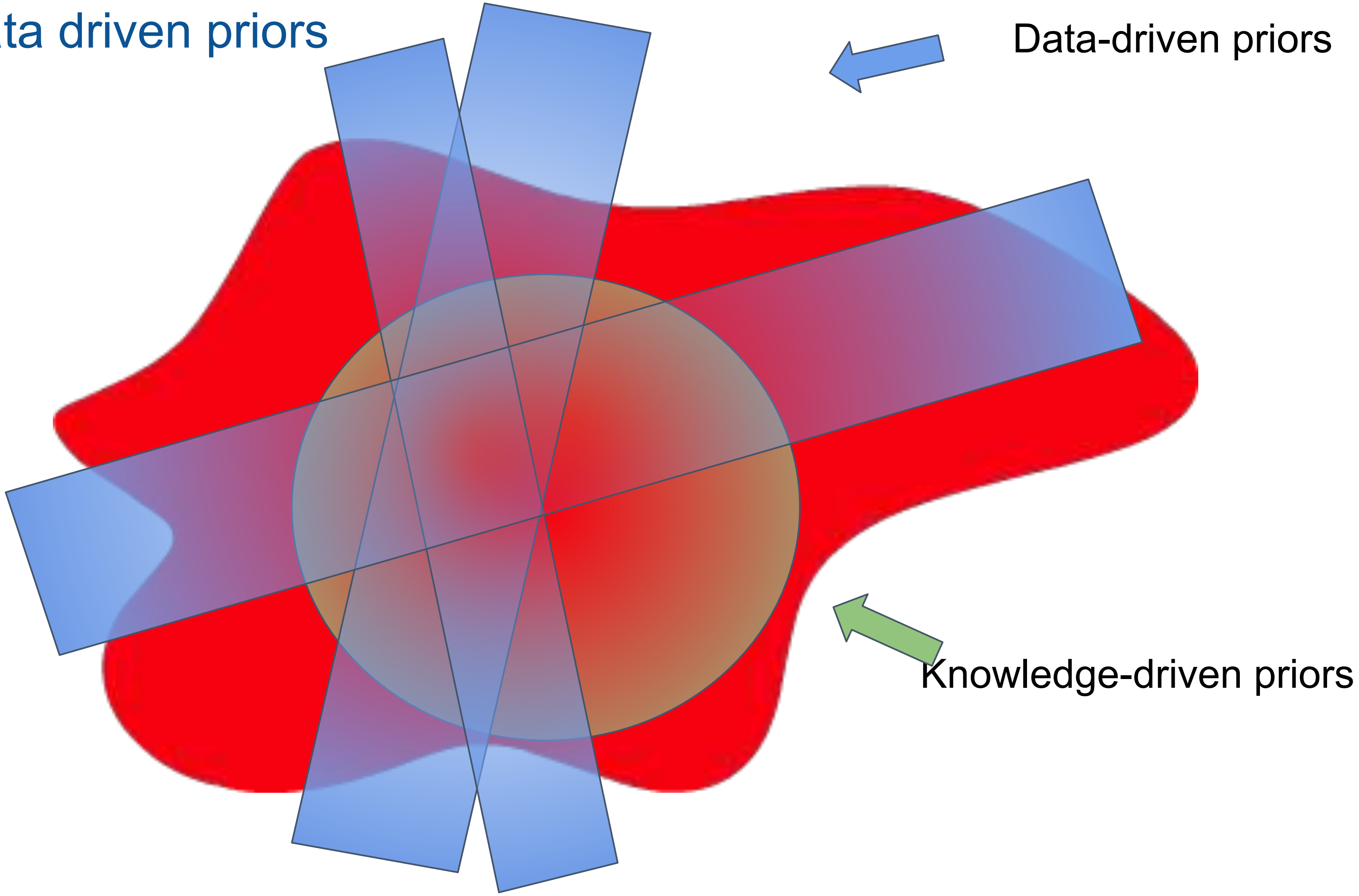
Constrained optimization (in CNNs)

Knowledge vs data driven priors



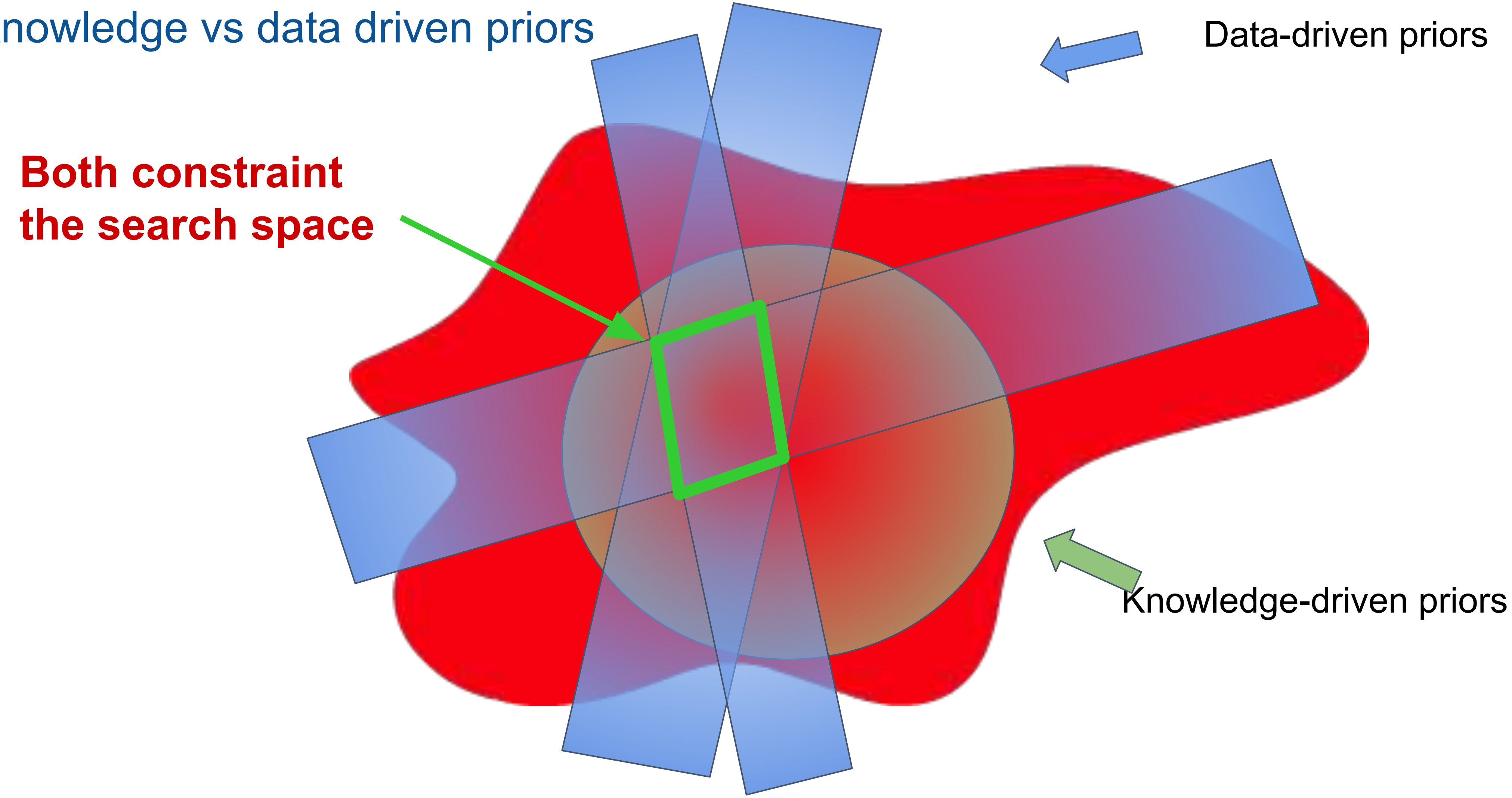
Constrained optimization (in CNNs)

Knowledge vs data driven priors



Constrained optimization (in CNNs)

Knowledge vs data driven priors



Constrained optimization (in CNNs)

Knowledge-driven priors

Common priors in natural images

Target Size

- Pathak et al., Constrained convolutional neural networks for weakly supervised segmentation, ICCV 2015
- Xu et al., Learning to Segment Under Various Forms of Weak Supervision, CVPR 2015
- Zhang et al., Curriculum Domain Adaptation for Semantic Segmentation of Urban Scenes. ICCV’17

Constrained optimization (in CNNs)

Knowledge-driven priors

Common priors in natural images

Target Location

- Remez et al. Learning to segment via cut-and-paste. ECCV 2018
- Georgakis et al Synthesizing training data for object detection in indoor scenes. RSS 2017

Constrained optimization (in CNNs)

Knowledge-driven priors

Common priors in natural images

Number of instances

- Deselaers et al. Localizing objects while learning their appearance. ECCV 2010

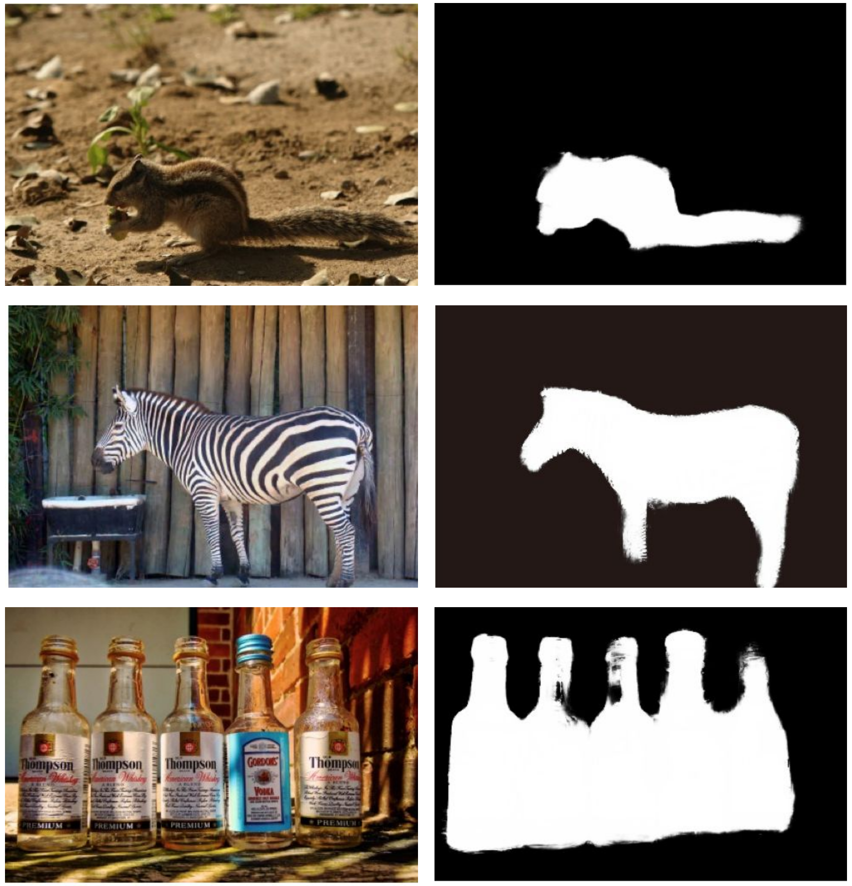
Constrained optimization (in CNNs)

Knowledge-driven priors

Common priors in natural images

Contrast
Foreground/Background

Saliency



Images from Hou et al, CVPR'17

- Hou et al. Deeply supervised salient object detection with short connections. CVPR 2017
- Li et al. Instance-level salient object segmentation. CVPR 2017

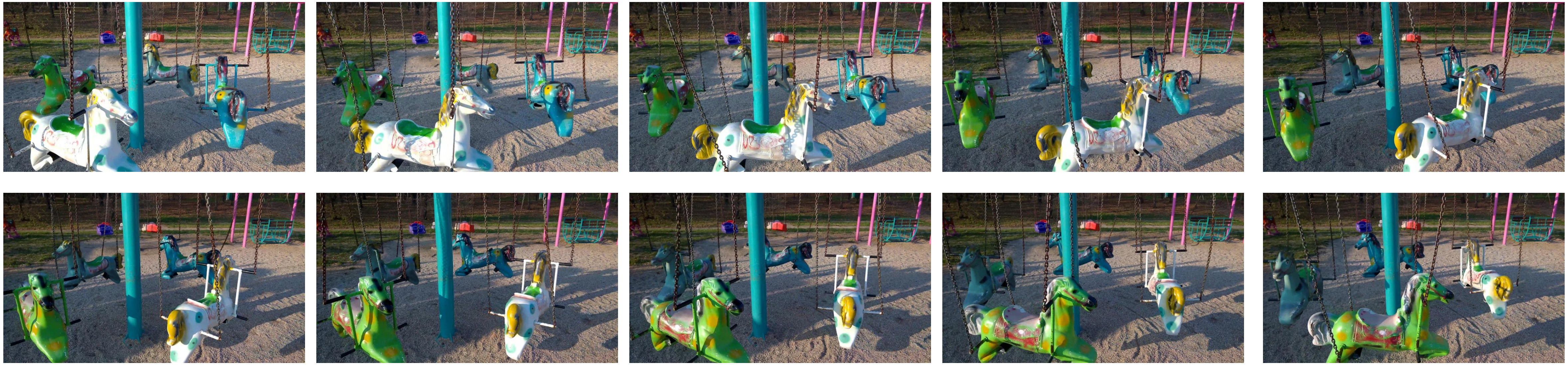
Constrained optimization (in CNNs)

Knowledge-driven priors

Common priors in natural images

Motion

Images from the DAVIS Challenge Dataset

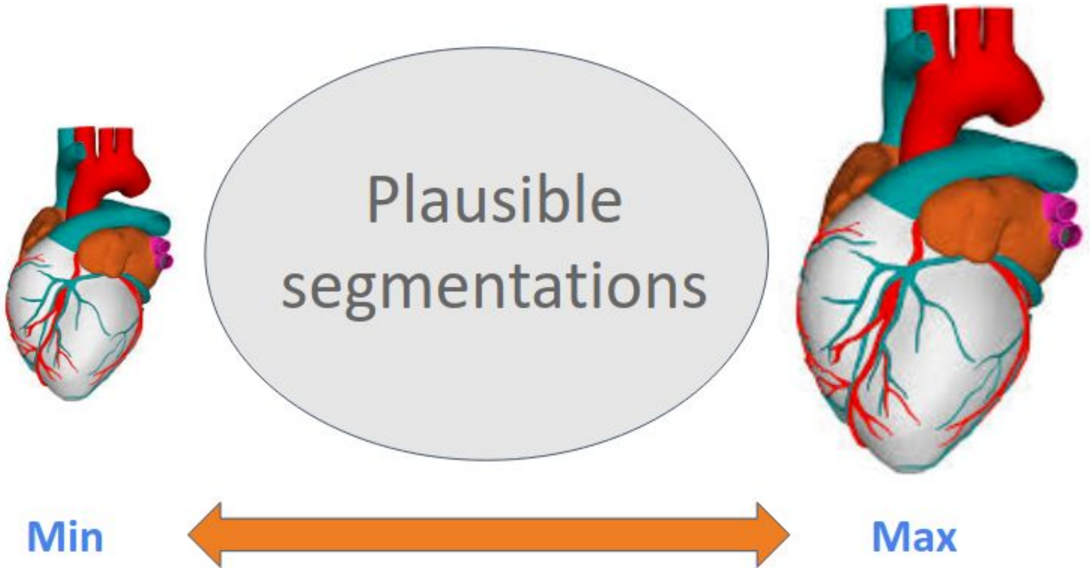


- Tokmakov et al. Weakly-supervised semantic segmentation using motion cues. ECCV 2016
- Pathak et al. Learning features by watching objects move. CVPR 2017

Constrained optimization (in CNNs)

Knowledge-driven priors

What about priors in the medical domain?

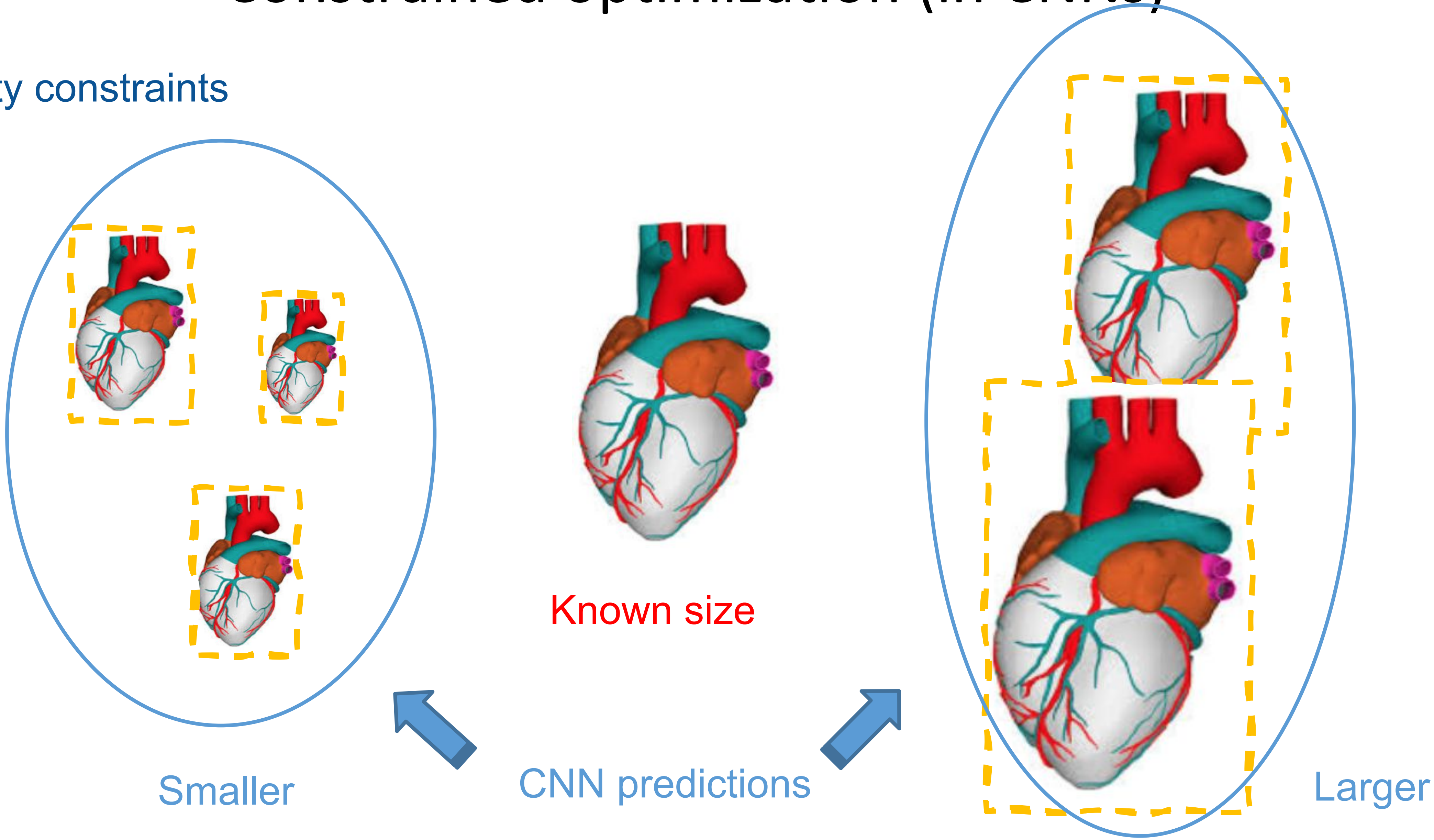


Anatomical priors

Partial labeled data
(exploit target relationships)

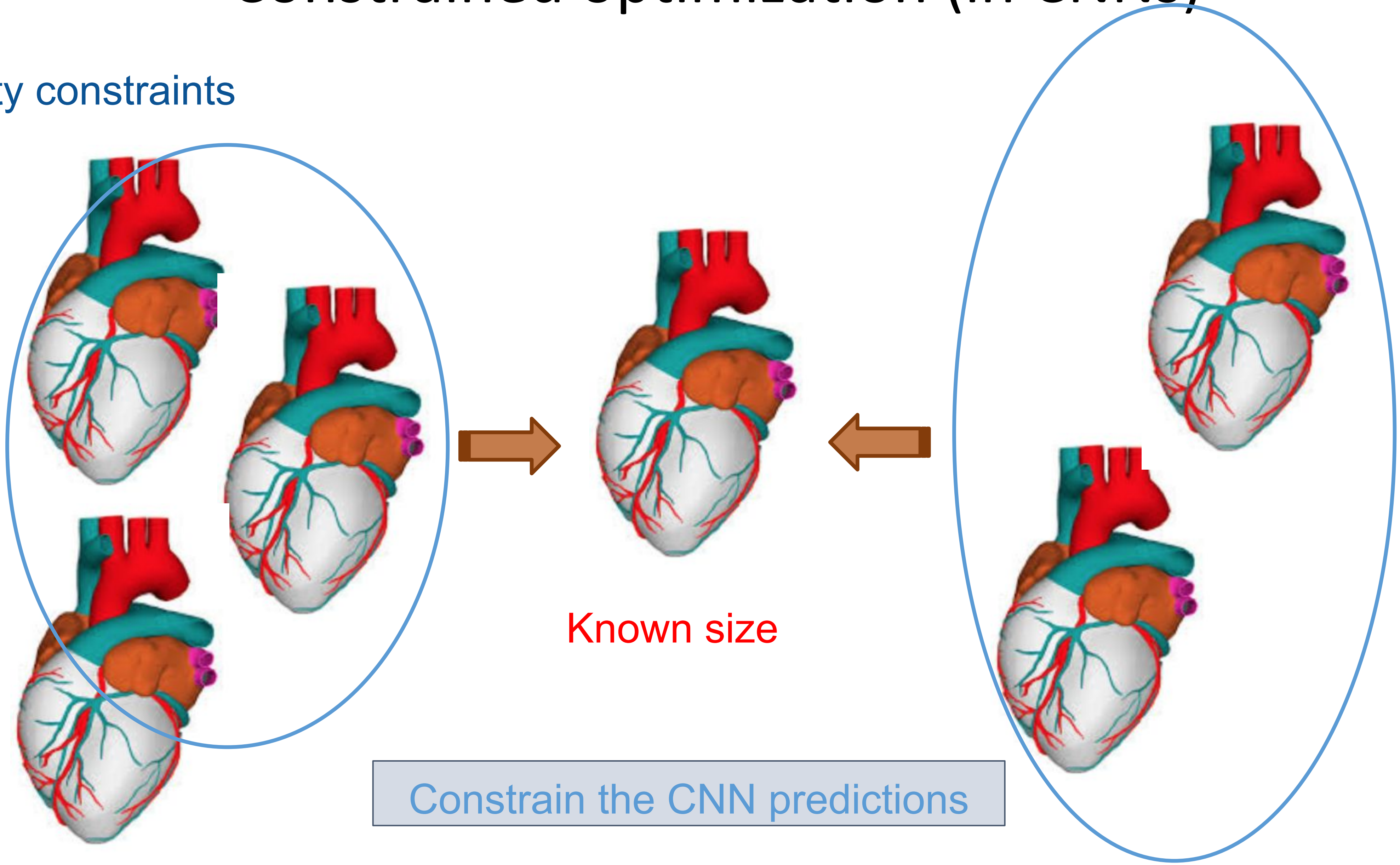
Constrained optimization (in CNNs)

Equality constraints



Constrained optimization (in CNNs)

Equality constraints

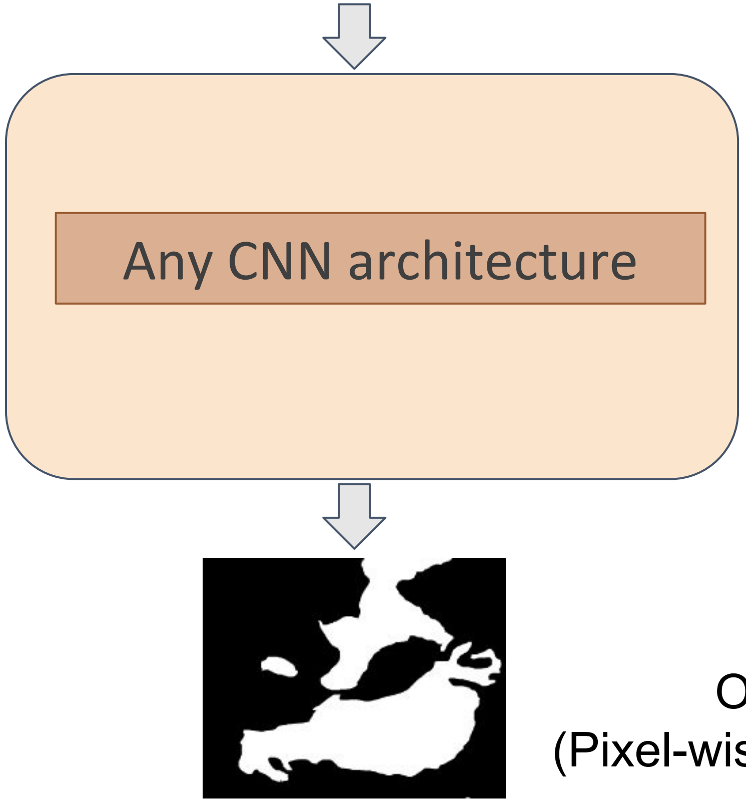


Constrain the CNN predictions

Constrained optimization (in CNNs)

Equality constraints (e.g, L2 penalty)

Input
(Histology image)



Output
(Pixel-wise prediction)

Jia Z,et al. Constrained deep weak supervision for histopathology image segmentation. IEEE TMI 2017

Constrained optimization (in CNNs)

Equality constraints (e.g, L2 penalty)

Input
(Histology image)

Any CNN architecture

Output
(Pixel-wise prediction)

Additional term

$$\boxed{l_{ac} = \boldsymbol{I}(Y_i = 1) \sum_i (v_i - a_i)^2}$$

Reminder: ½ KL is an upper bound on L_2 for simplex vectors (Pinsker's inequality)

Predicted relative size (%)

$$v_i = \frac{1}{N} \sum_{p \in \Omega} s_\theta^{p,1}$$

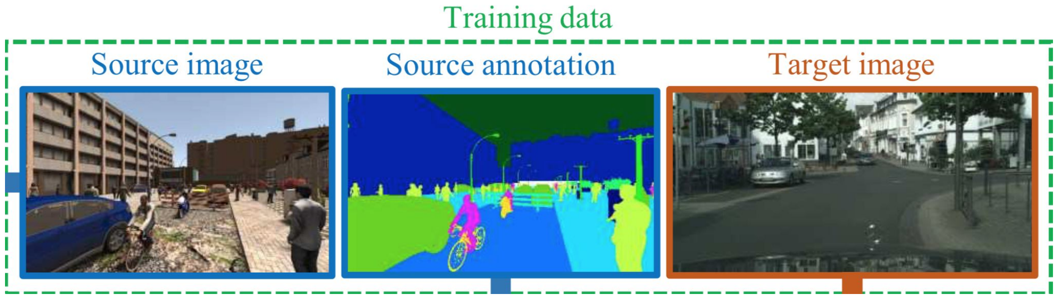
Predicted size given by experts

Jia Z,et al. Constrained deep weak supervision for histopathology image segmentation. IEEE TMI 2017

Constrained optimization (in CNNs)

Equality constraints (e.g, KL)

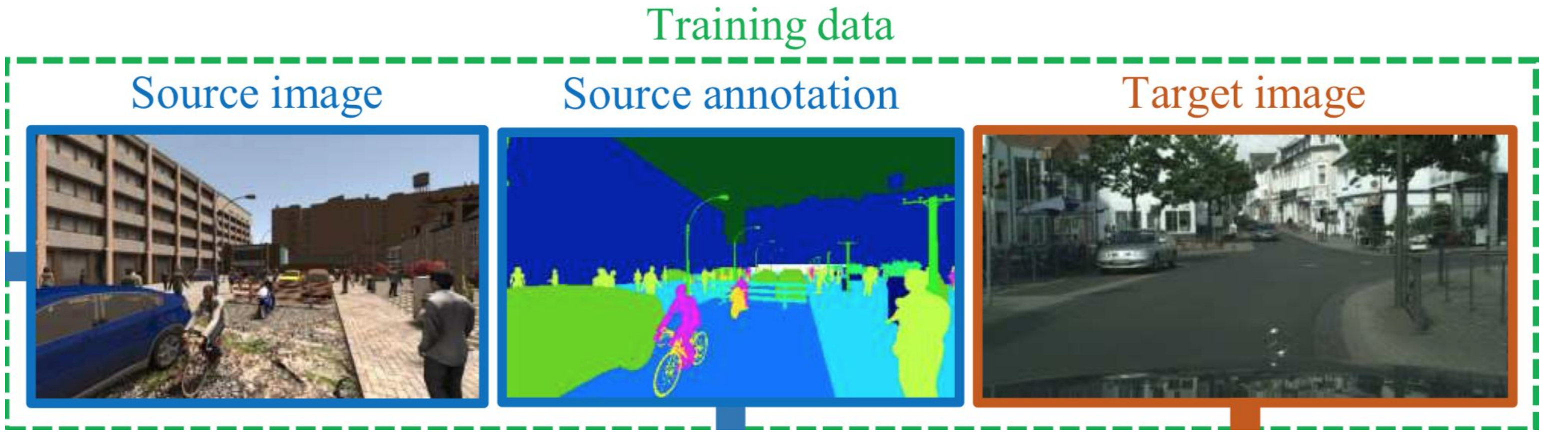
Unsupervised domain adaptation



Partially labeled data

Constrained optimization (in CNNs)

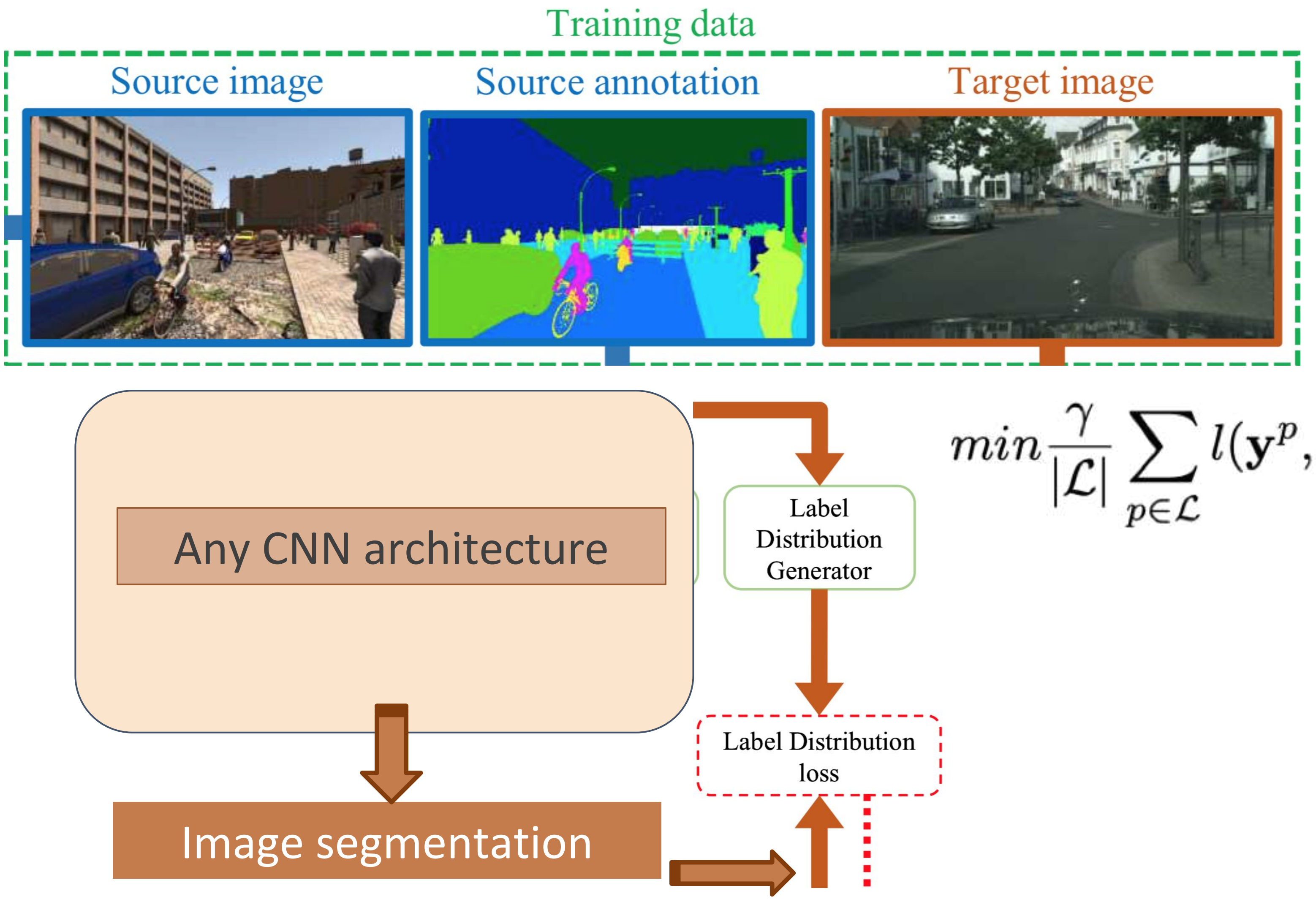
Equality constraints (e.g, KL): Curriculum DA



Zhang et al., Curriculum Domain Adaptation for Semantic Segmentation of Urban Scenes. ICCV'17

Constrained optimization (in CNNs)

Equality constraints (e.g, KL): Curriculum DA

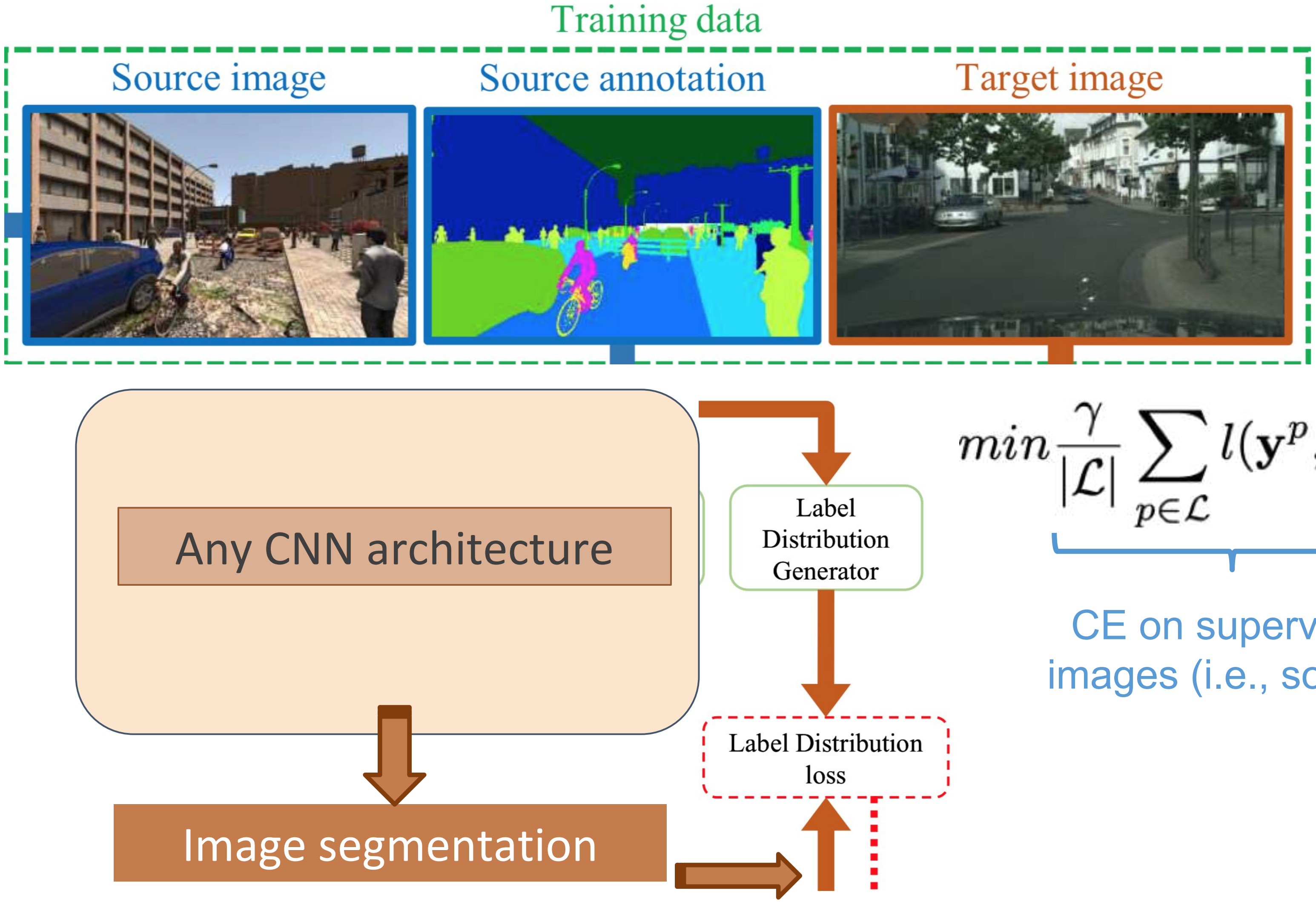


$$min \frac{\gamma}{|\mathcal{L}|} \sum_{p \in \mathcal{L}} l(\mathbf{y}^p, s_\theta^p) + \frac{1-\gamma}{|\mathcal{U}|} \sum_{q \in \mathcal{U}} \sum_k \mathbf{C}(a^{q,k}, \hat{a}^{q,k})$$

Zhang et al., Curriculum Domain Adaptation for Semantic Segmentation of Urban Scenes. ICCV’17

Constrained optimization (in CNNs)

Equality constraints (e.g, KL): Curriculum DA

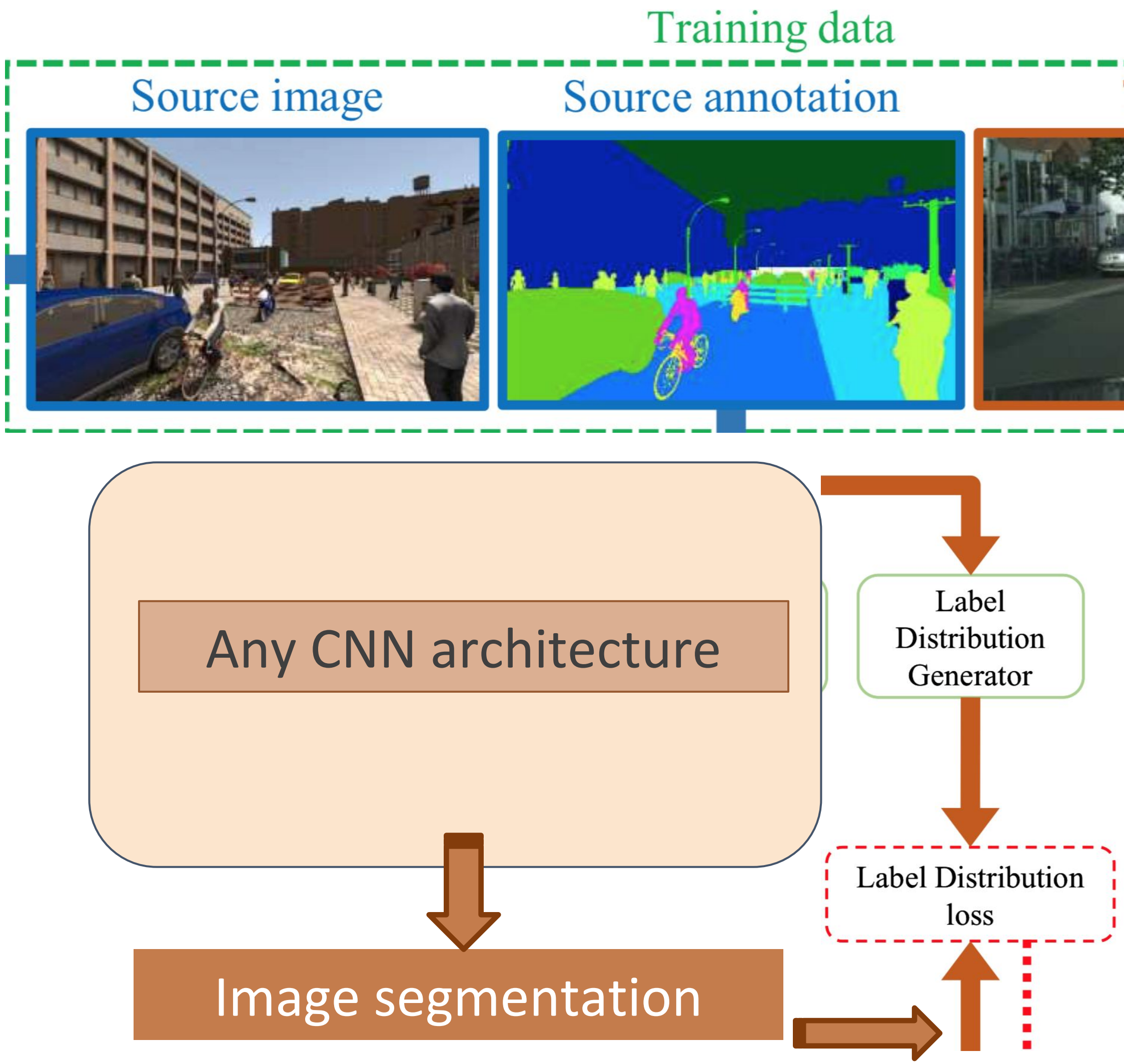


$$\underbrace{min \frac{\gamma}{|\mathcal{L}|} \sum_{p \in \mathcal{L}} l(\mathbf{y}^p, s_\theta^p)}_{\text{CE on supervised images (i.e., source)}} + \frac{1-\gamma}{|\mathcal{U}|} \sum_{q \in \mathcal{U}} \sum_{k} \mathbf{C}(a^{q,k}, \hat{a}^{q,k})$$

Zhang et al., Curriculum Domain Adaptation for Semantic Segmentation of Urban Scenes. ICCV'17

Constrained optimization (in CNNs)

Equality constraints (e.g, KL): Curriculum DA



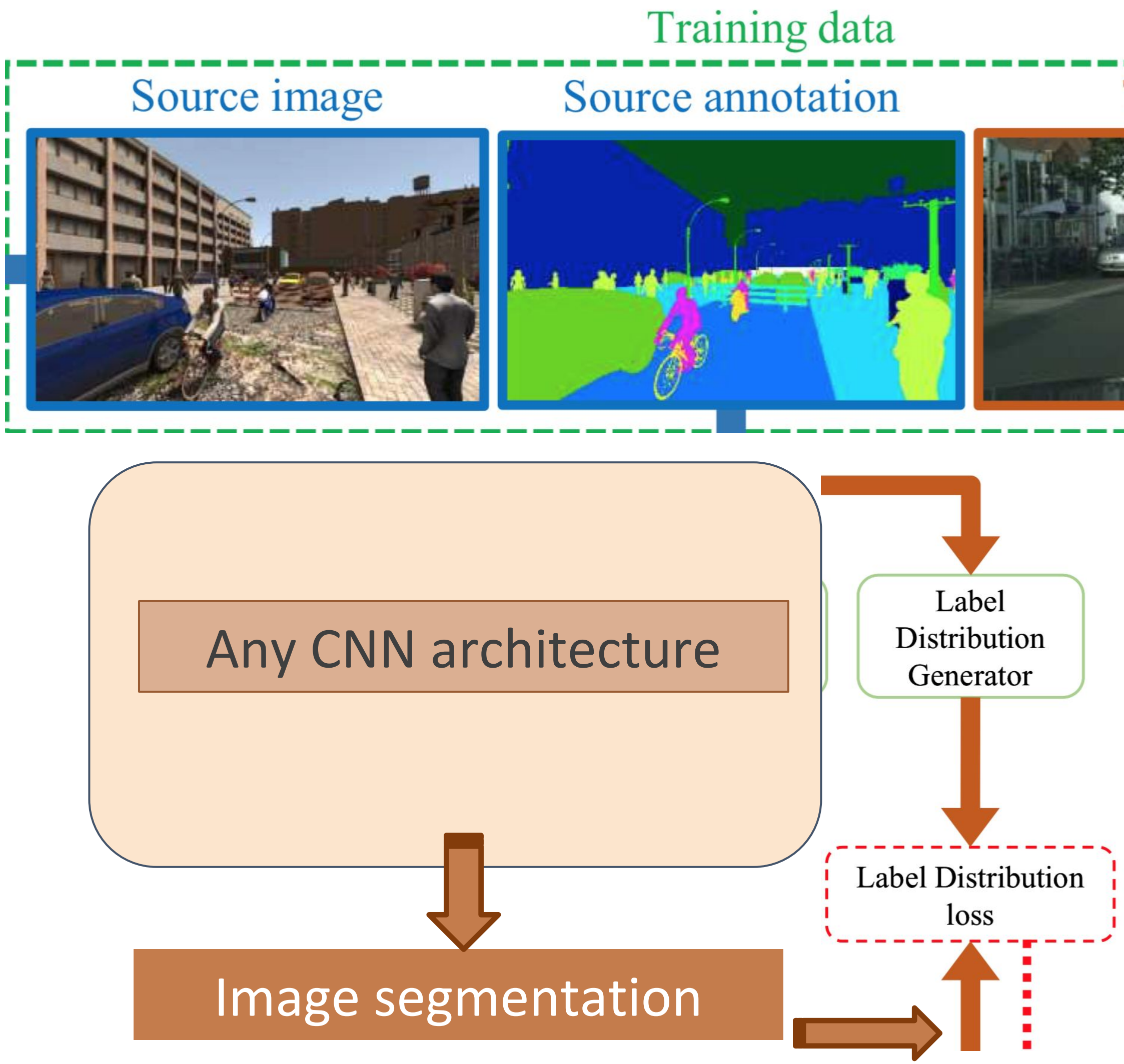
Additional term

$$min \underbrace{\frac{\gamma}{|\mathcal{L}|}\sum_{p\in\mathcal{L}} l(\mathbf{y}^p, s^p_\theta)}_{\text{CE on supervised images (i.e., source)}} + \frac{1-\gamma}{|\mathcal{U}|}\sum_{q\in\mathcal{U}}\sum_{k} \mathbf{C}(a^{q,k}, \hat{a}^{q,k})$$

Zhang et al., Curriculum Domain Adaptation for Semantic Segmentation of Urban Scenes. ICCV'17

Constrained optimization (in CNNs)

Equality constraints (e.g, KL): Curriculum DA



Additional term

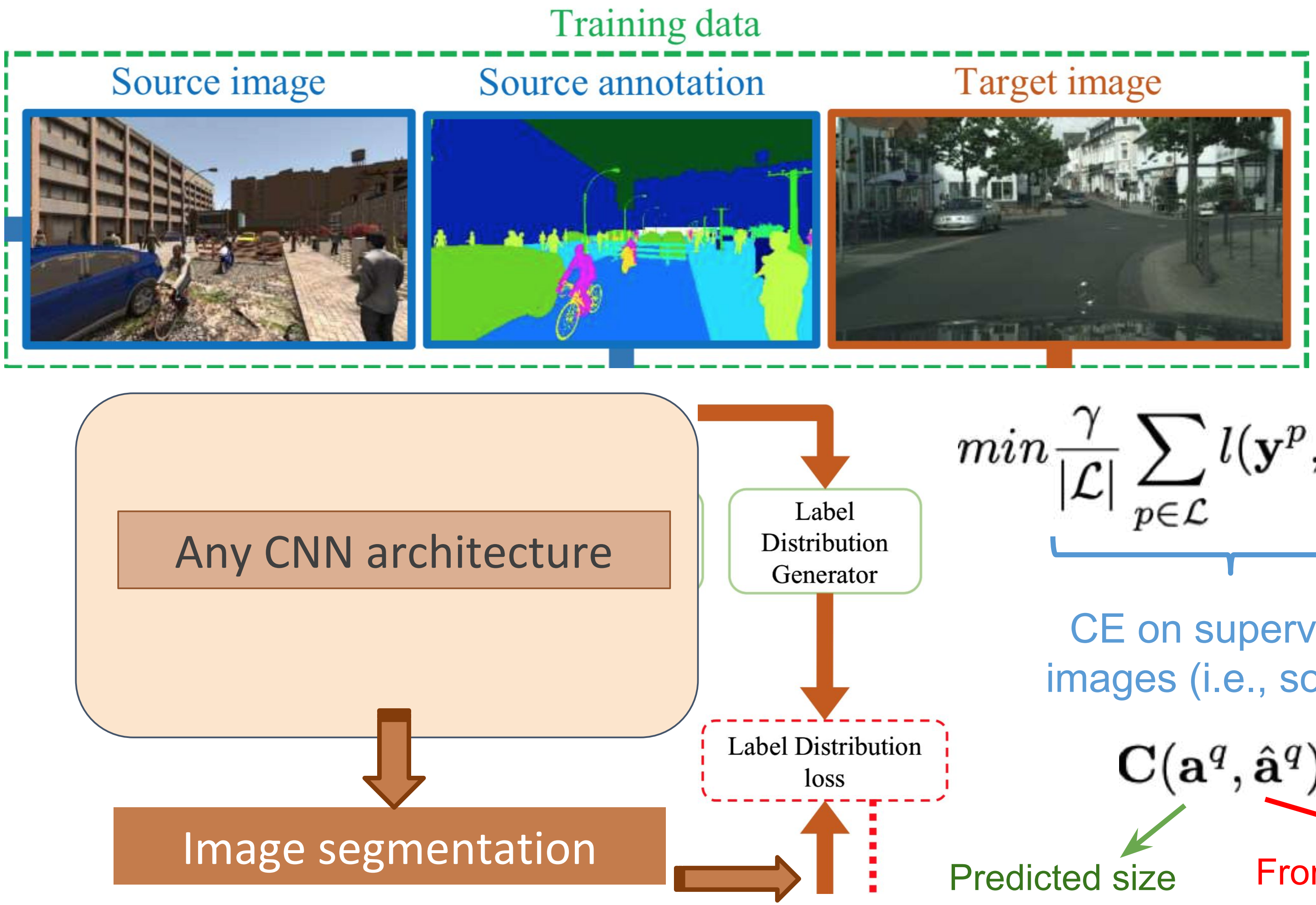
$$\min \underbrace{\frac{\gamma}{|\mathcal{L}|}\sum_{p\in\mathcal{L}} l(\mathbf{y}^p, s^p_\theta)}_{\text{CE on supervised images (i.e., source)}} + \frac{1-\gamma}{|\mathcal{U}|}\sum_{q\in\mathcal{U}}\sum_{k} \mathbf{C}(a^{q,k}, \hat{a}^{q,k})$$

$$\mathbf{C}(\mathbf{a}^q, \hat{\mathbf{a}}^q) = H(\mathbf{a}^q) + KL(\mathbf{a}^q, \hat{\mathbf{a}}^q)$$

Zhang et al., Curriculum Domain Adaptation for Semantic Segmentation of Urban Scenes. ICCV'17

Constrained optimization (in CNNs)

Equality constraints (e.g, KL): Curriculum DA



Additional term

$$\min \underbrace{\frac{\gamma}{|\mathcal{L}|}\sum_{p\in\mathcal{L}} l(\mathbf{y}^p, s^p_\theta)}_{\text{CE on supervised images (i.e., source)}} + \frac{1-\gamma}{|\mathcal{U}|}\sum_{q\in\mathcal{U}}\sum_{k} \mathbf{C}(a^{q,k}, \hat{a}^{q,k})$$

$$\mathbf{C}(\mathbf{a}^q, \hat{\mathbf{a}}^q) = H(\mathbf{a}^q) + KL(\mathbf{a}^q, \hat{\mathbf{a}}^q)$$

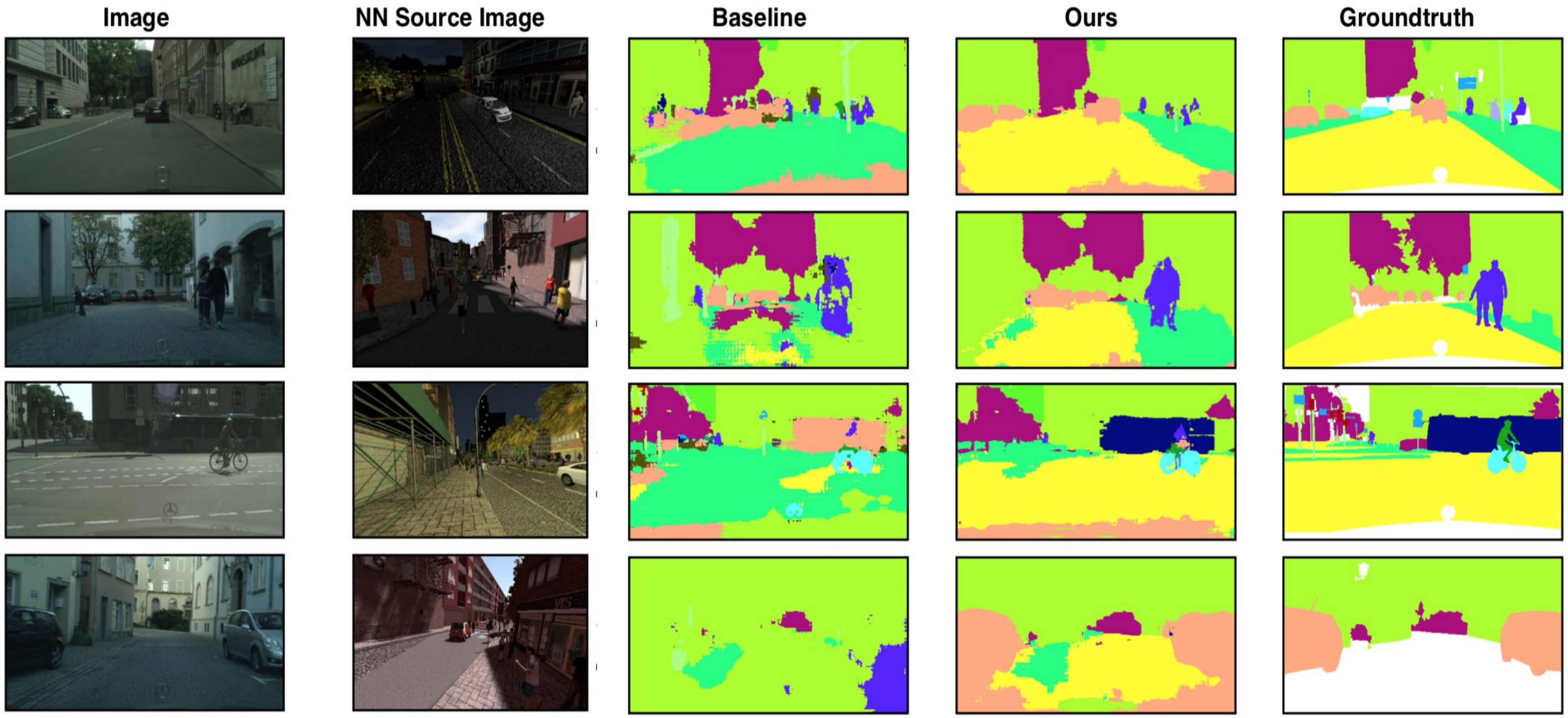
Predicted size

From predicted image

Zhang et al., Curriculum Domain Adaptation for Semantic Segmentation of Urban Scenes. ICCV’17

Constrained optimization (in CNNs)

Equality constraints (e.g, KL): Curriculum DA



Images from [Zhang et al., Curriculum Domain Adaptation for Semantic Segmentation of Urban Scenes. ICCV'17]

Constrained optimization (in CNNs)

Equality constraints (e.g, KL): Partial annotations

[Zhou et al., Prior-aware Neural Network for Partially-Supervised Multi-Organ Segmentation, ICCV’19]

Constrained optimization (in CNNs)

Equality constraints (e.g, KL): Partial annotations

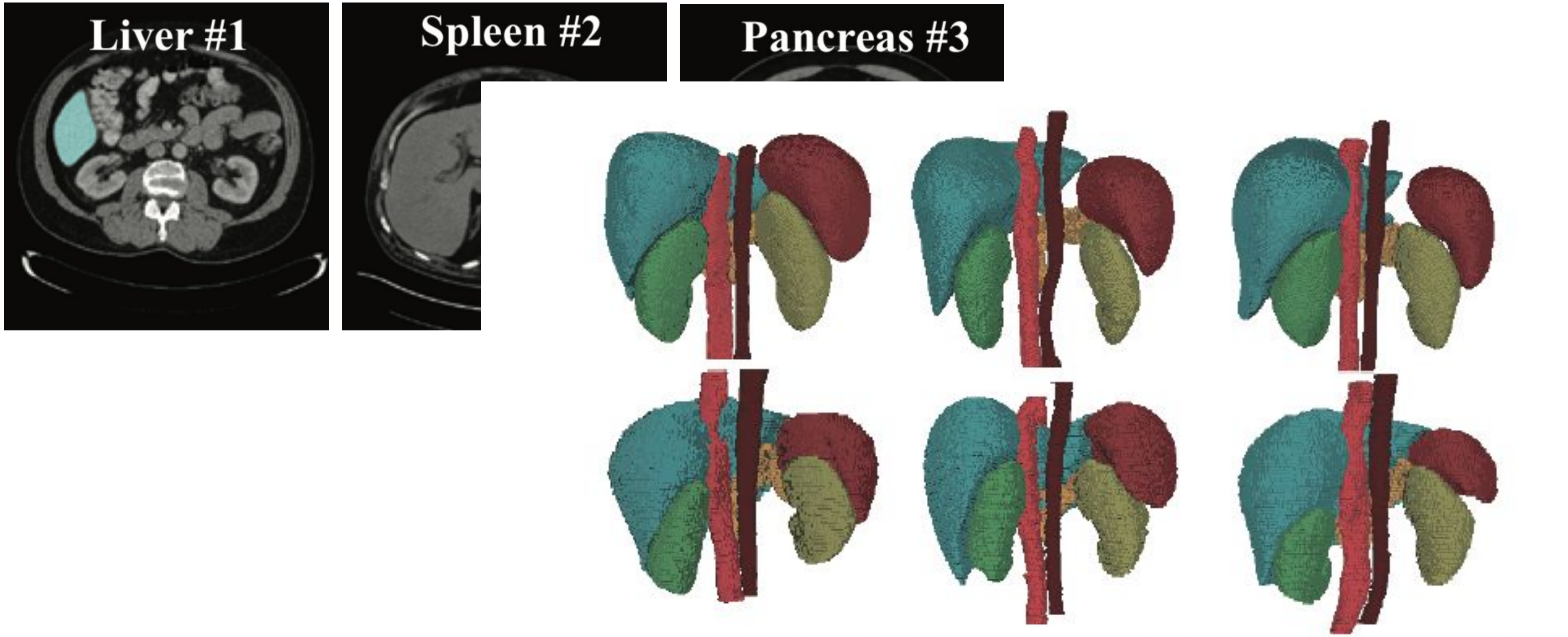


Figure 1. 3D Visualization of several abdominal organs (liver, spleen, left kidney, right kidney, aorta, inferior vena cava) to show the similarity of patient-wise abdominal organ size distributions.

Prior on the proportion

Images from [Zhou et al., Prior-aware Neural Network for Partially-Supervised Multi-Organ Segmentation, ICCV'19]

Constrained optimization (in CNNs)

Equality constraints (e.g, KL): Partial annotations

Main objective:

$$\underbrace{min \frac{1}{|\mathcal{L}|} \sum_{p \in \mathcal{L}} l(\mathbf{y}^p, \mathbf{s}_\theta^p) + \lambda_1 \frac{1}{|\mathcal{P}|} \sum_{q \in \mathcal{P}} l(\mathbf{y}^q, \mathbf{s}_\theta^q)} + \lambda_2 \mathcal{J}(\theta)$$

Fully labeled images

Partially labeled images

Prior-aware loss

[Zhou et al., Prior-aware Neural Network for Partially-Supervised Multi-Organ Segmentation, ICCV'19]

Constrained optimization (in CNNs)

Equality constraints (e.g, KL): Partial annotations

Prior-aware loss

Averaged predicted distribution

$$\hat{\mathbf{p}} = \frac{1}{N} \sum_{p \in \mathcal{P}} \mathbf{s}_\theta^p$$

$$[s_\theta^{p,0}, s_\theta^{p,1}, \ldots, s_\theta^{p,|K|}]$$

On partially labeled images

[Zhou et al., Prior-aware Neural Network for Partially-Supervised Multi-Organ Segmentation, ICCV'19]

Constrained optimization (in CNNs)

Equality constraints (e.g, KL): Partial annotations

Prior-aware loss

Embed prior knowledge

$$KL(\mathbf{q}|\hat{\mathbf{p}})$$

Real label distribution Average predicted distribution

[Zhou et al., Prior-aware Neural Network for Partially-Supervised Multi-Organ Segmentation, ICCV'19]

Constrained optimization (in CNNs)

Equality constraints (e.g, KL): Partial annotations

Prior-aware loss

KL can be expanded

$$-\sum_{c=0}^{|K|}\{q^c log\frac{1}{N}\sum_{p\in\mathcal{P}}\mathbf{s}_\theta^{p,c} + (1-q^c)log(1-\frac{1}{N}\sum_{p\in\mathcal{P}}\mathbf{s}_\theta^{p,c})\} + const$$

[Zhou et al., Prior-aware Neural Network for Partially-Supervised Multi-Organ Segmentation, ICCV’19]

Constrained optimization (in CNNs)

Equality constraints (e.g, KL): Partial annotations

Prior-aware loss

KL can be expanded

$$-\sum_{c=0}^{|K|}\{q^c log \frac{1}{N}\sum_{p\in\mathcal{P}} \mathbf{s}_\theta^{p,c} + (1-q^c) log(1-\frac{1}{N}\sum_{p\in\mathcal{P}} \mathbf{s}_\theta^{p,c})\} + const$$

This is problematic (average distribution of p̂ organ sizes inside log!!)

[Zhou et al., Prior-aware Neural Network for Partially-Supervised Multi-Organ Segmentation, ICCV'19]

Constrained optimization (in CNNs)

Equality constraints (e.g, KL): Partial annotations

Stochastic primal-dual gradient

(split terms updated independently)

Prior-aware loss

KL can be expanded

$$-\sum_{c=0}^{|K|}\{q^c log\frac{1}{N}\sum_{p\in\mathcal{P}}\mathbf{s}_\theta^{p,c} + (1-q^c)log(1-\frac{1}{N}\sum_{p\in\mathcal{P}}\mathbf{s}_\theta^{p,c})\} + const$$

This is problematic (average distribution of p̂ organ sizes inside log!!)

[Zhou et al., Prior-aware Neural Network for Partially-Supervised Multi-Organ Segmentation, ICCV'19]

Constrained optimization (in CNNs)

Equality constraints (e.g, KL): Partial annotations

primal

dual

$$-\log \alpha = \max_{\beta} \left(\alpha\beta + 1 + \log(-\beta)\right)$$

Stochastic primal-dual gradient

(split terms updated independently)

Prior-aware loss

KL can be expanded

$$-\sum_{c=0}^{|K|}\{q^c log \frac{1}{N}\sum_{p\in\mathcal{P}} \mathbf{s}_{\theta}^{p,c} + (1-q^c) log(1-\frac{1}{N}\sum_{p\in\mathcal{P}} \mathbf{s}_{\theta}^{p,c})\} + const$$

This is problematic (average distribution of p̂ organ sizes inside log!!)

[Zhou et al., Prior-aware Neural Network for Partially-Supervised Multi-Organ Segmentation, ICCV'19]

Constrained optimization (in CNNs)

Equality constraints (e.g, KL): Partial annotations

$$-\log \bar{p}^l = \max_{\nu^l} \left(\bar{p}^l \nu^l + 1 + \log(-\nu^l) \right)$$
$$-\log(1-\bar{p}^l) = \max_{\mu^l} \left((1-\bar{p}^l)\mu^l + 1 + \log(-\mu^l) \right),$$

Stochastic primal-dual gradient

(split terms updated independently)

Prior-aware loss

KL can be expanded

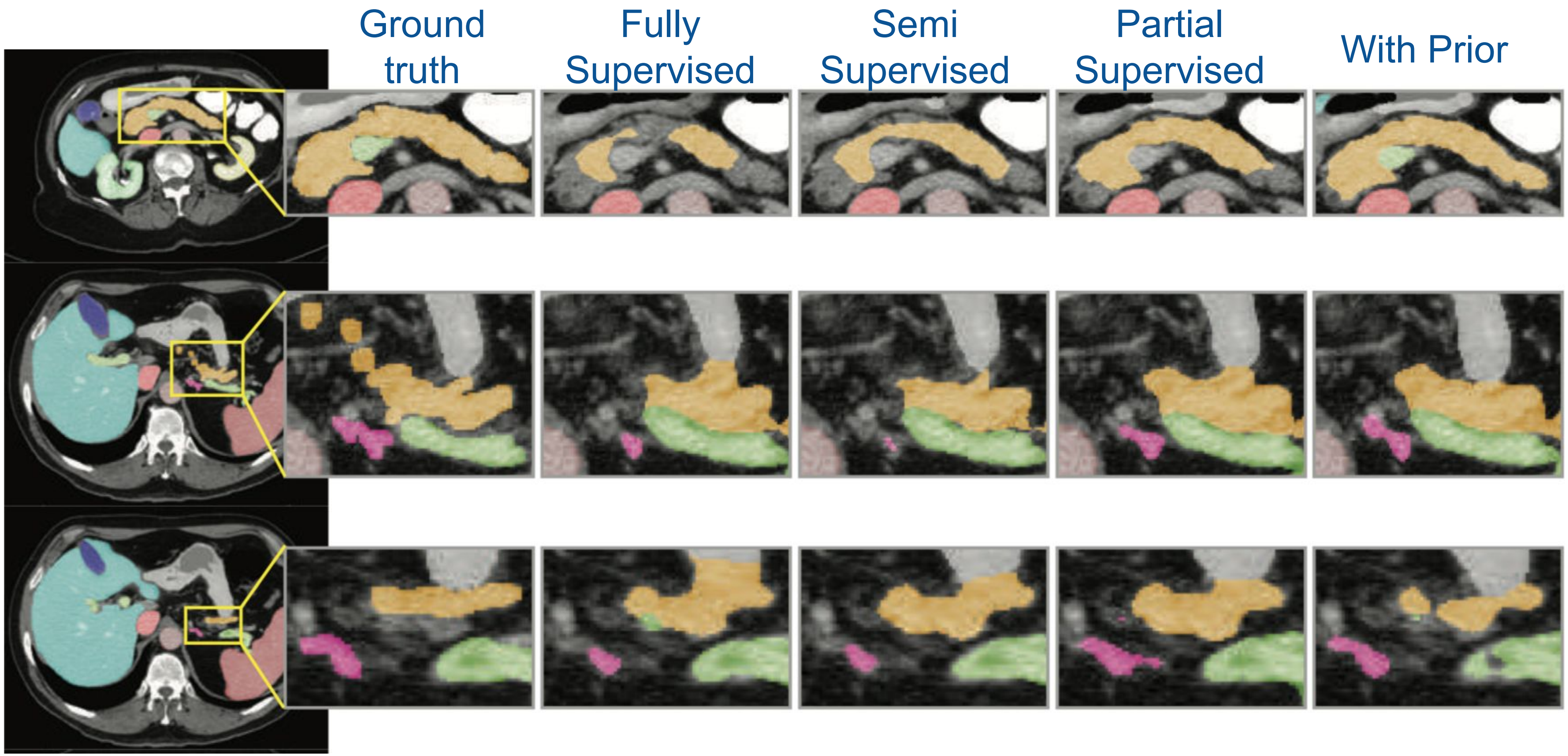
$$-\sum_{c=0}^{|K|}\{q^c log \frac{1}{N}\sum_{p\in\mathcal{P}} \mathbf{s}_\theta^{p,c} + (1-q^c) log(1-\frac{1}{N}\sum_{p\in\mathcal{P}} \mathbf{s}_\theta^{p,c})\} + const$$

This is problematic (average distribution of p̂ organ sizes inside log!!)

[Zhou et al., Prior-aware Neural Network for Partially-Supervised Multi-Organ Segmentation, ICCV’19]

Constrained optimization (in CNNs)

Equality constraints (e.g, KL): Partial annotations

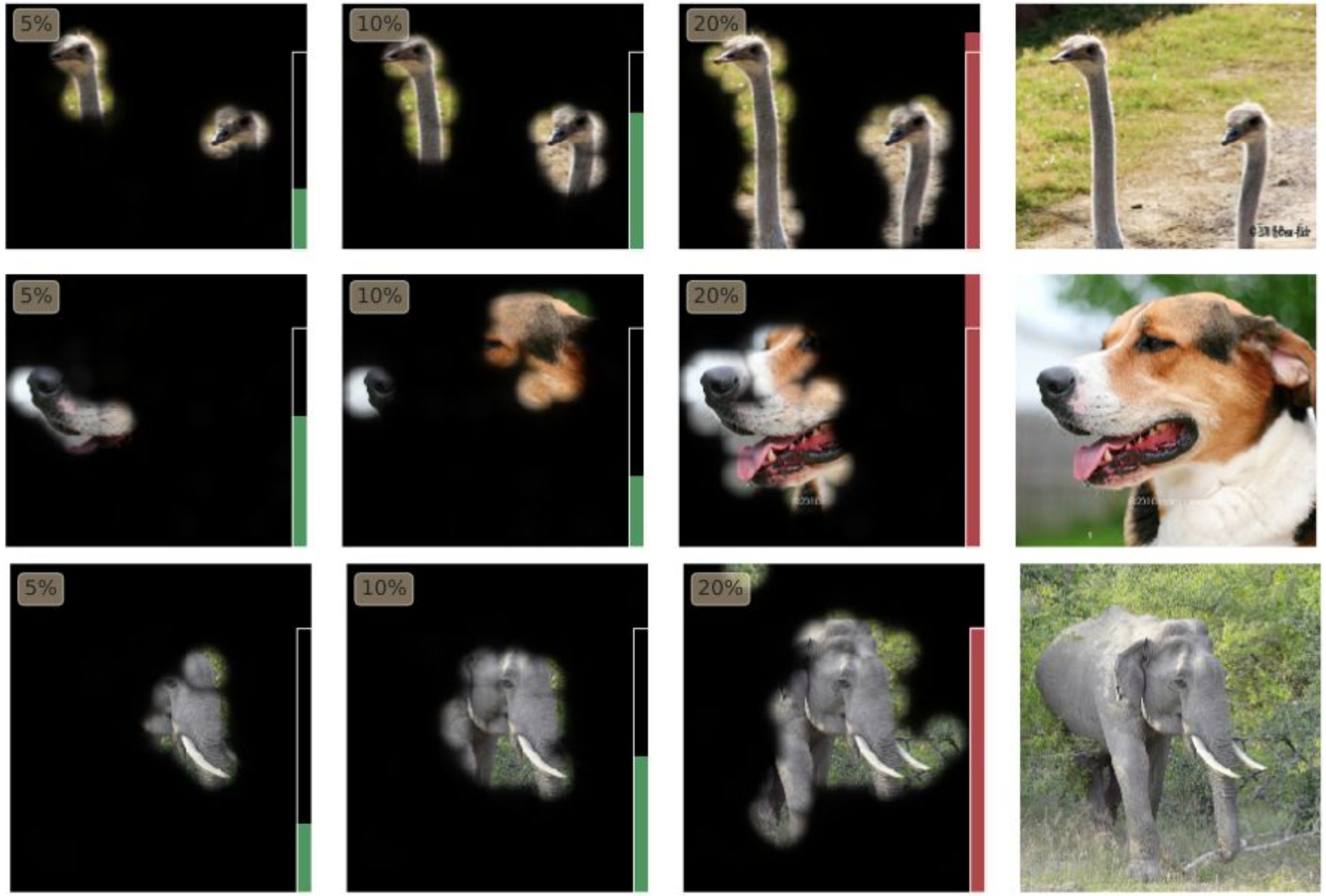


Images from [Zhou et al., Prior-aware Neural Network for Partially-Supervised Multi-Organ Segmentation, ICCV’19]

Constrained optimization (in CNNs)

Equality constraints

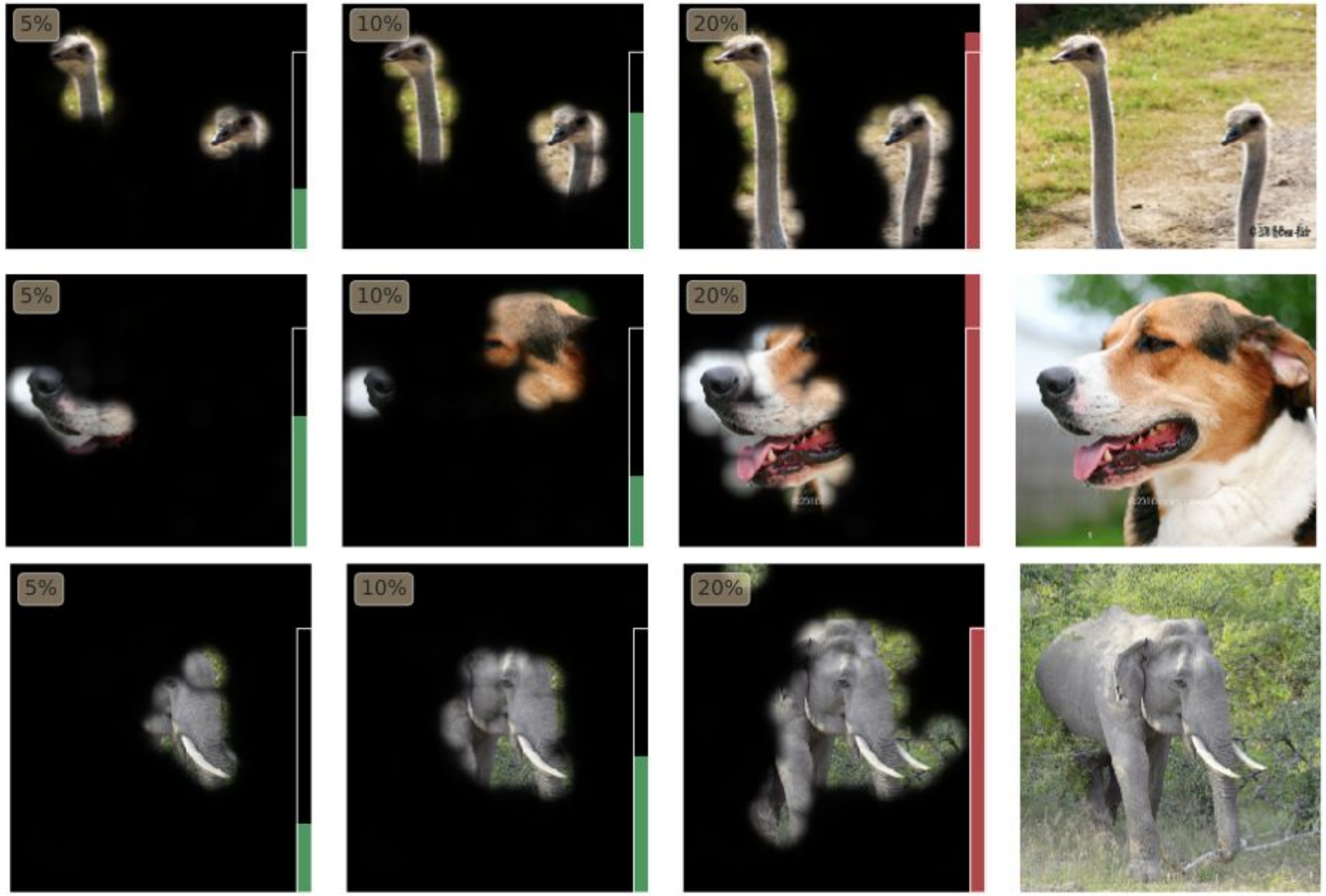
Extremal perturbations



Images from [Fong et al., Understanding Deep Networks via Extremal Perturbations and Smooth Masks, ICCV'19]

Constrained optimization (in CNNs)

Equality constraints



Extremal perturbations

1 - Relax the mask

$$\mathbf{m} \in [0,1]^{|\Omega|}$$

2 - Vectorize and sort **m** (non-decreasing order)

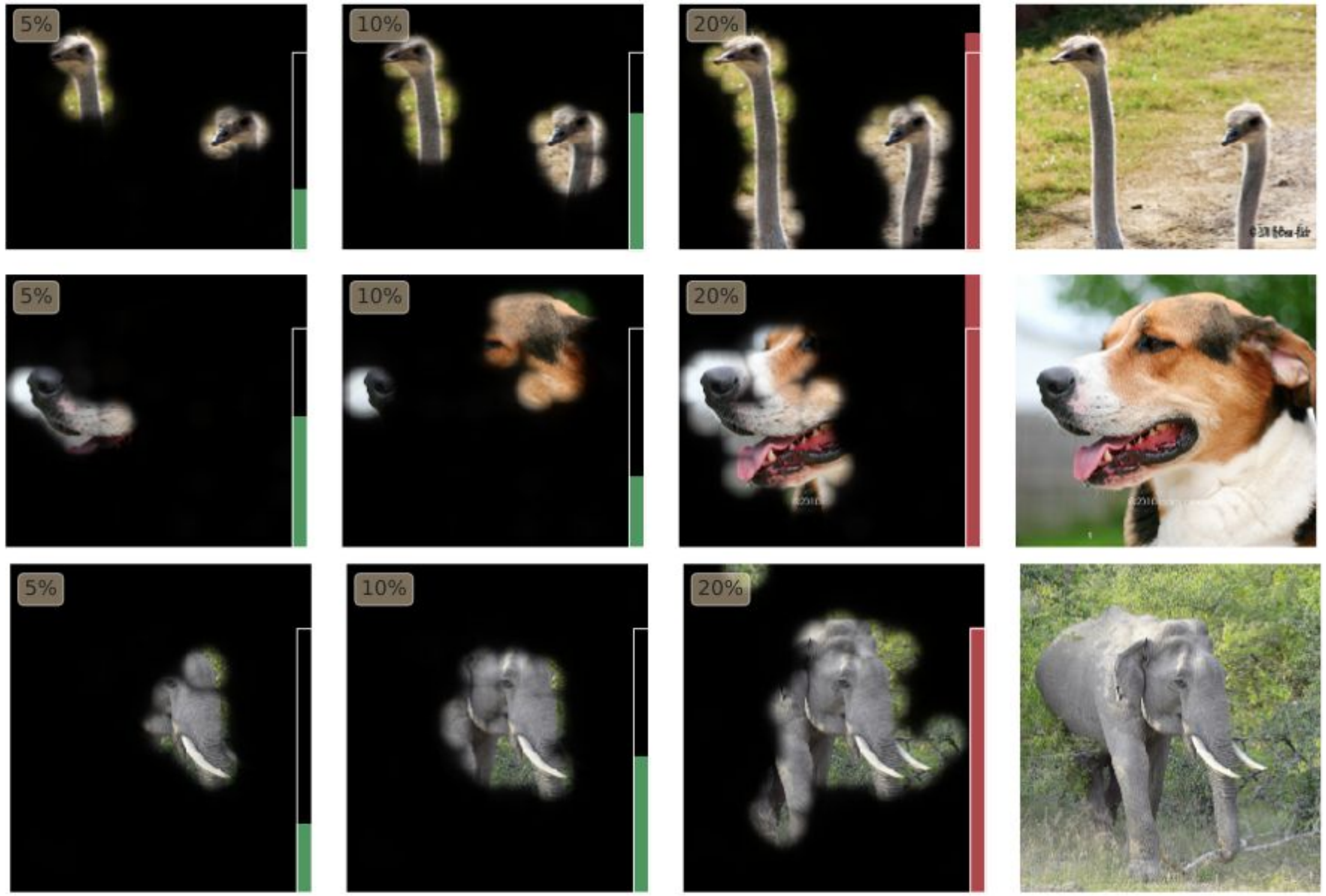
3 - If **m** satisfies the area constraints (a) ***exactly***

$$\mathbf{r}_a \in [0,1]^{|\Omega|} = [\underbrace{0,0,...,0}_{(1-a)|\Omega|},\underbrace{1,1,...,1}_{(a)|\Omega|}]$$

Images from [Fong et al., Understanding Deep Networks via Extremal Perturbations and Smooth Masks, ICCV'19]

Constrained optimization (in CNNs)

Equality constraints



Extremal perturbations

1 - Relax the mask

$$\mathbf{m} \in [0,1]^{|\Omega|}$$

2 - Vectorize and sort **m** (non-decreasing order)

3 - If **m** satisfies the area constraints (a) ***exactly***

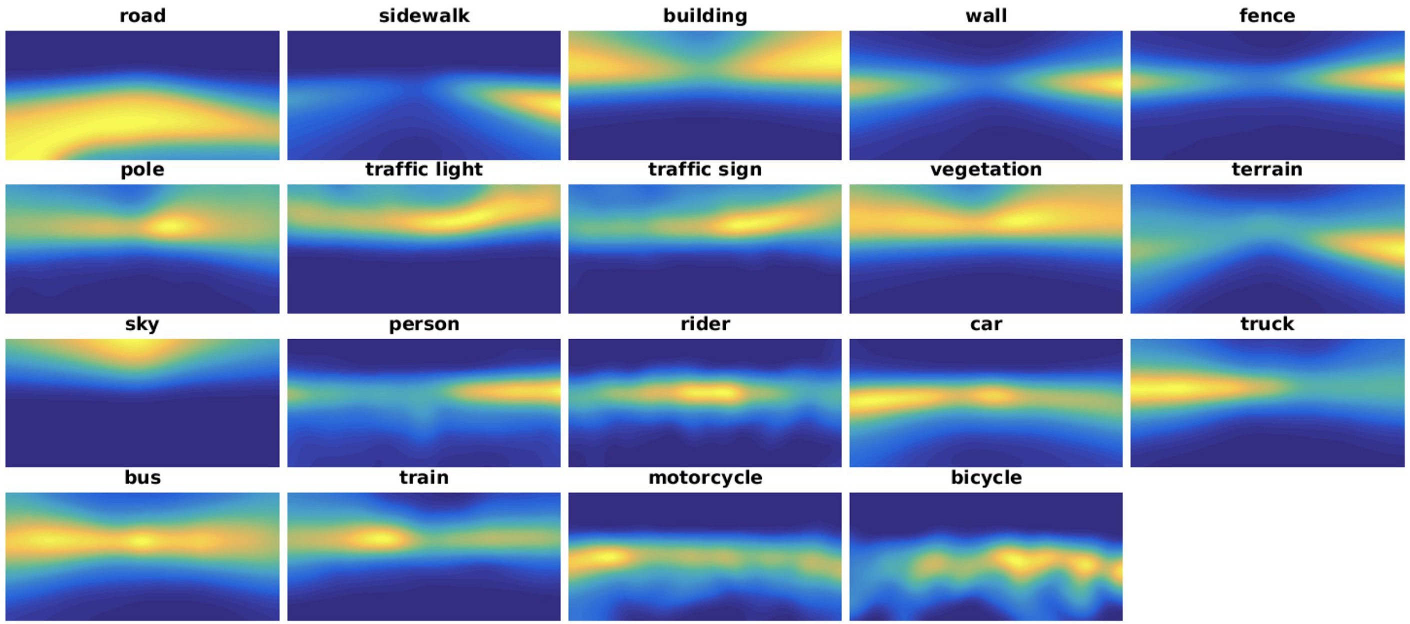
$$\mathbf{r}_a \in [0,1]^{|\Omega|} = [\underbrace{0, 0, \ldots, 0}_{(1-a)|\Omega|}, \underbrace{1, 1, \ldots, 1}_{(a)|\Omega|}]$$

$$\boxed{R_a(\mathbf{m}) = \|vecsort(\mathbf{m}) - r_a\|^2}$$

Images from [Fong et al., Understanding Deep Networks via Extremal Perturbations and Smooth Masks, ICCV'19]

Constrained optimization (in CNNs)

Equality constraints (at pixel-level)

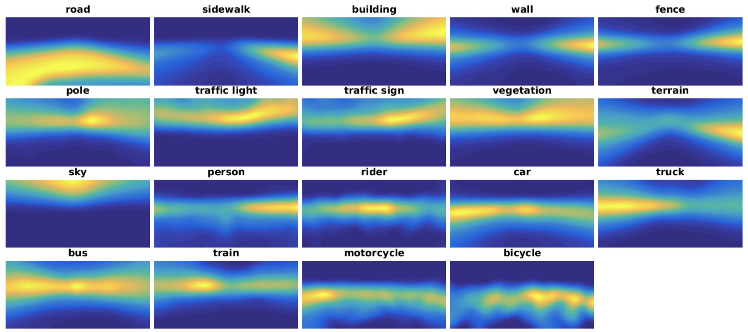


Spatial priors on GTA5

Images from [Zou et al., Unsupervised Domain Adaptation for Semantic Segmentation via Class-Balanced Self-Training, ECCV'18]

Constrained optimization (in CNNs)

Equality constraints (at pixel-level)



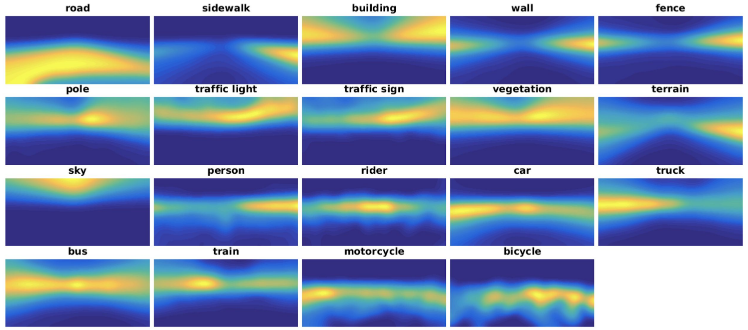
Objective:

$$-\sum_{p \in \mathcal{L}} l(\mathbf{y}^p, s_\theta^p) - \sum_{q \in \mathcal{U}} \sum_{c=1}^{K} (\hat{y}^{q,c} log(prior^{q,c} s_\theta^{q,c}) + w_c \hat{y}^{q,c})$$

Zou et al., Unsupervised Domain Adaptation for Semantic Segmentation via Class-Balanced Self-Training, ECCV'18

Constrained optimization (in CNNs)

Equality constraints (at pixel-level)



Proposals

Objective:

$$-\sum_{p\in\mathcal{L}} l(\mathbf{y}^p, s_\theta^p) - \sum_{q\in\mathcal{U}} \sum_{c=1}^{K} (\hat{y}^{q,c} log(prior^{q,c} s_\theta^{q,c}) + w_c \hat{y}^{q,c})$$

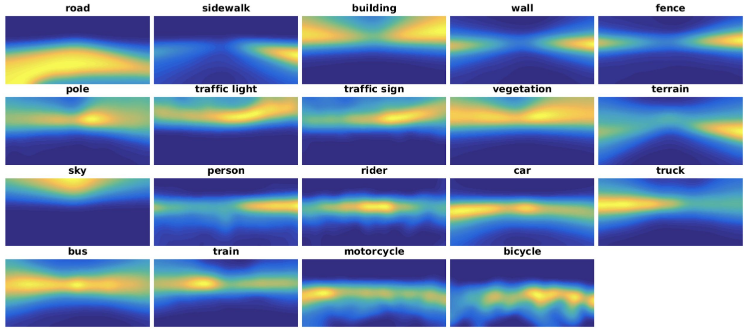
Target images

Source images

Zou et al., Unsupervised Domain Adaptation for Semantic Segmentation via Class-Balanced Self-Training, ECCV'18

Constrained optimization (in CNNs)

Equality constraints (at pixel-level)



Objective:

$$-\sum_{p\in\mathcal{L}} l(\mathbf{y}^p, s_\theta^p) - \sum_{q\in\mathcal{U}}\sum_{c=1}^{K}(\underbrace{\hat{y}^{q,c} log(prior^{q,c} s_\theta^{q,c})} + w_c \hat{y}^{q,c})$$

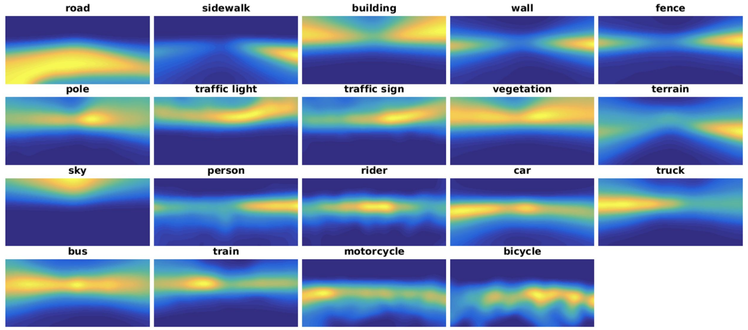
This becomes two KL

$$KL(\hat{y}^{q,c}|prior^{q,c}) \qquad KL(\hat{y}^{q,c}|s_\theta^{q,c})$$

Zou et al., Unsupervised Domain Adaptation for Semantic Segmentation via Class-Balanced Self-Training, ECCV’18

Constrained optimization (in CNNs)

Equality constraints (at pixel-level)



Objective:

$$-\sum_{p\in\mathcal{L}} l(\mathbf{y}^p, s_\theta^p) - \sum_{q\in\mathcal{U}}\sum_{c=1}^{K}(\hat{y}^{q,c} log(prior^{q,c} s_\theta^{q,c}) + w_c \hat{y}^{q,c})$$

Weights the proposals

Zou et al., Unsupervised Domain Adaptation for Semantic Segmentation via Class-Balanced Self-Training, ECCV’18

Constrained optimization (in CNNs)

Equality constraints (at pixel-level)

Spatial prior

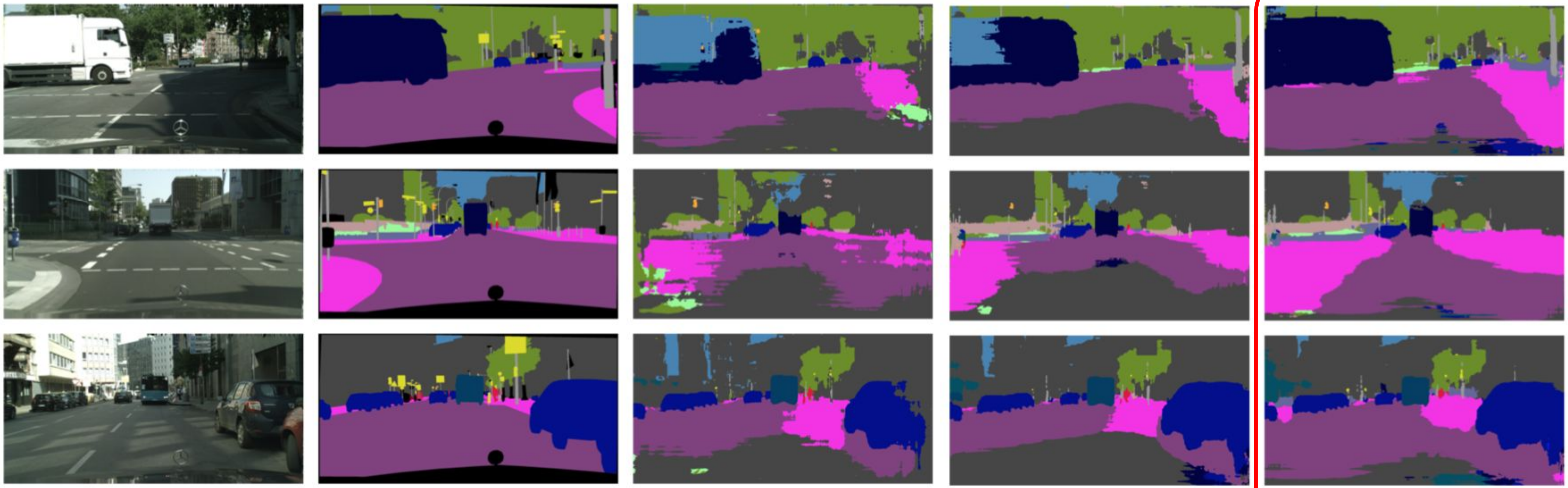


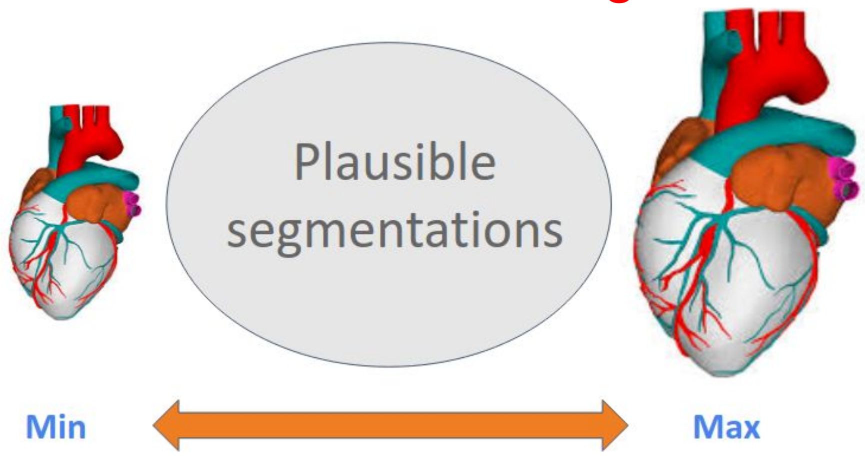
Image | Ground Truth | Only source | CBST | CBST-SP

Images from [Zou et al., Unsupervised Domain Adaptation for Semantic Segmentation via Class-Balanced Self-Training, ECCV’18]

Constrained optimization (in CNNs)

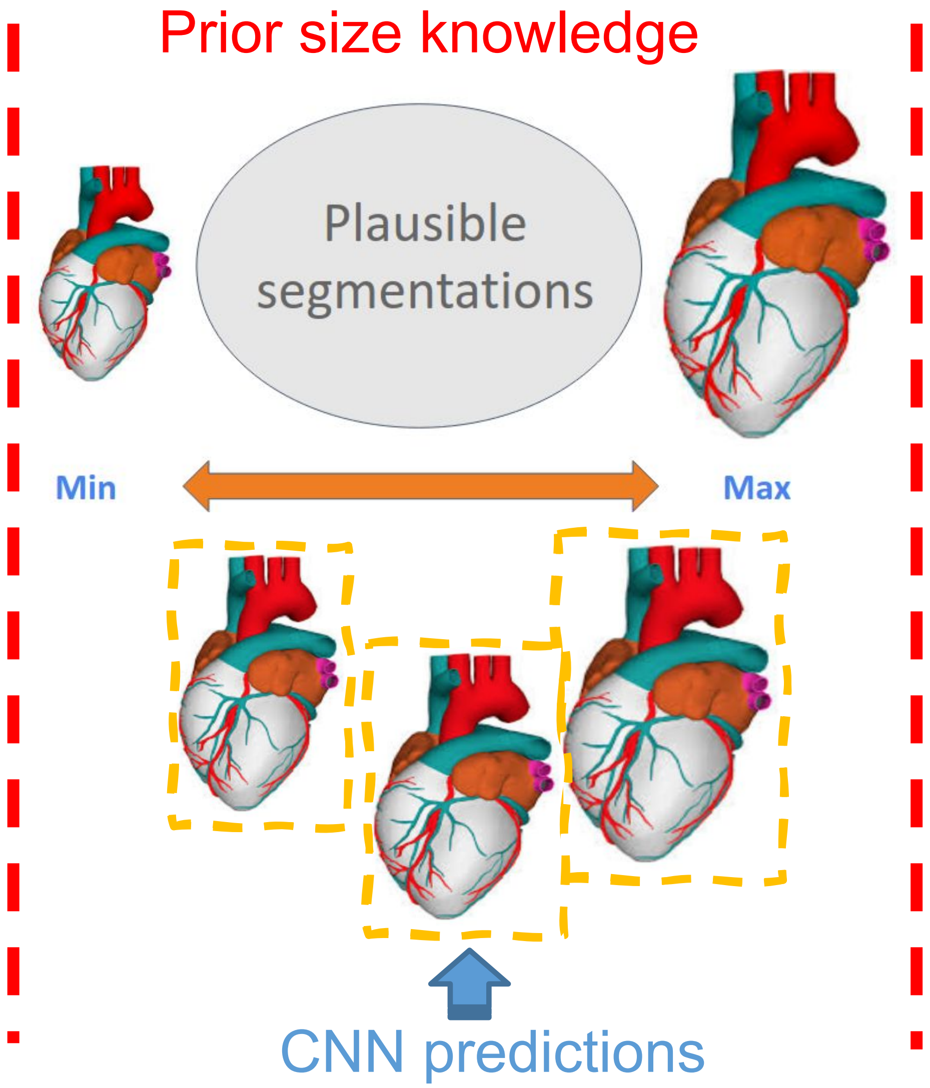
Inequality constraints

Prior size knowledge



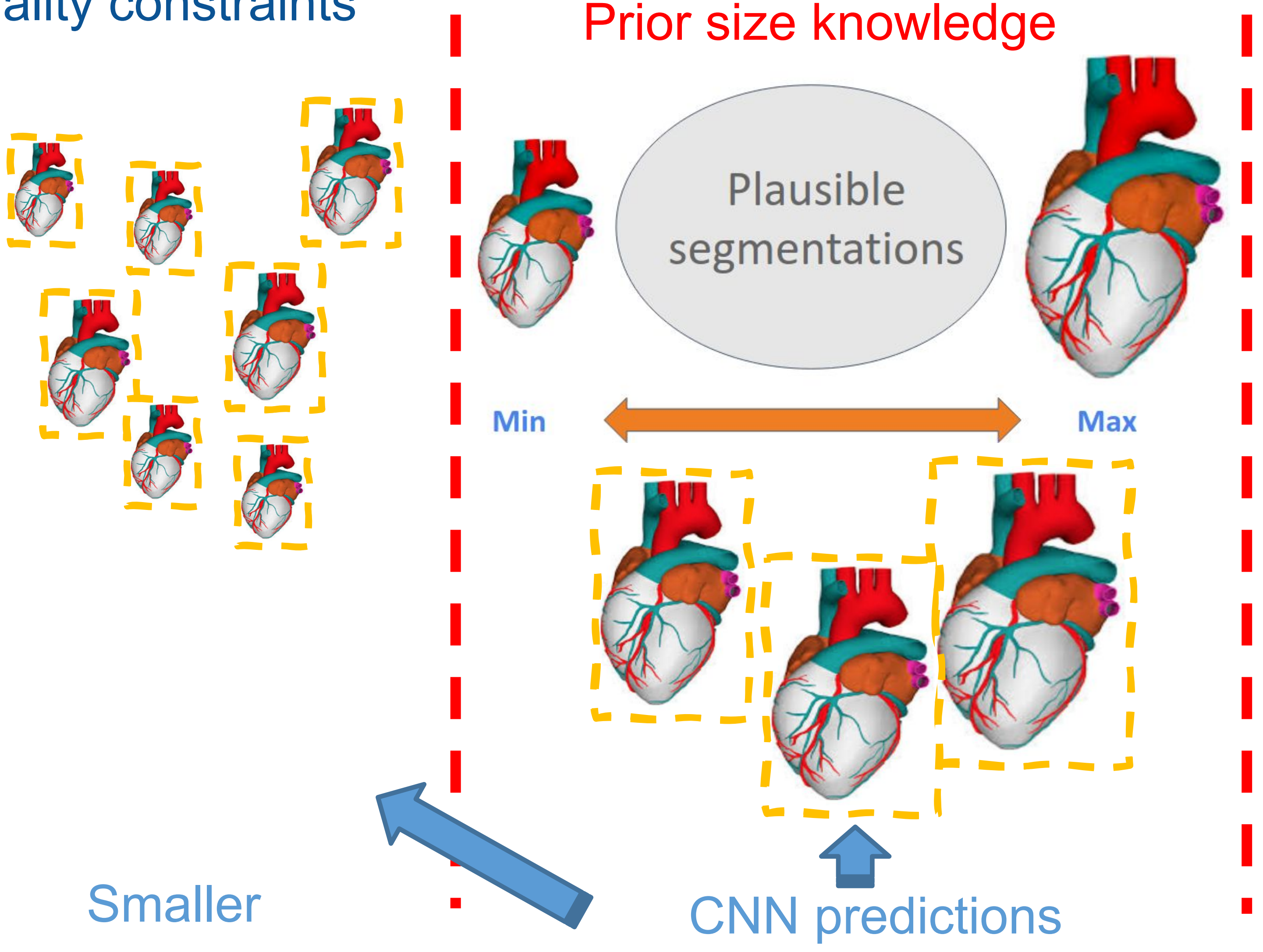
Constrained optimization (in CNNs)

Inequality constraints

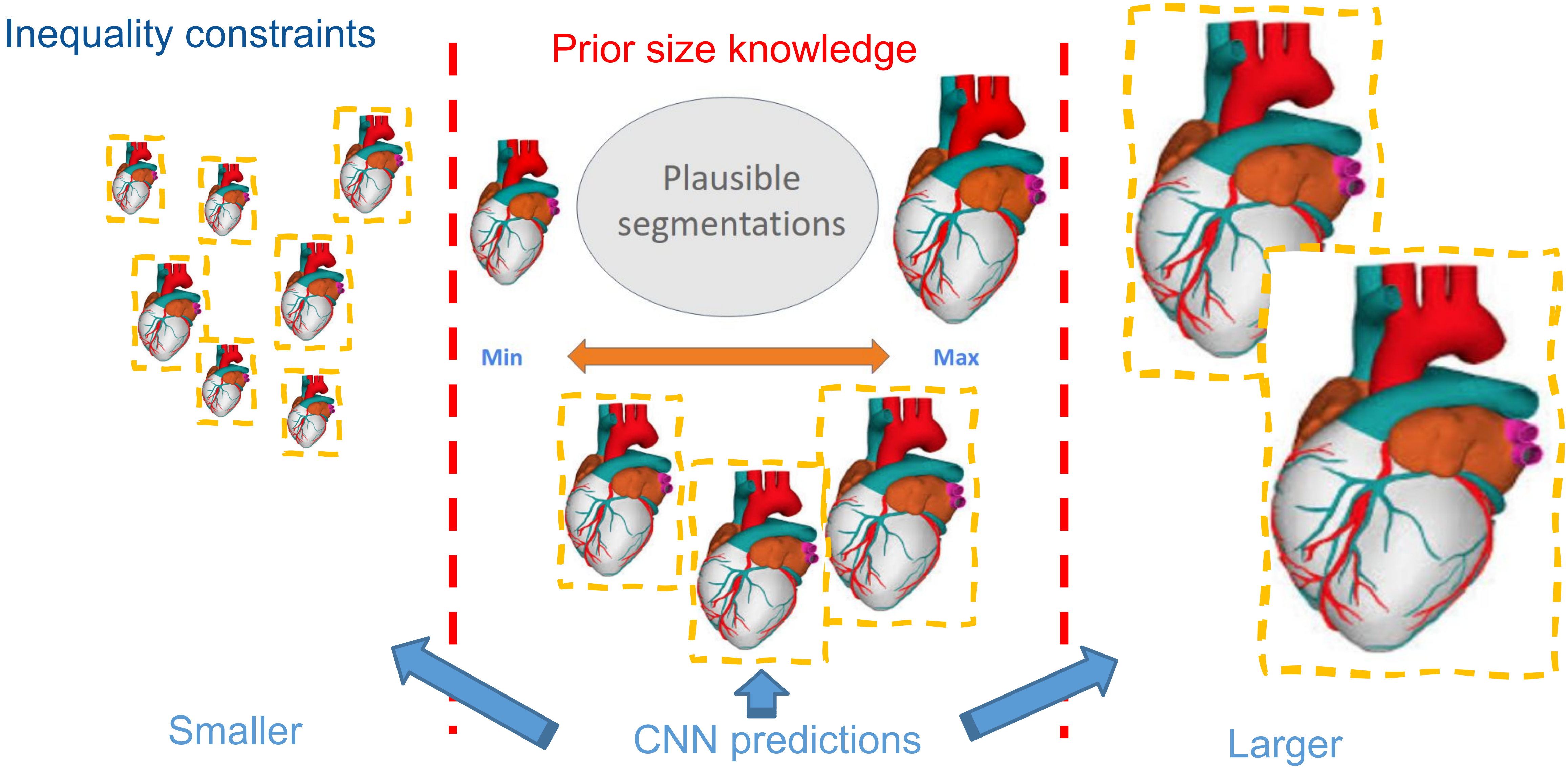


Constrained optimization (in CNNs)

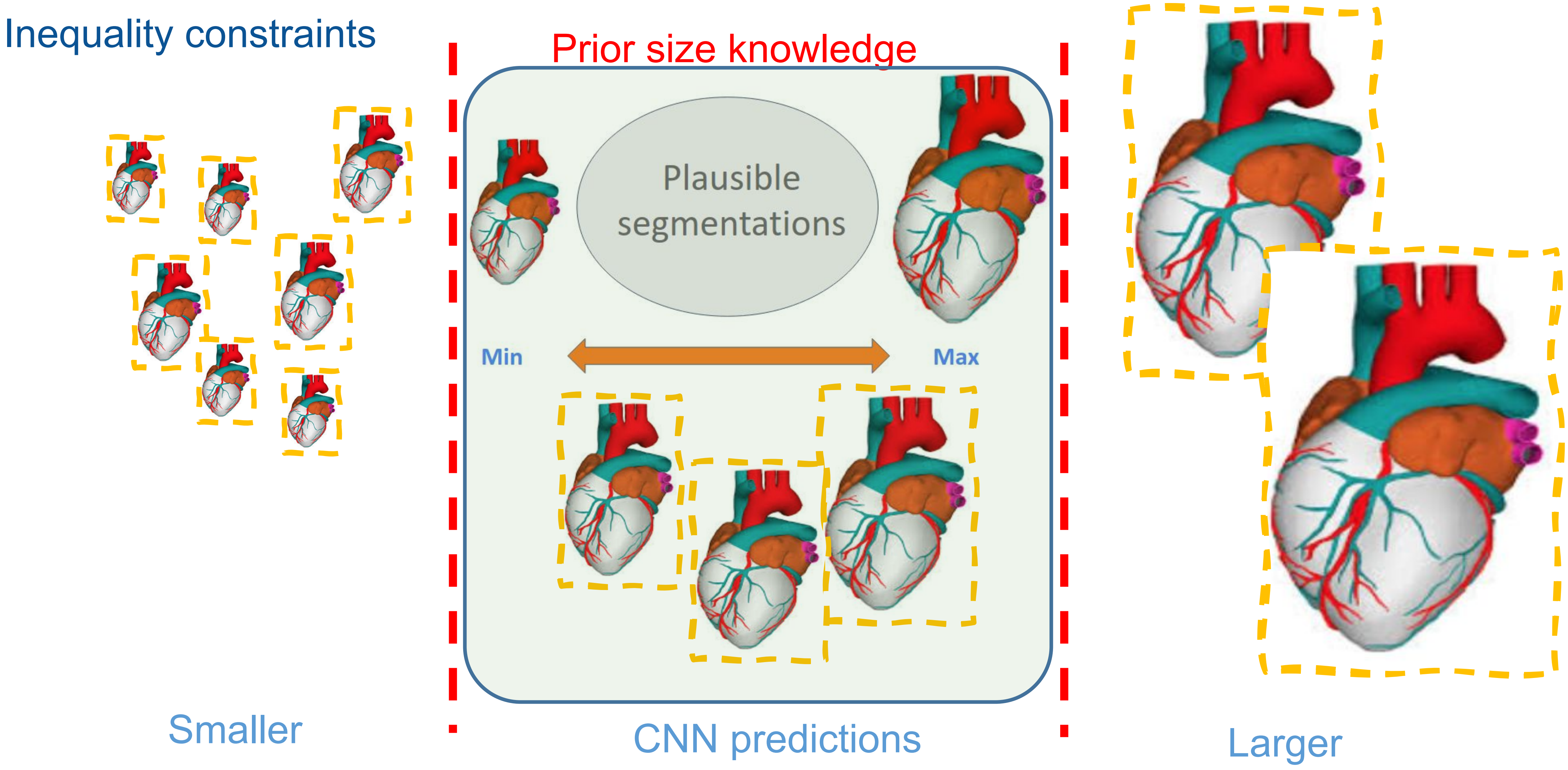
Inequality constraints



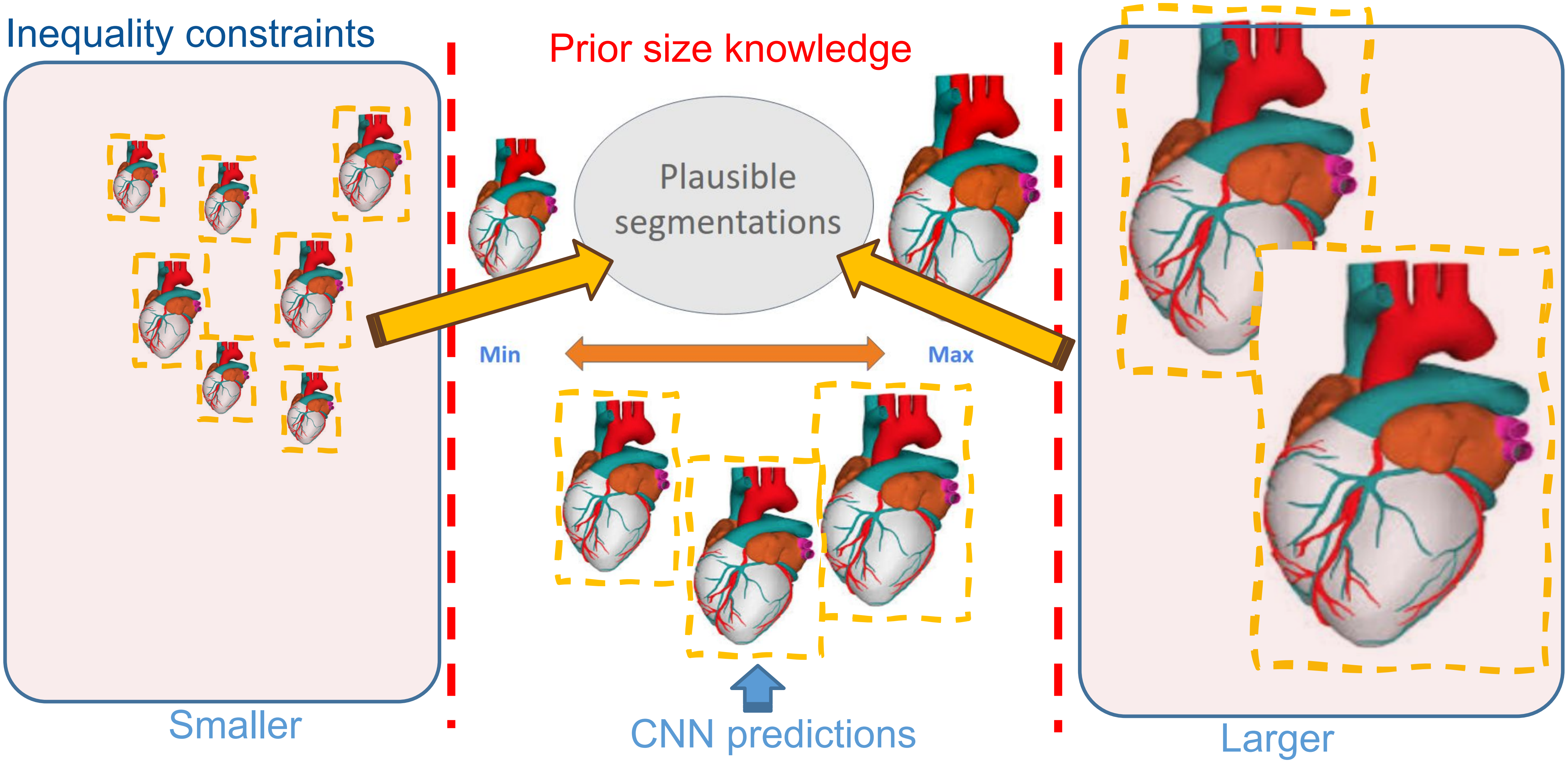
Constrained optimization (in CNNs)



Constrained optimization (in CNNs)

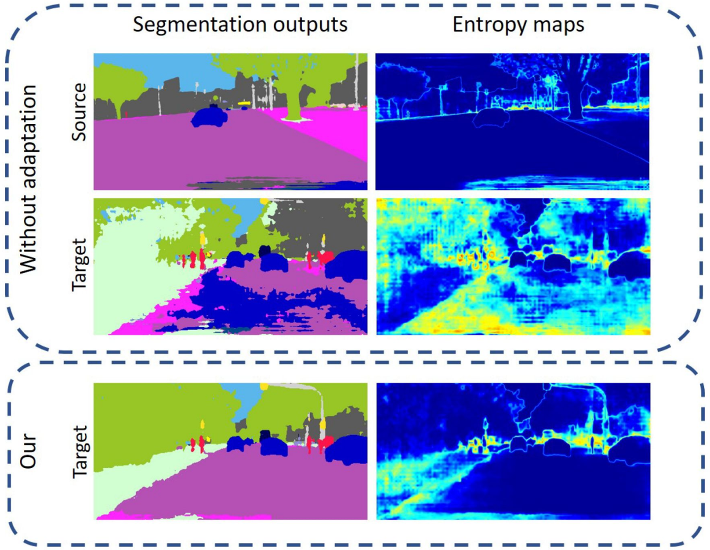


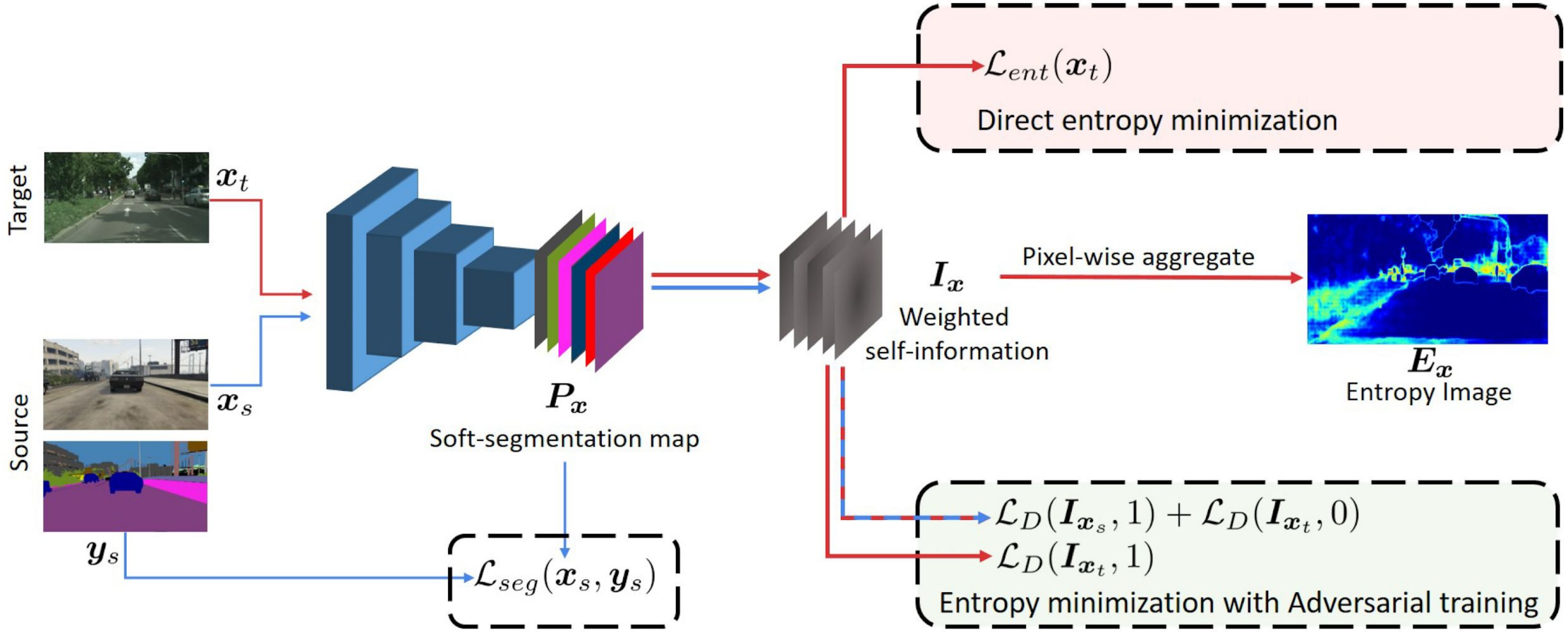
Constrained optimization (in CNNs)



Constrained optimization (in CNNs)

Inequality constraints

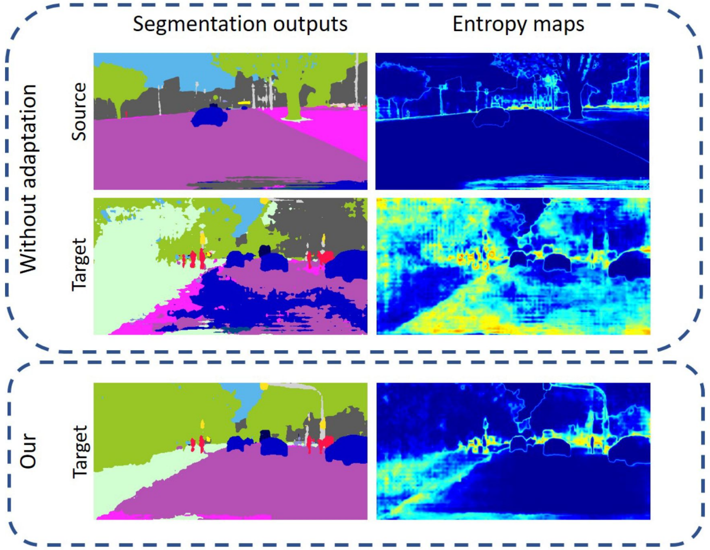


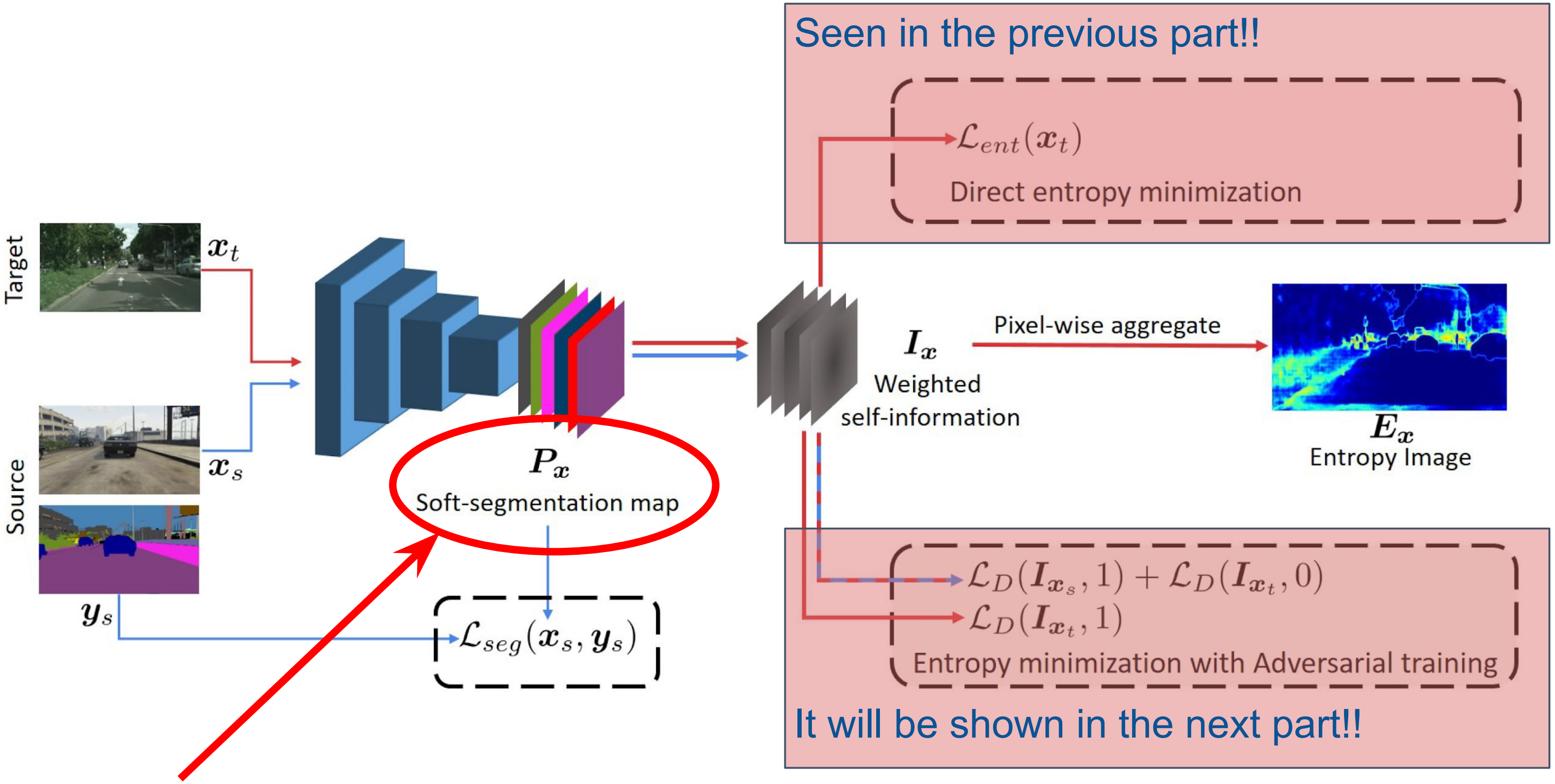


Images from [Vu et al., ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation. CVPR'19]

Constrained optimization (in CNNs)

Inequality constraints



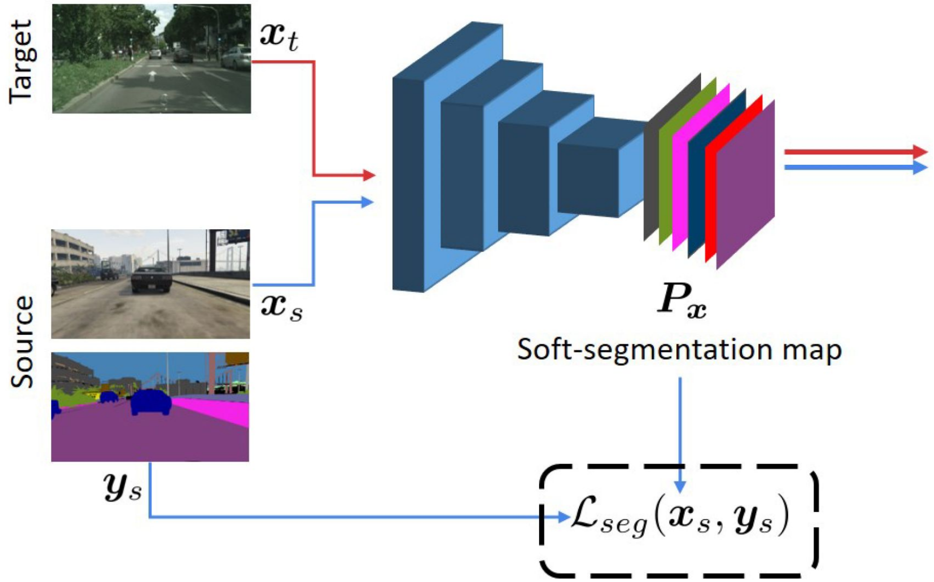


We focus on this now

Images from [Vu et al., ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation. CVPR'19]

Constrained optimization (in CNNs)

Inequality constraints



Class-ratio priors

$$\mathcal{L}_{cp}(\boldsymbol{x}_t) = \sum_{c=1}^{C} \max\left(0, \mu \boldsymbol{p}_s^{(c)} - \mathbb{E}_c(\boldsymbol{P}_{\boldsymbol{x}_t}^{(c)})\right)$$

It relaxes the class prior constraint

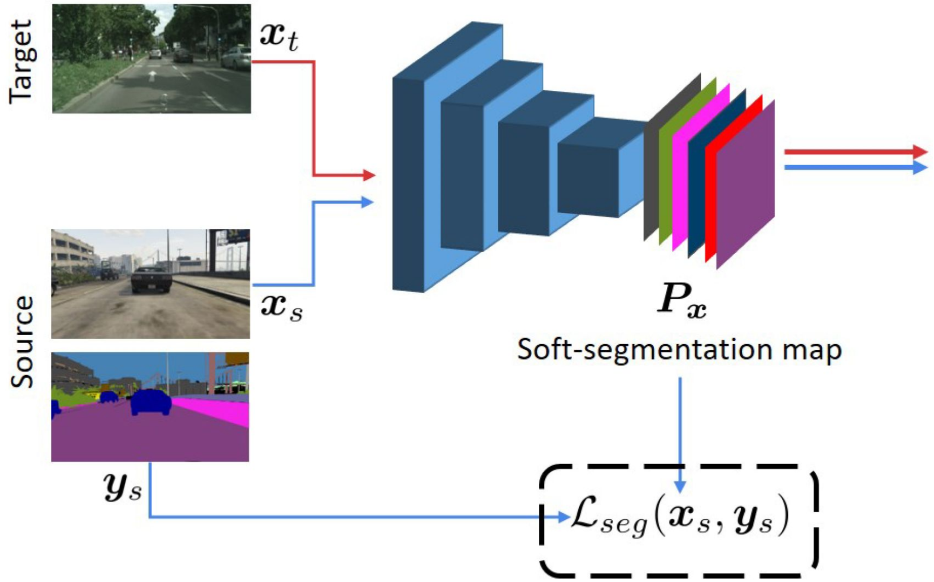
$\ell 1$-normalized histogram (source)

Estimated size on the prediction

Vu et al., ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation. CVPR'19

Constrained optimization (in CNNs)

Inequality constraints



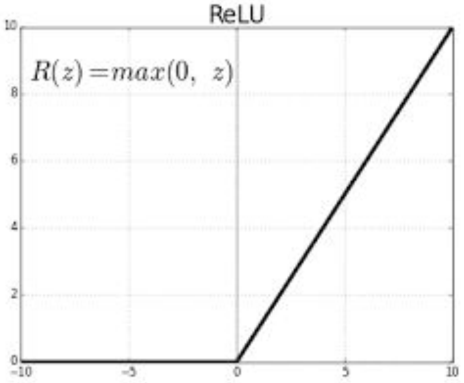
Class-ratio priors

$$\mathcal{L}_{cp}(\boldsymbol{x}_t) = \sum_{c=1}^{C} \max\left(0, \mu \boldsymbol{p}_s^{(c)} - \mathbb{E}_c(\boldsymbol{P}_{\boldsymbol{x}_t}^{(c)})\right)$$

It relaxes the class prior constraint

ℓ1-normalized histogram (source)

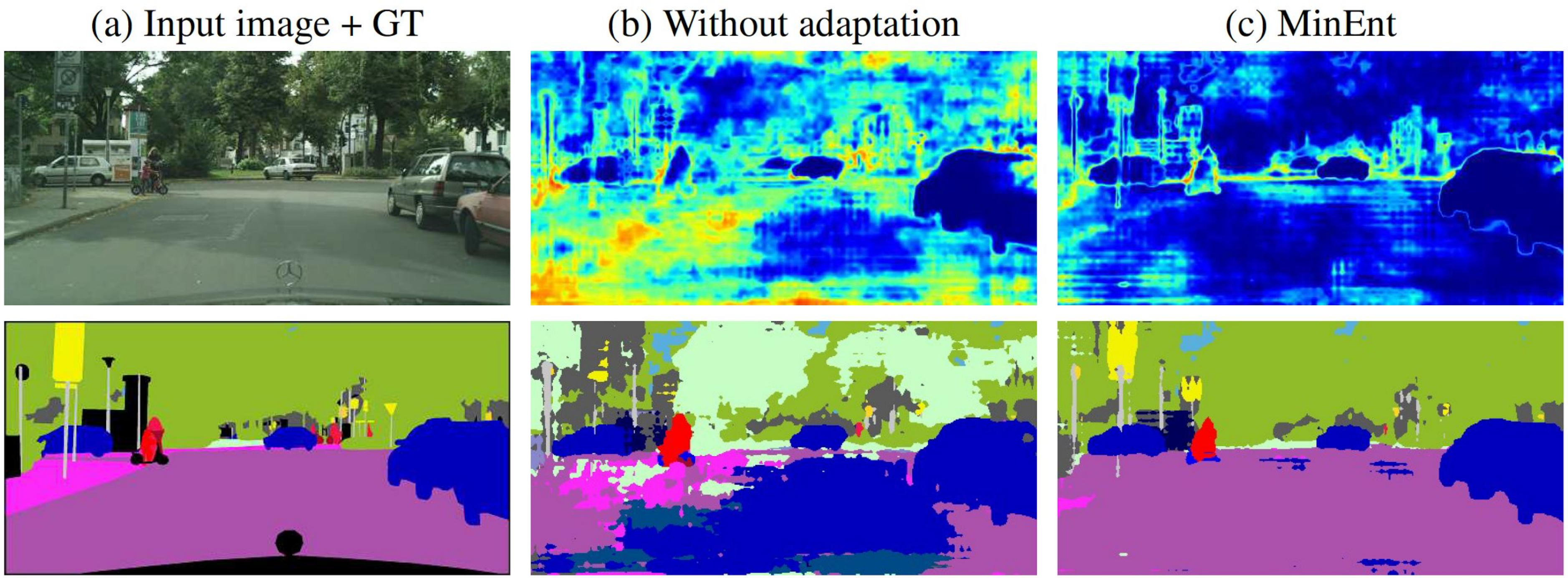
Estimated size on the prediction



Vu et al., ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation. CVPR'19

Constrained optimization (in CNNs)

Inequality constraints



Images from [Vu et al., ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation. CVPR'19]

Constrained optimization (in CNNs)

Inequality constraints

Information is given in the form of image-tags

Suppression $\sum_{p \in \Omega} s_\theta^{p,c} \leq 0 \quad \forall c \notin C$

"Person"

[Pathak et al., Constrained Convolutional Neural Networks for Weakly Supervised Segmentation, ICCV'15]

Constrained optimization (in CNNs)

Inequality constraints

Information is given in the form of image-tags

Inclusion (or existence)

$$\sum_{p\in\Omega} s_\theta^{p,c} \geq 1 \quad \forall c \in C$$

“Car”

[Pathak et al., Constrained Convolutional Neural Networks for Weakly Supervised Segmentation, ICCV’15]

Constrained optimization (in CNNs)

Inequality constraints

Information is given in the form of image-tags

Target Size
a > 1

$$\sum_{p \in \Omega} s_\theta^{p,c} \geq a \quad \forall c \in C$$

“Car”

[Pathak et al., Constrained Convolutional Neural Networks for Weakly Supervised Segmentation, ICCV’15]

Constrained optimization (in CNNs)

How we can benefit from this in the medical domain?

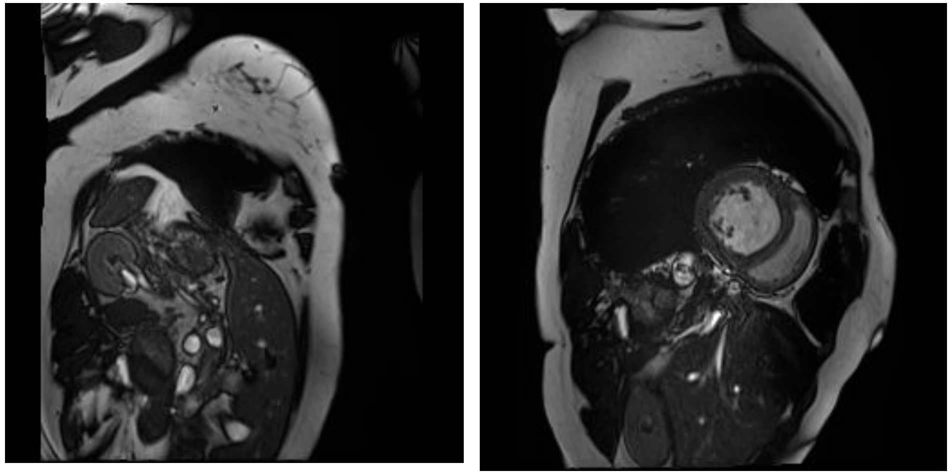


Image-tag information

$$\sum_{p \in \Omega} s_\theta^{p,c} \leq 0$$

For negative image tags

[Kervadec et al., Constrained-CNN losses for weakly supervised segmentation. MedIA 2019]

Constrained optimization (in CNNs)

How we can benefit from this in the medical domain?

No cavity | Cavity

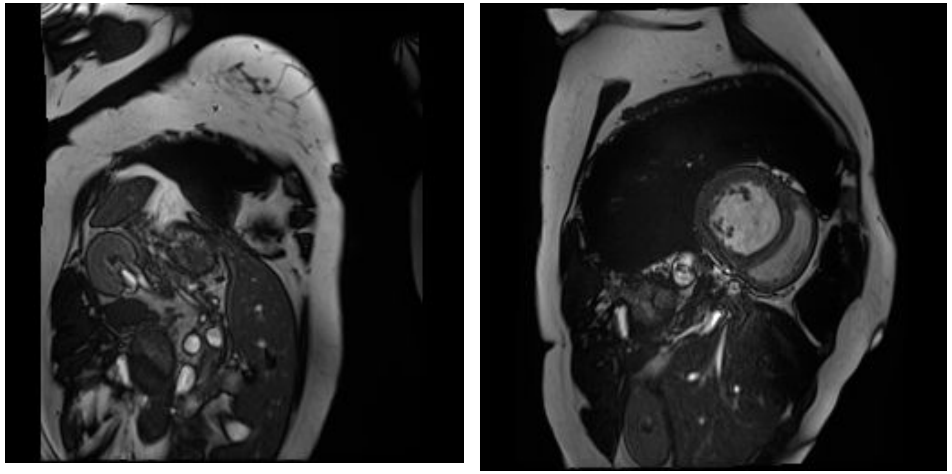
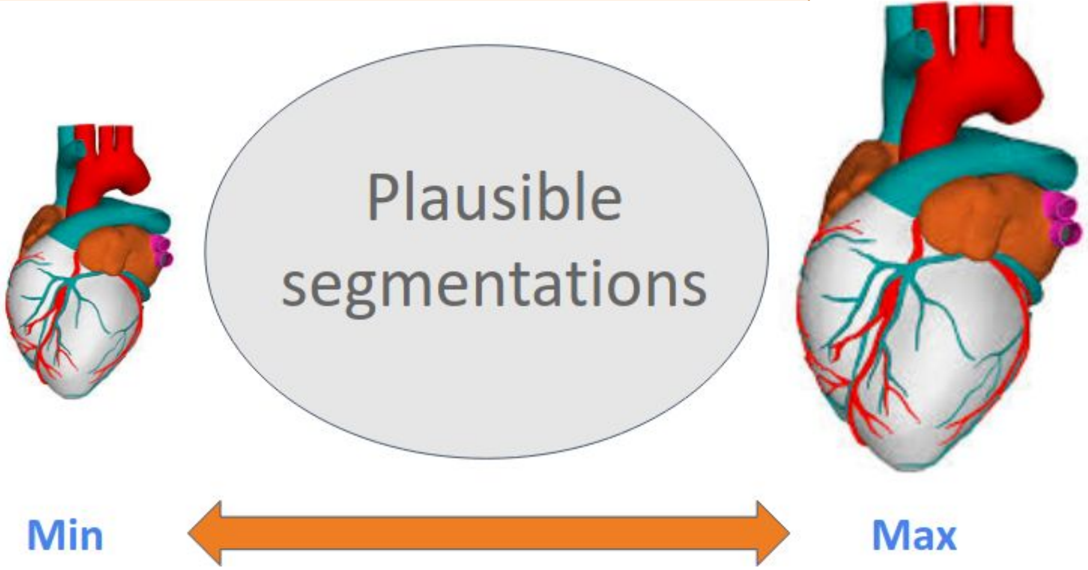


Image-tag information

$$\sum_{p \in \Omega} s_{\theta}^{p,c} \leq 0$$

For negative image tags



Size information

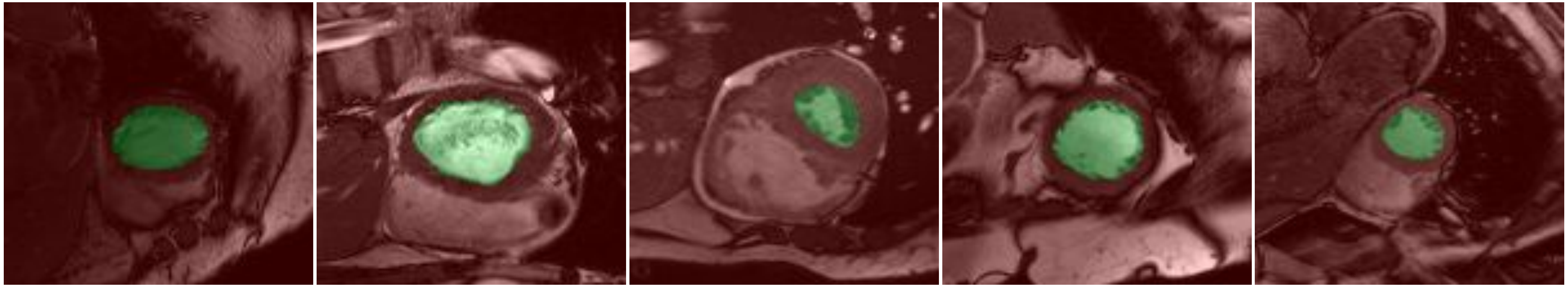
$$min \leq \sum_{p \in \Omega} s_{\theta}^{p,c} \leq max$$

For positive image tags

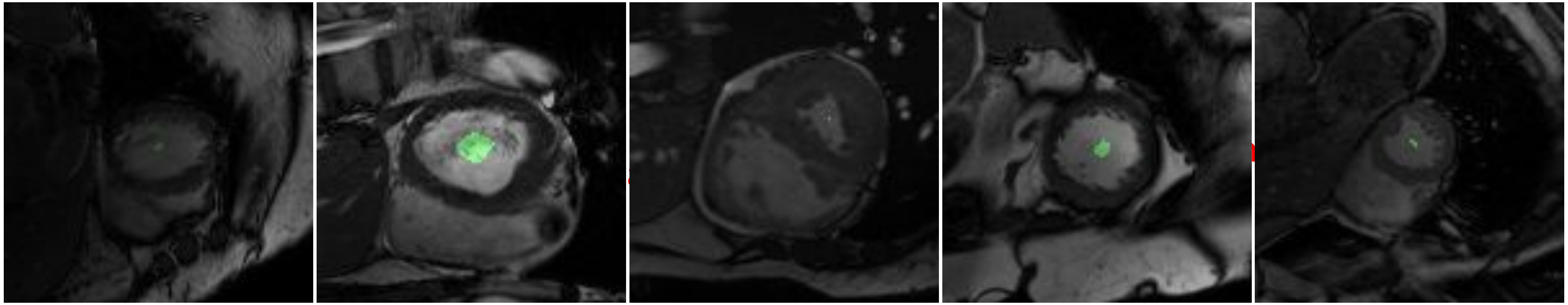
[Kervadec et al., Constrained-CNN losses for weakly supervised segmentation. MedIA 2019]

Constrained optimization (in CNNs)

Inequality constraints (e.g, L2 penalty)



Full annotations



Partial annotations for cross-entropy

[Kervadec et al., Constrained-CNN losses for weakly supervised segmentation. MedIA 2019]

Constrained optimization (in CNNs)

Inequality constraints (e.g, L2 penalty)

Objective

$$\min_{\boldsymbol{\theta}} \ \mathcal{H}(S) \quad \text{s.t} \quad a \leq \sum_{p \in \Omega} s_{\theta}^{p,c} \leq b \quad \Rightarrow \quad \mathcal{H}(S) + \lambda \mathcal{C}\left(V_S\right)$$

[Kervadec et al., Constrained-CNN losses for weakly supervised segmentation. MedIA 2019]

Constrained optimization (in CNNs)

Inequality constraints (e.g, L2 penalty)

Objective

$$\min_{\boldsymbol{\theta}} \mathcal{H}(S) \quad \text{s.t} \quad a \leq \sum_{p \in \Omega} s_\theta^{p,c} \leq b \Rightarrow \mathcal{H}(S) + \lambda \mathcal{C}\left(V_S\right)$$

$$\mathcal{H}(S) = -\sum_{p \in \mathcal{L}} log(s_\theta^p)$$

On annotated pixels

[Kervadec et al., Constrained-CNN losses for weakly supervised segmentation. MedIA 2019]

Constrained optimization (in CNNs)

Inequality constraints (e.g, L2 penalty)

Objective

$$\min_{\boldsymbol{\theta}} \mathcal{H}(S) \quad \text{s.t} \quad a \leq \sum_{p \in \Omega} s_{\theta}^{p,c} \leq b \Rightarrow \mathcal{H}(S) + \lambda \mathcal{C}(V_S)$$

$$\mathcal{H}(S) = -\sum_{p \in \mathcal{L}} log(s_{\theta}^{p})$$

On annotated pixels

$$\mathcal{C}(V_S) = \begin{cases} (V_S - a)^2, & \text{if } V_S < a \\ (V_S - b)^2, & \text{if } V_S > b \\ 0, & \text{otherwise} \end{cases}$$

[Kervadec et al., Constrained-CNN losses for weakly supervised segmentation. MedIA 2019]

Constrained optimization (in CNNs)

Inequality constraints (e.g, L2 penalty)

Objective

$$\min_{\boldsymbol{\theta}} \ \mathcal{H}(S) \quad \text{s.t} \quad a \leq \sum_{p \in \Omega} s_{\theta}^{p,c} \leq b \quad \Rightarrow \quad \mathcal{H}(S) + \lambda \mathcal{C}\left(V_S\right)$$

Back-propagation

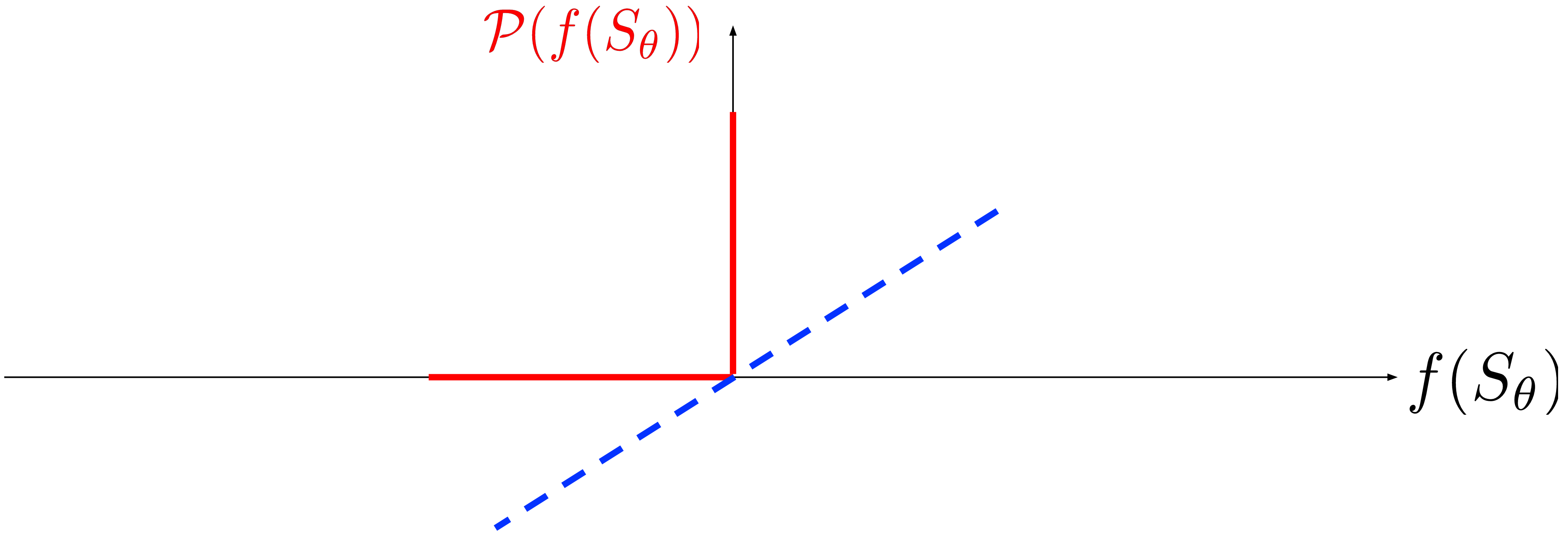
$$-\frac{\partial \mathcal{C}(V_S)}{\partial \boldsymbol{\theta}} \propto \begin{cases} (a - V_S) \frac{\partial S_p}{\partial \boldsymbol{\theta}}, & \text{if } V_S < a \\ (b - V_S) \frac{\partial S_p}{\partial \boldsymbol{\theta}}, & \text{if } V_S > b \\ 0, & \text{otherwise} \end{cases}$$

[Kervadec et al., Constrained-CNN losses for weakly supervised segmentation. MedIA 2019]

Constrained optimization (in CNNs)

Inequality constraints (Why not Lagrangian primal/dual?)

$$\mathcal{L}(S_\theta, \lambda) = \mathcal{E}(\theta) + \lambda f(S_\theta)$$

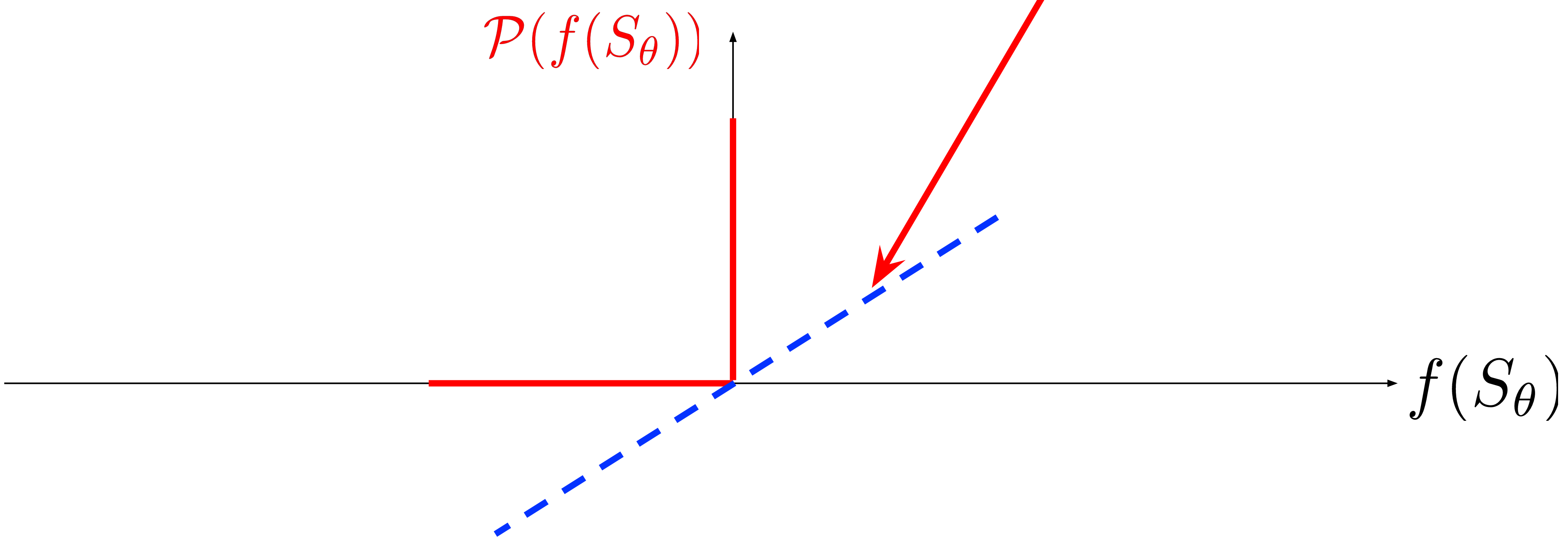


[Boyd and Vandenberghe, Convex optimization, 2004]

Constrained optimization (in CNNs)

Inequality constraints (Why not Lagrangian primal/dual?)

$$\mathcal{L}(S_\theta, \lambda) = \mathcal{E}(\theta) + \lambda f(S_\theta)$$

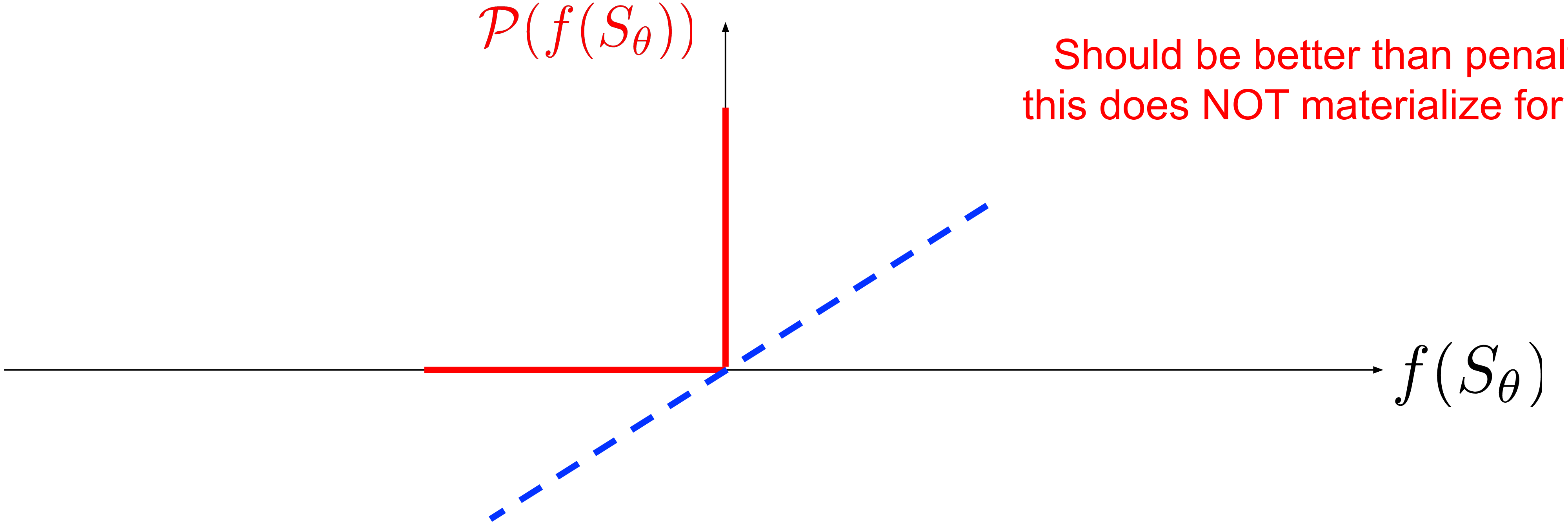


[Boyd and Vandenberghe, Convex optimization, 2004]

Constrained optimization (in CNNs)

Inequality constraints (Why not Lagrangian primal/dual?)

$$\mathcal{L}(S_\theta, \lambda) = \mathcal{E}(\theta) + \lambda f(S_\theta)$$



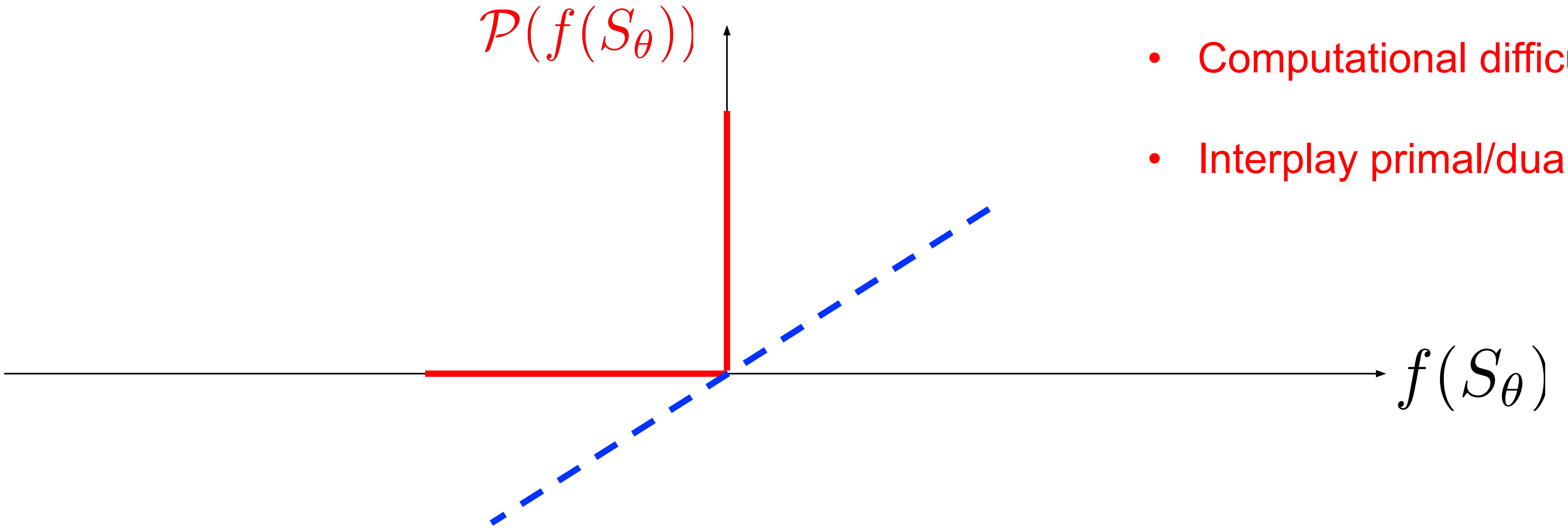
Should be better than penalties, but this does NOT materialize for CNNs!!!!!

[Boyd and Vandenberghe, Convex optimization, 2004]

Constrained optimization (in CNNs)

Inequality constraints (Why not Lagrangian primal/dual?)

$$\mathcal{L}(S_\theta, \lambda) = \mathcal{E}(\theta) + \lambda f(S_\theta)$$



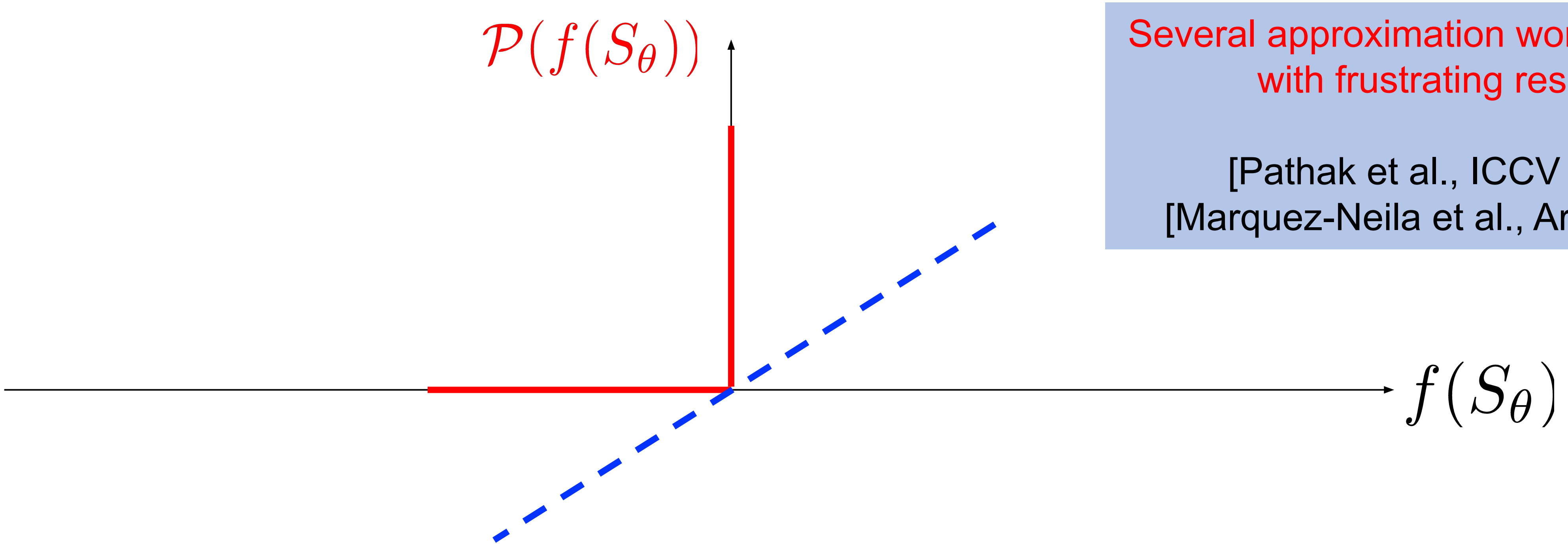
- Computational difficulty
- Interplay primal/dual

[Boyd and Vandenberghe, Convex optimization, 2004]

Constrained optimization (in CNNs)

Inequality constraints (Why not Lagrangian primal/dual?)

$$\mathcal{L}(S_\theta, \lambda) = \mathcal{E}(\theta) + \lambda f(S_\theta)$$



Several approximation works recently, with frustrating results:

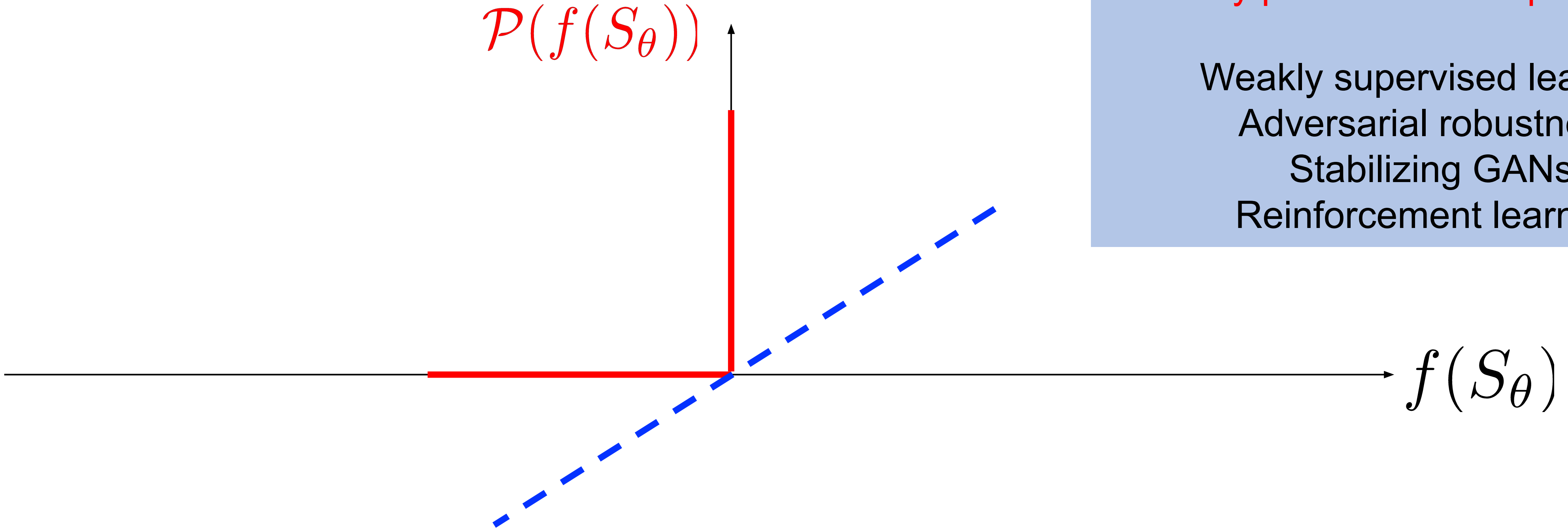
[Pathak et al., ICCV 2015]
[Marquez-Neila et al., ArXiv 2017]

[Boyd and Vandenberghe, Convex optimization, 2004]

Constrained optimization (in CNNs)

Inequality constraints (Why not Lagrangian primal/dual?)

$$\mathcal{L}(S_\theta, \lambda) = \mathcal{E}(\theta) + \lambda f(S_\theta)$$



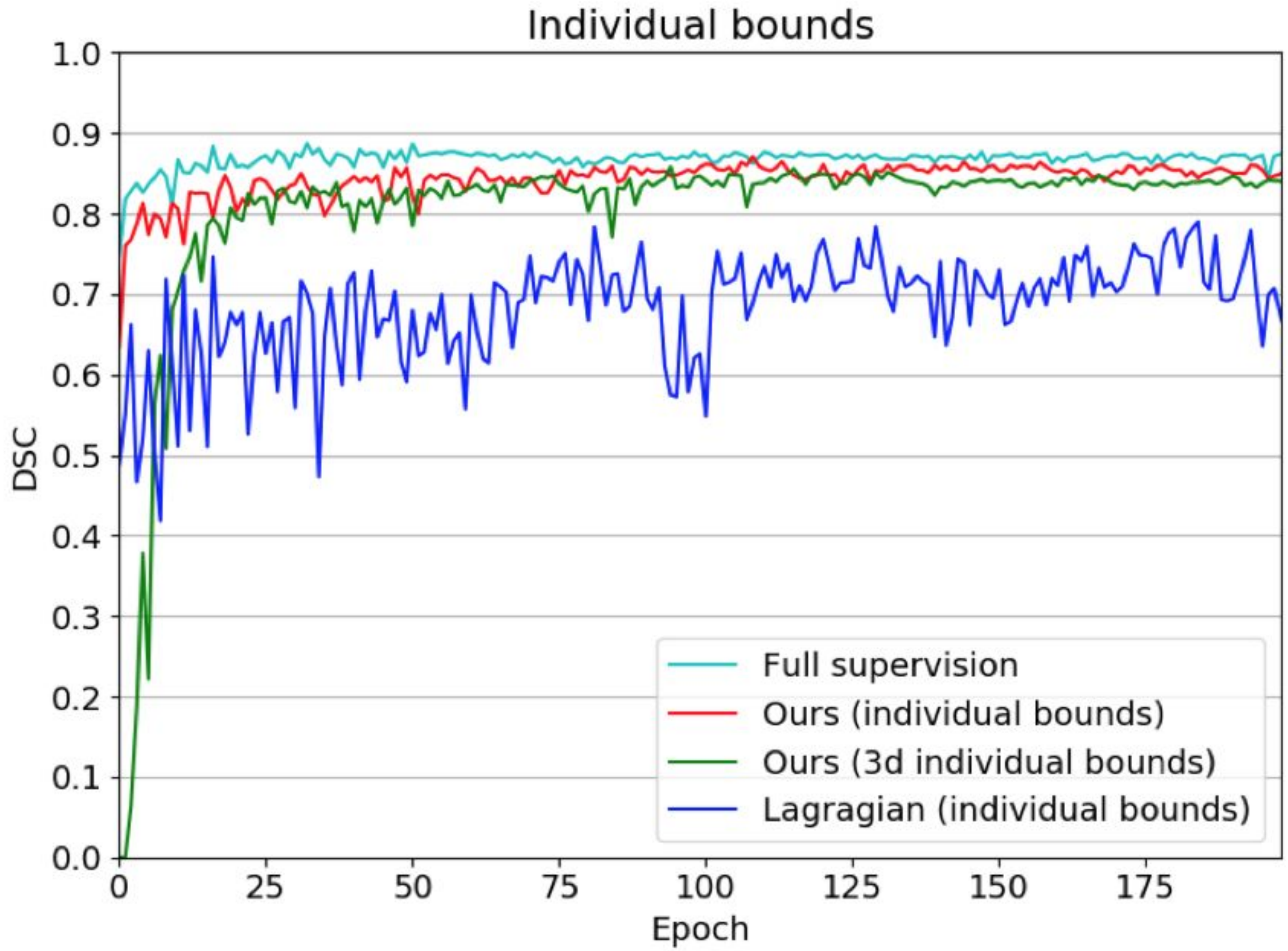
Mostly penalties in deep networks:

Weakly supervised learning
Adversarial robustness
Stabilizing GANs
Reinforcement learning

[Boyd and Vandenberghe, Convex optimization, 2004]

Constrained optimization (in CNNs)

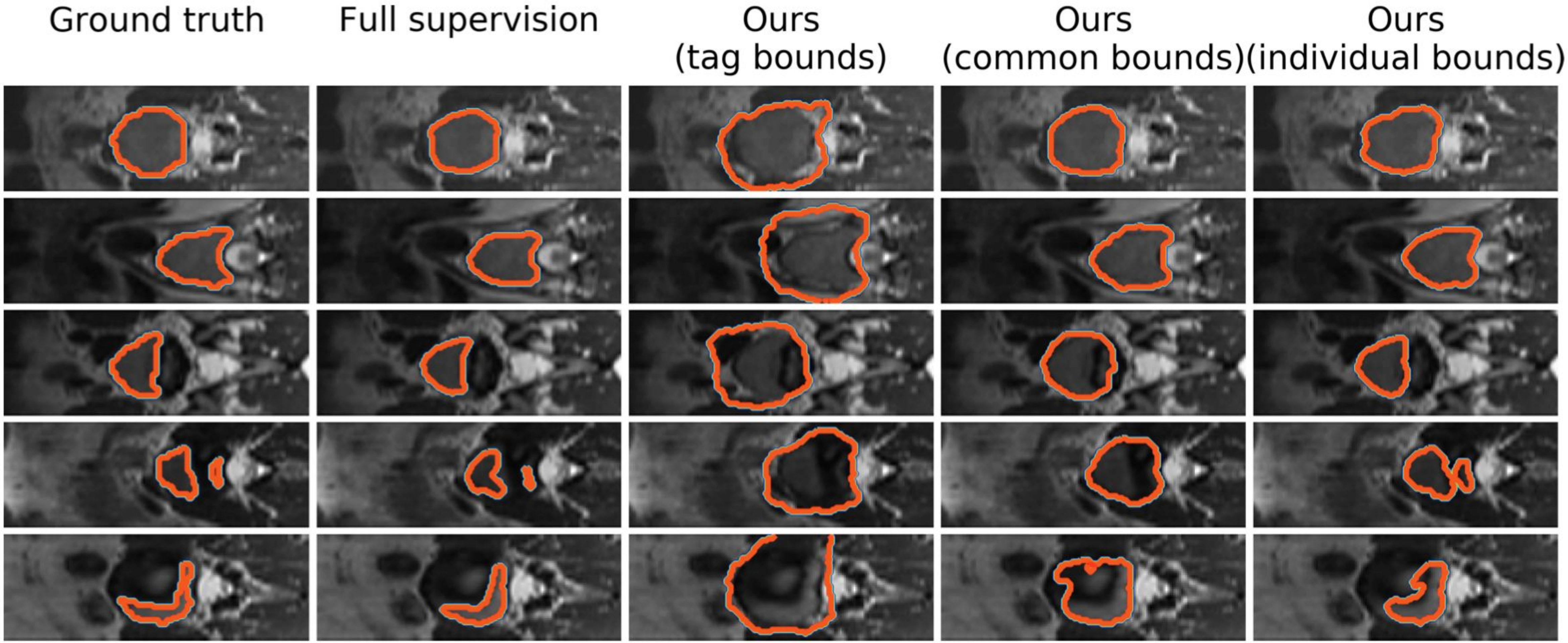
Inequality constraints (e.g, L2 penalty)



[Kervadec et al., Constrained-CNN losses for weakly supervised segmentation. MedIA 2019]

Constrained optimization (in CNNs)

Inequality constraints (e.g, L2 penalty)



[Kervadec et al., Constrained-CNN losses for weakly supervised segmentation. MedIA 2019]

Constrained optimization (in CNNs)

Inequality constraints (e.g, L2 penalty)

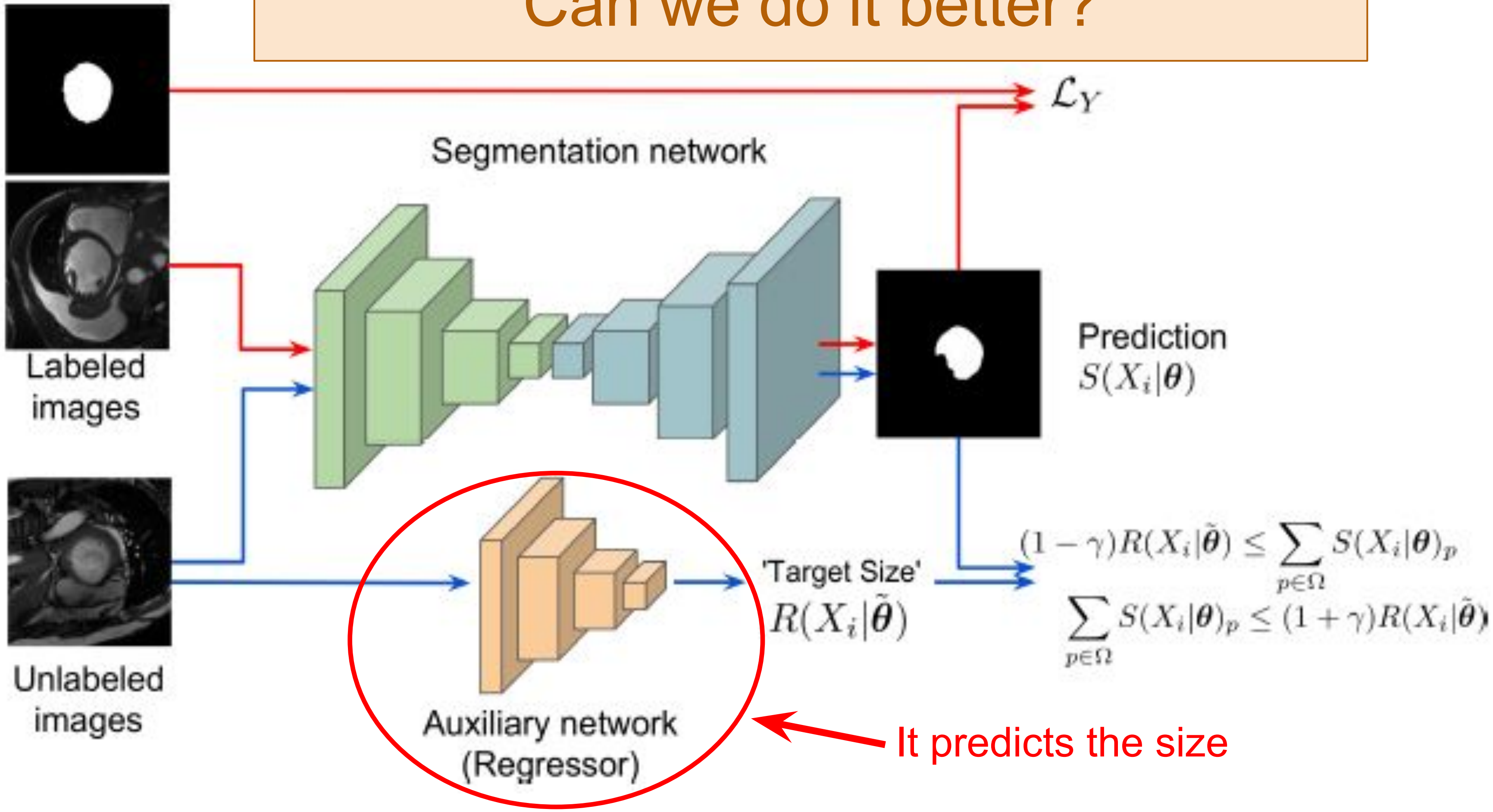
Can we do it better?

[Kervadec et al., Curriculum semi-supervised segmentation. MICCAI'19]

Constrained optimization (in CNNs)

Inequality constraints (e.g, L2 penalty)

Can we do it better?



It predicts the size

[Kervadec et al., Curriculum semi-supervised segmentation. MICCAI'19]

Constrained optimization (in CNNs)

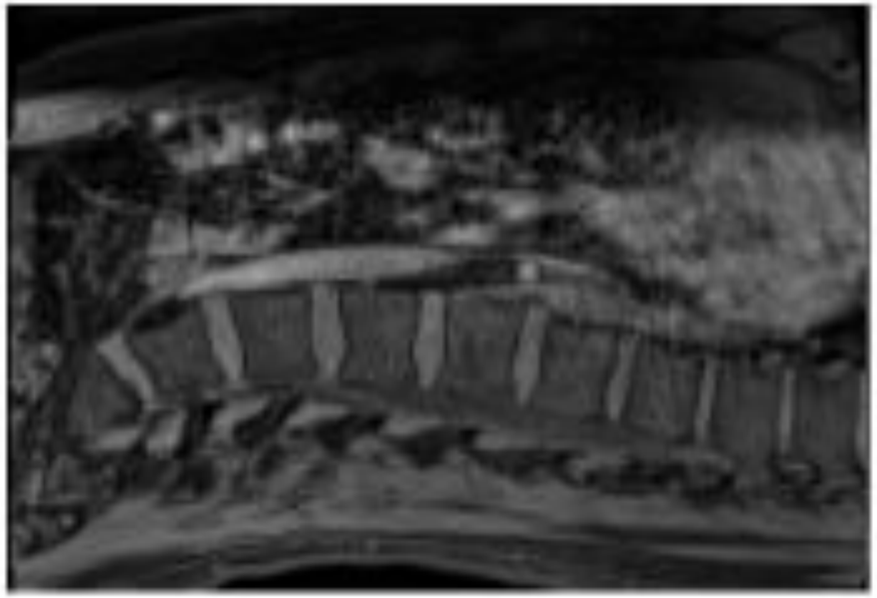
Inequality constraints (e.g, L2 penalty)

Also useful in Domain adaptation!

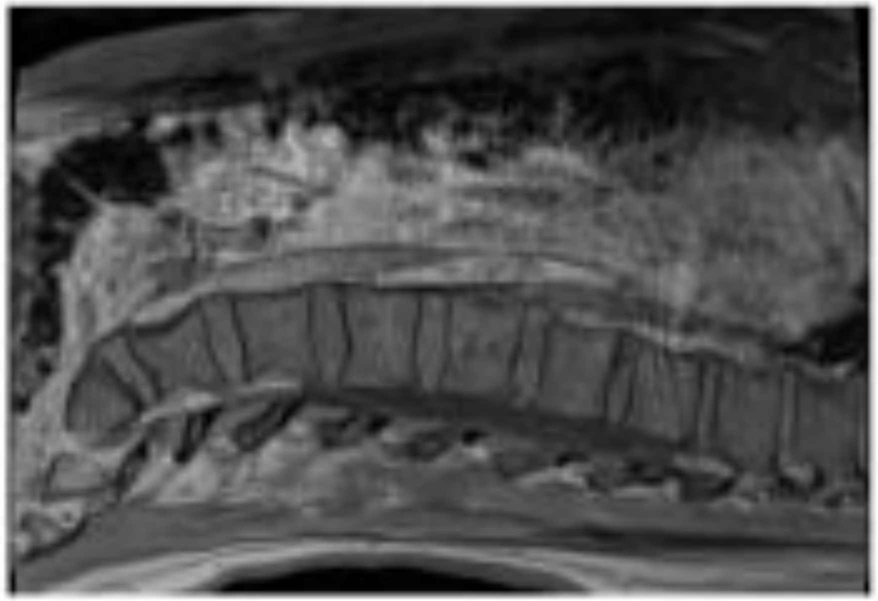
Constrained optimization (in CNNs)

Inequality constraints (e.g, L2 penalty)

Source (labeled) modality (WAT)



Target (unlabeled) modality (IP)

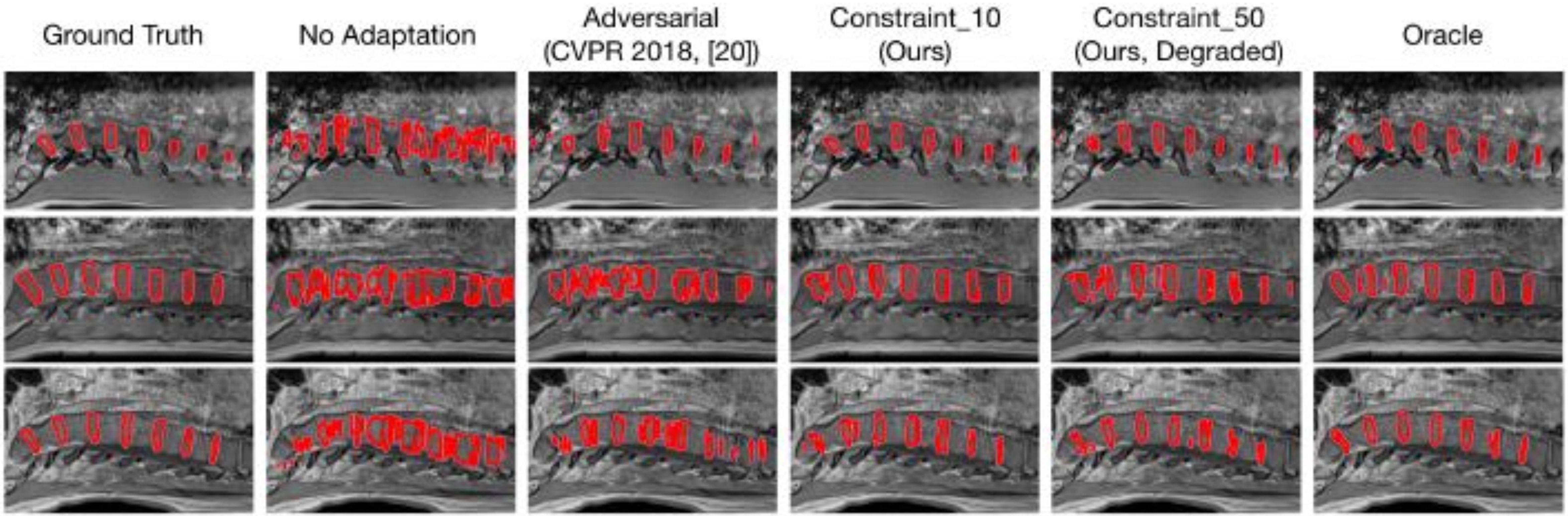


Dataset from IVD 2018 MICCAI Challenge

[Bateson et al., Constrained domain adaptation for segmentation. MICCAI'19]

Constrained optimization (in CNNs)

Inequality constraints (e.g, L2 penalty)

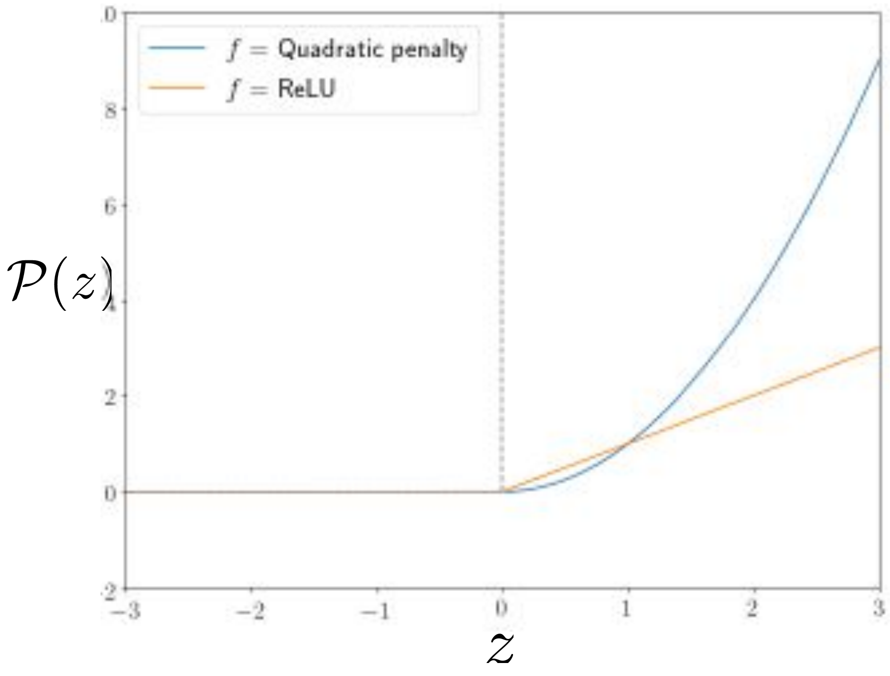


[Bateson et al., Constrained domain adaptation for segmentation. MICCAI’19]

Limitations of penalties

$$\min_{\theta} \mathcal{J}(\theta)$$

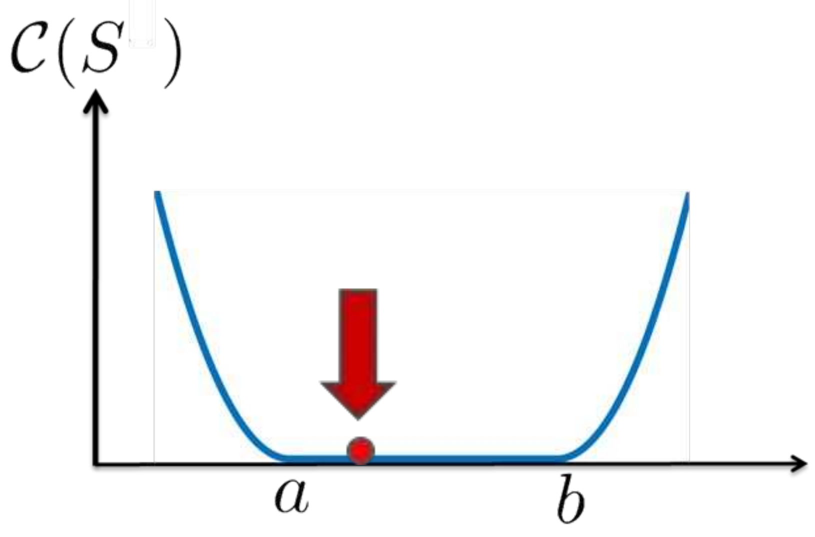
$$\mathcal{J}(\theta) = \mathcal{E}(\theta) + \lambda \mathcal{P}(f(S_\theta))$$



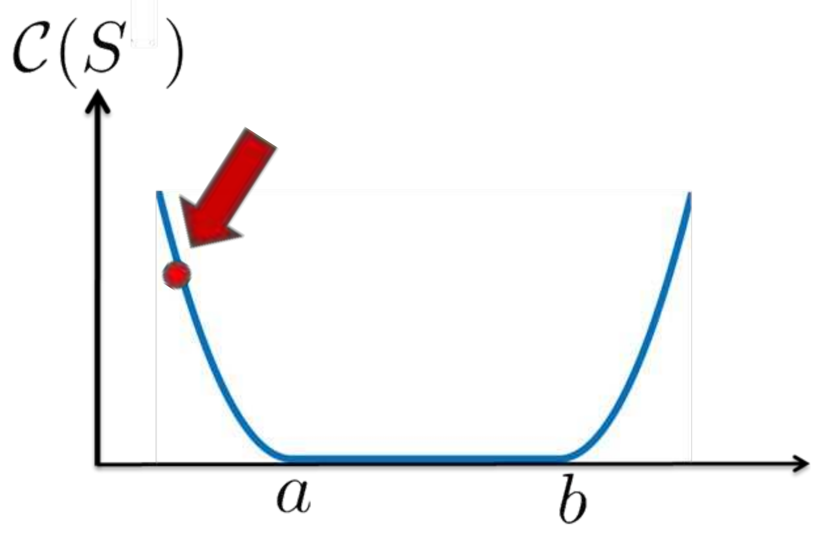
Problem: How to set the weight?
Multiples constraints make it worse?

Limitations of penalties

Multiple constraints



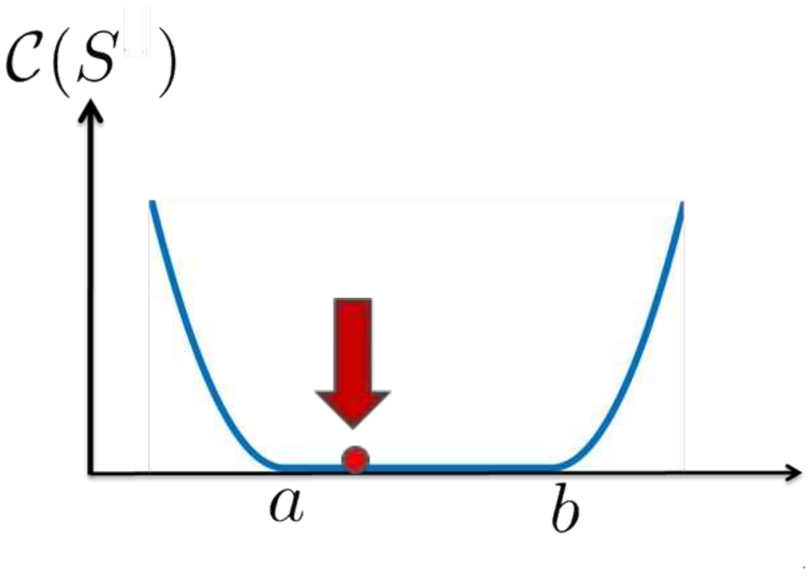
Constraint A satisfied

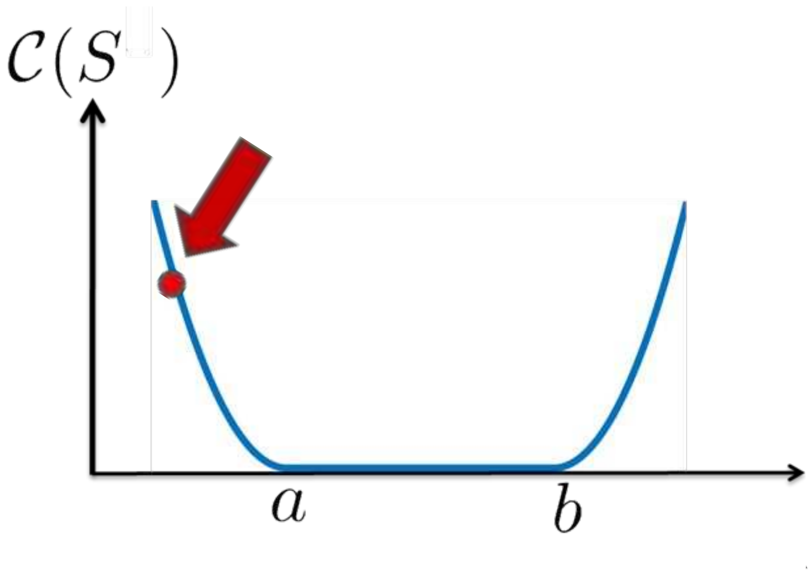


Constraint B violated

Limitations of penalties

Multiple constraints





Constraint A satisfied

Constraint B violated

Gradient is 0

Gradient is **NOT** 0

Log-barrier extensions: Approximates Lagrangian optimization but **NO** explicit dual steps

Lagrangian optimization can deal with these limitations:

- it finds automatically the optimal weights of the constraints.
- it acts as a barrier for satisfied constraints.
- it guarantees constraint satisfaction when feasible solutions exist.

[Kervadec et al., Log-barrier constrained CNNs, ArXiv 2019]

Log-barrier extensions: Approximates Lagrangian optimization but **NO** explicit dual steps

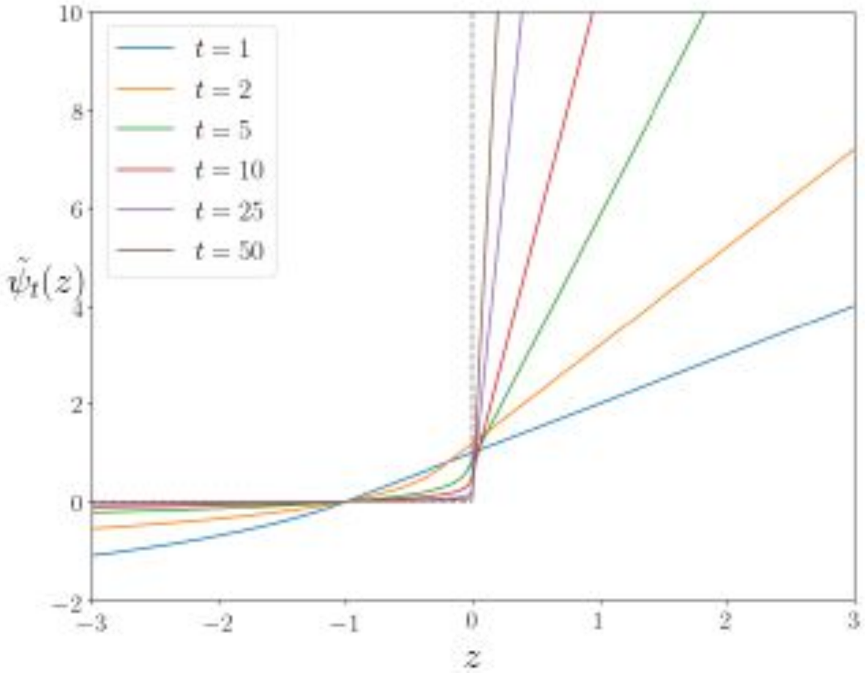
Lagrangian optimization can deal with these limitations:

- it finds automatically the optimal weights of the constraints.
- it acts as a barrier for satisfied constraints.
- it guarantees constraint satisfaction when feasible solutions exist.

$$\min_{\boldsymbol{\theta}} \mathcal{E}(\boldsymbol{\theta}) + \sum_{i=1}^{N} \tilde{\psi}_t\left(f_i(S_{\boldsymbol{\theta}})\right) \quad \Rightarrow \quad \tilde{\psi}_t(z) = \begin{cases} -\frac{1}{t}\log(-z) & \text{if } z \leq -\frac{1}{t^2} \\ tz - \frac{1}{t}\log(\frac{1}{t^2}) + \frac{1}{t} & \text{otherwise} \end{cases}$$

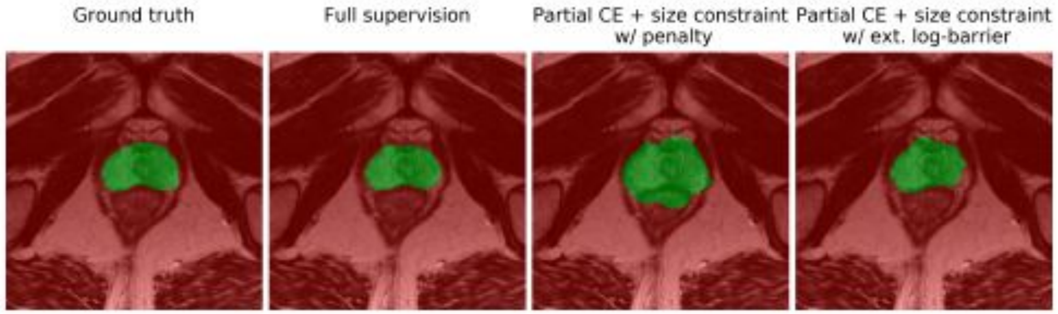
[Kervadec et al., Constrained Deep Networks: Lagrangian Optimization via Log-Barrier Extensions, ArXiv 2019]

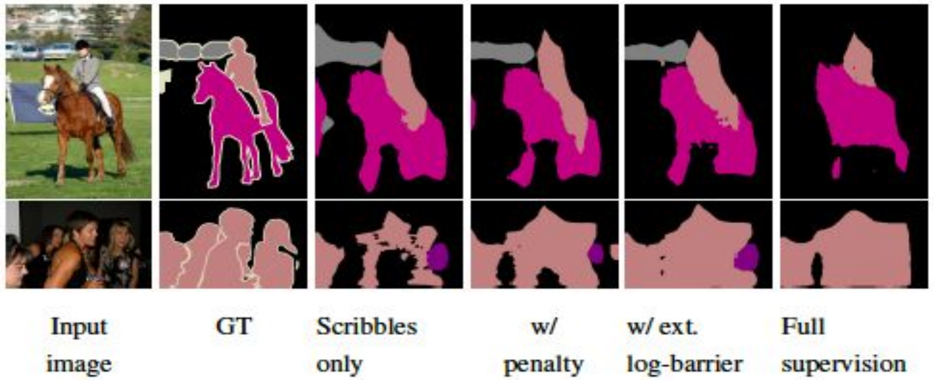
Log-barrier extensions: Approximates Lagrangian optimization but **NO** explicit dual steps



Promising results

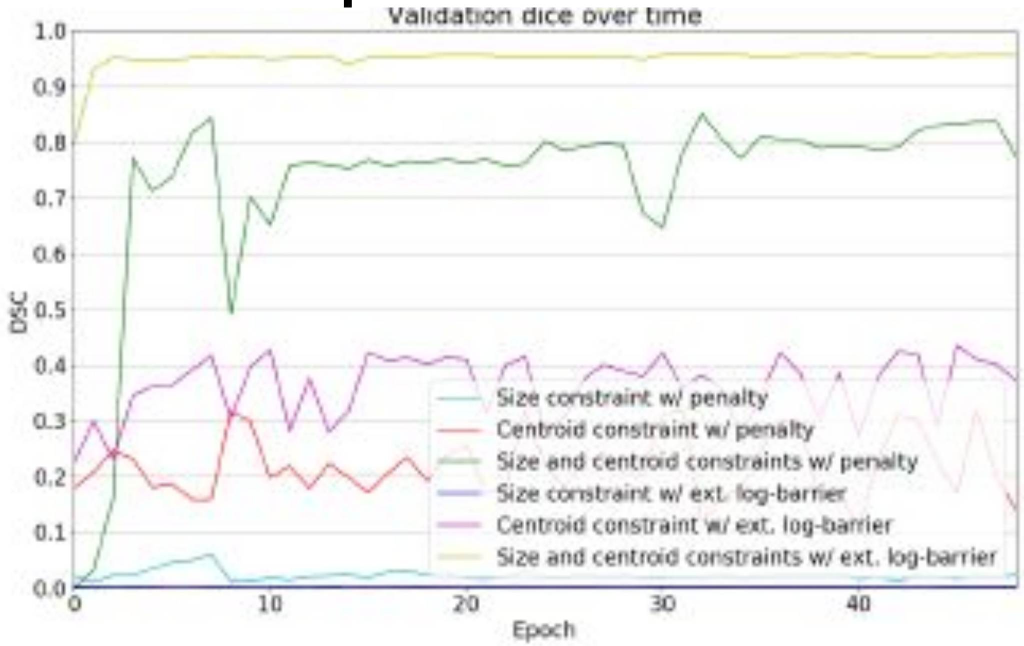
	Dataset	
Method	PROMISE12 (DSC)	VOC2012 (mIoU)
Partial cross-entropy	0.032 (0.015)	48.48 (14.88)
w/ penalty [12]	0.830 (0.057)	52.22 (14.94)
w/ extended log-barrier	**0.852** (0.038)	**53.40** (14.62)
Full supervision	0.891 (0.032)	59.87 (16.94)





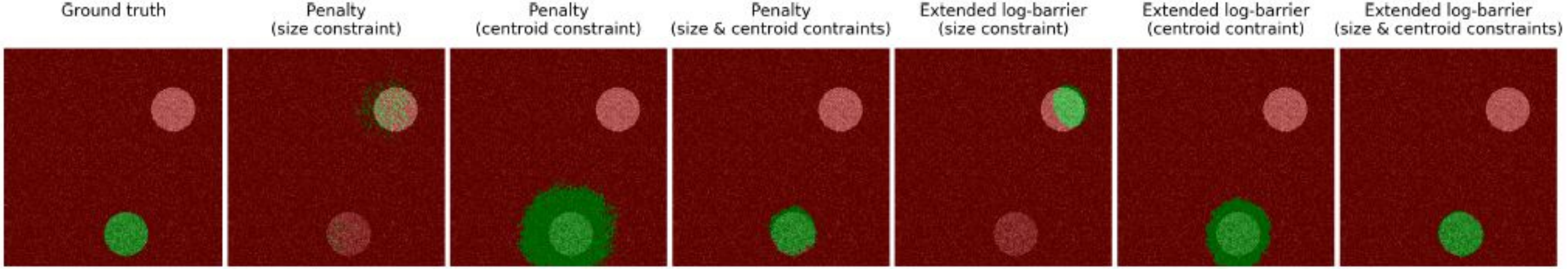
[Kervadec et al., Constrained Deep Networks: Lagrangian Optimization via Log-Barrier Extensions, ArXiv 2019]

Log-barrier extensions: Approximates Lagrangian optimization but **NO** explicit dual steps



Promising results

	Constraints		
Method	Size	Centroid	Size & Centroid
Penalty [12]	0.0601	0.3197	**0.8514**
Extended log-barrier	0.0018	0.4347	**0.9574**

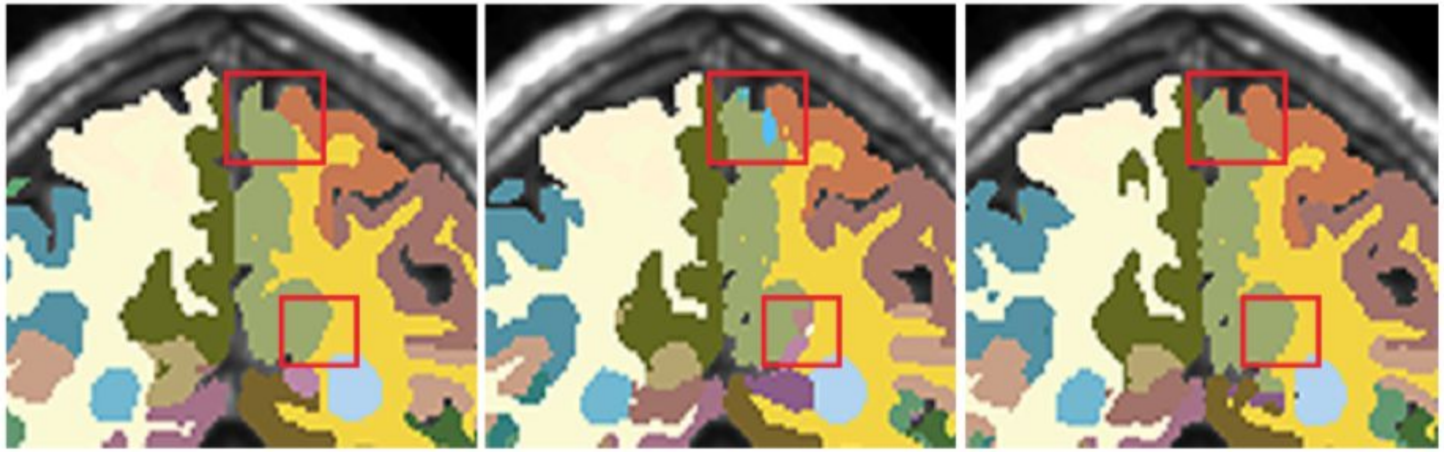


[Kervadec et al., Constrained Deep Networks: Lagrangian Optimization via Log-Barrier Extensions, ArXiv 2019]

Other constraints (Semi-supervision)

Connectivity of anatomical structures

Ground Truth w/o prior Prior



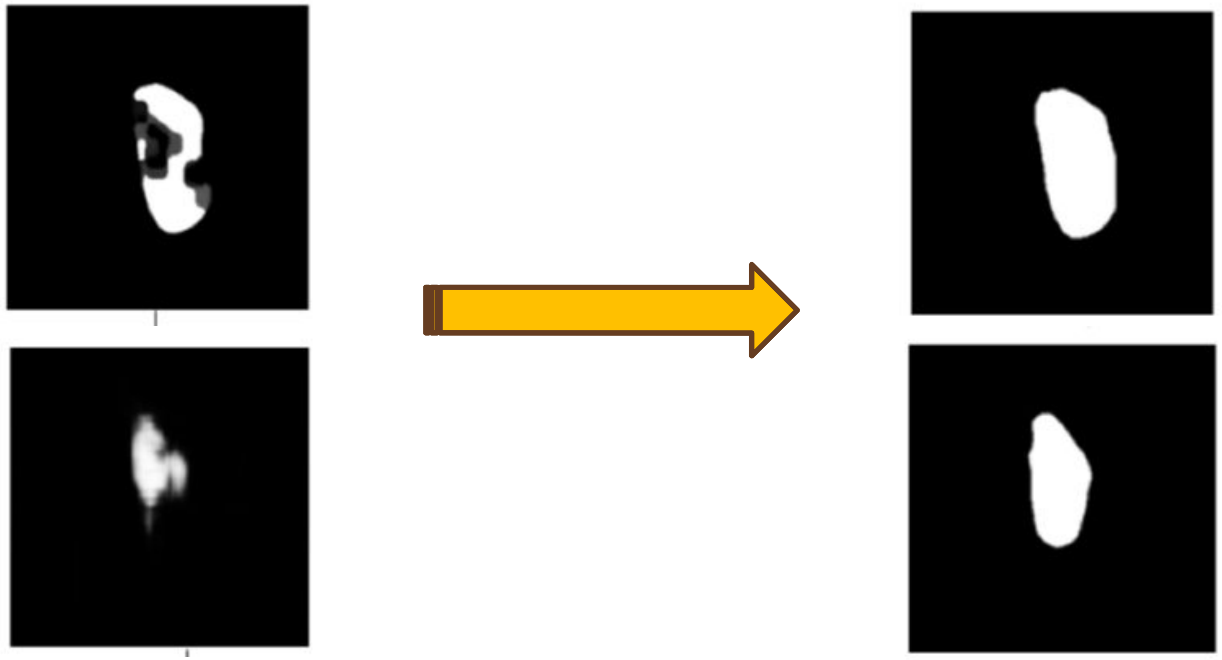
Images from [Ganaye et al. Semi-supervised learning for segmentation under semantic constraint. MICCAI'18]

Other constraints (Full-supervision)

Shape

Bad segmentations

Good segmentations

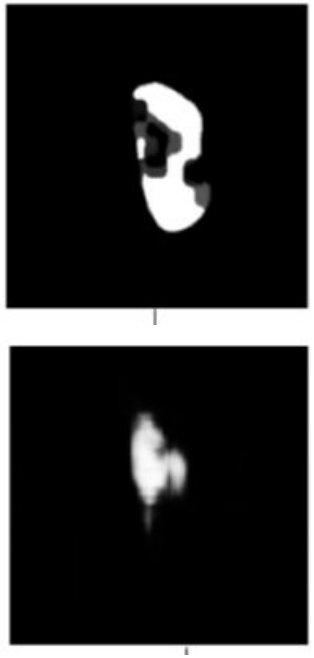


Ravishankar H, et al. Learning and incorporating shape models for semantic segmentation. MICCAI’17

Other constraints (Full-supervision)

Shape

Bad segmentations



Good segmentations

During training

Oktay et al., IEEE TMI’17

Post-processing

Larrazabal et al., MICCAI’19
Painchaud et al, MICCAI’19

Ravishankar H, et al. Learning and incorporating shape models for semantic segmentation. MICCAI’17

Take-home message

- Imposing constraints helps weakly-supervised segmentation learning by restricting plausible segmentations on unlabeled images

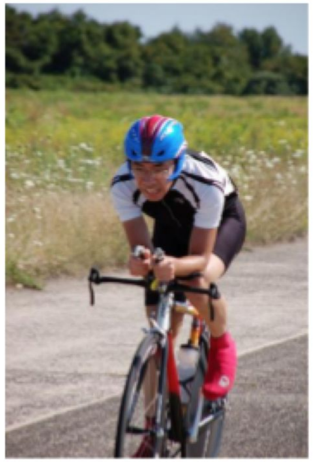
- Few constraints have been explored under low-labeled data regime

- Room for improvement (many opportunities)

Data-driven priors
Introduction

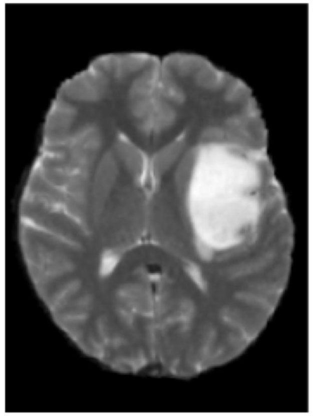
Data-driven priors (cues)

Image tags



Person

Bike



Tumor

Original Image

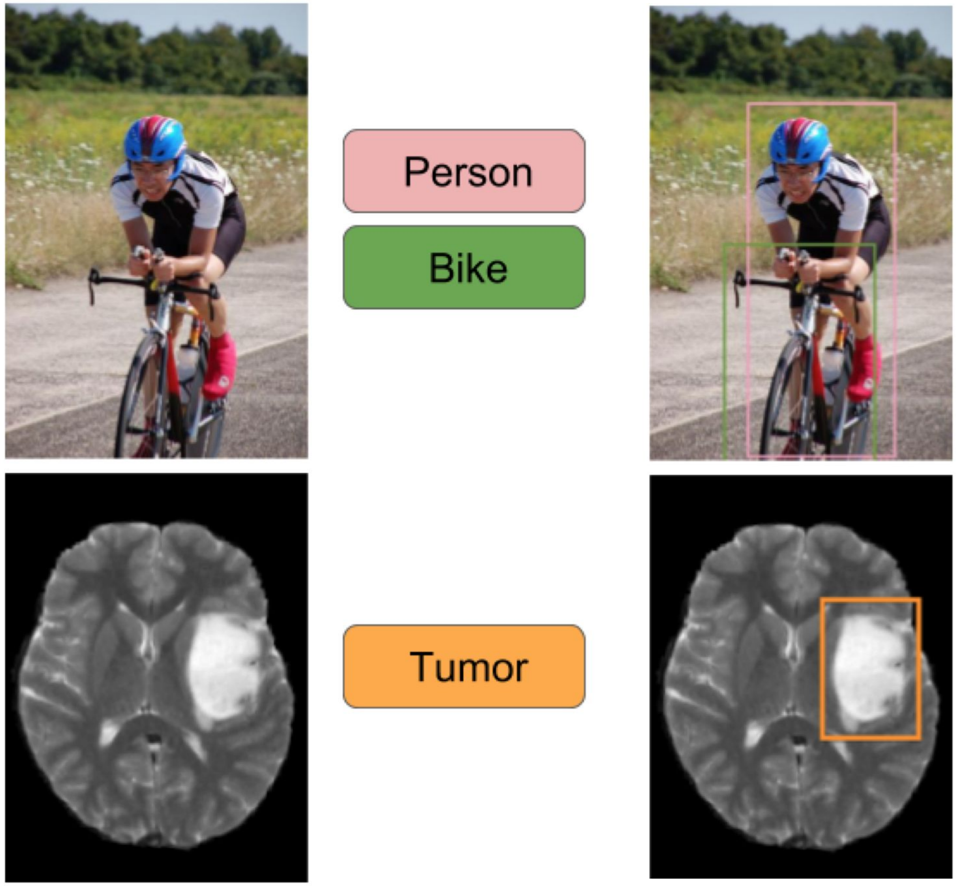
Image tags

- Pathak et al., Constrained convolutional neural networks for weakly supervised segmentation, ICCV 2015
- Kervadec et al., Constrained-CNN losses for weakly supervised segmentation, MedIA 2019.

Data-driven priors (cues)

Image tags

Bounding boxes



Original Image | Image tags | Bounding boxes

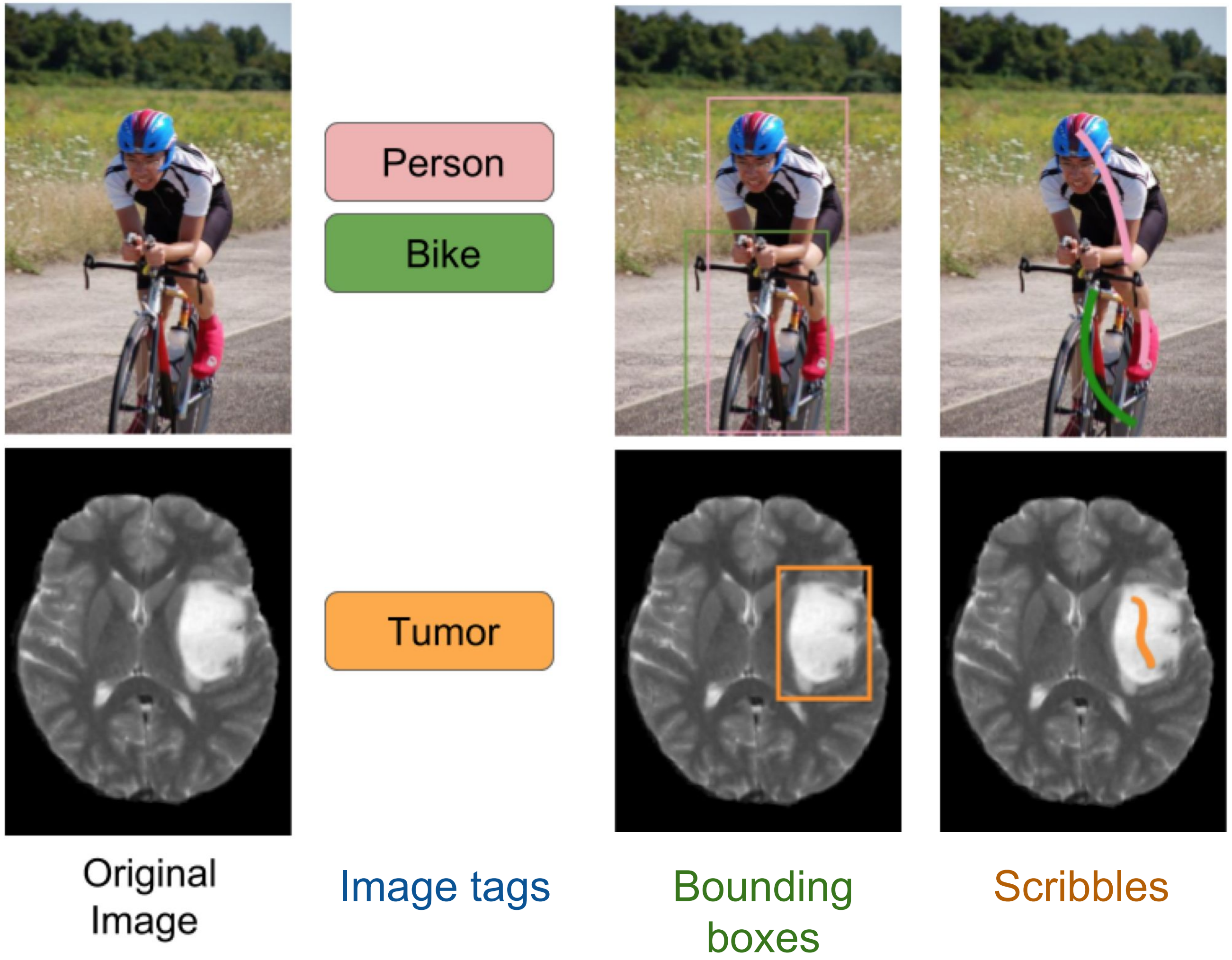
- Pathak et al., Constrained convolutional neural networks for weakly supervised segmentation, ICCV 2015
- Kervadec et al., Constrained-CNN losses for weakly supervised segmentation, MedIA 2019.

Data-driven priors (cues)

Image tags

Bounding boxes

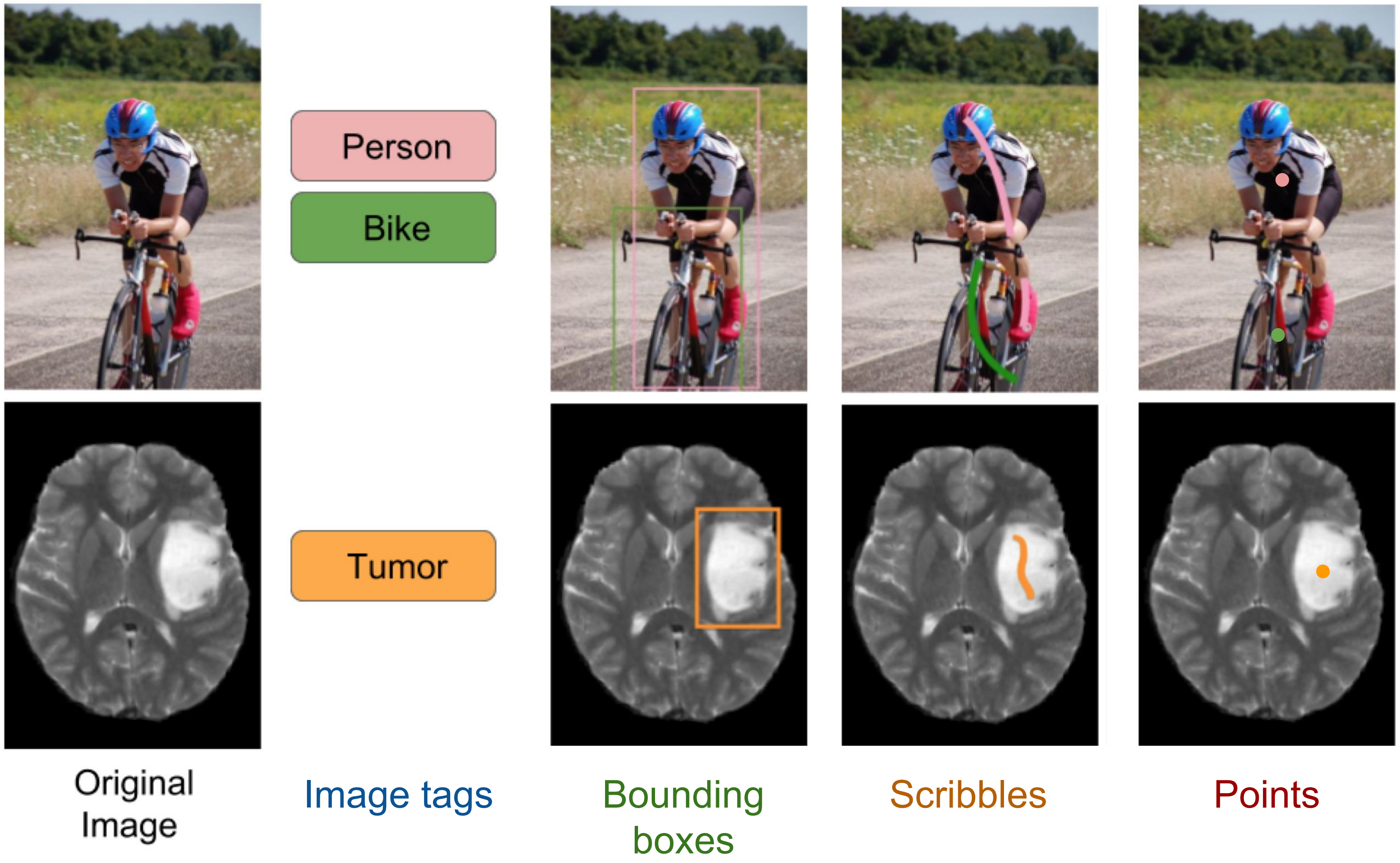
Scribbles



- Pathak et al., Constrained convolutional neural networks for weakly supervised segmentation, ICCV 2015
- Kervadec et al., Constrained-CNN losses for weakly supervised segmentation, MedIA 2019.

Data-driven priors (cues)

- Image tags
- Bounding boxes
- Scribbles
- Points



- Pathak et al., Constrained convolutional neural networks for weakly supervised segmentation, ICCV 2015
- Kervadec et al., Constrained-CNN losses for weakly supervised segmentation, MedIA 2019.

Data-driven priors (cues)

Another data-driven priors

Image captions

A boy jumping on a skateboard

Extreme points

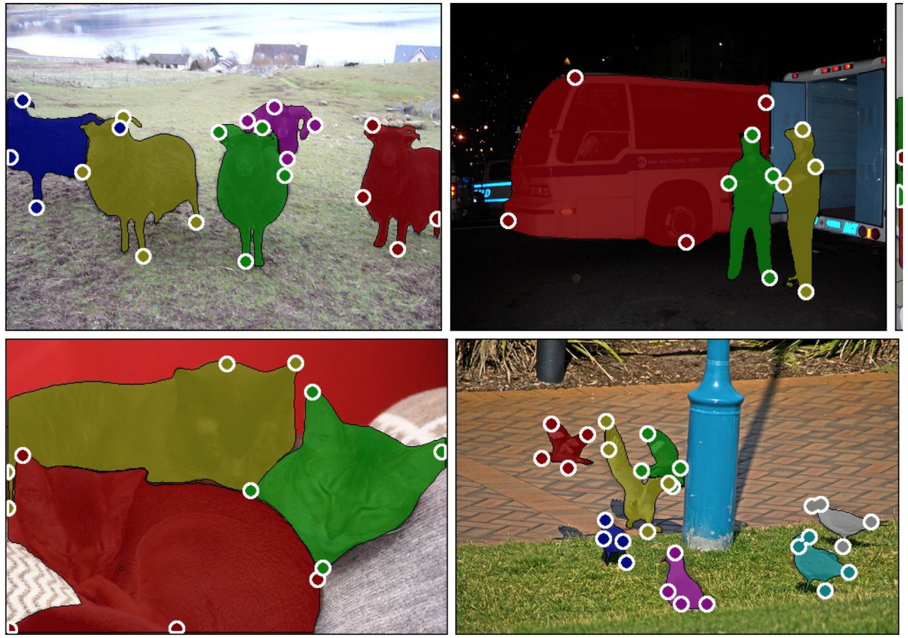
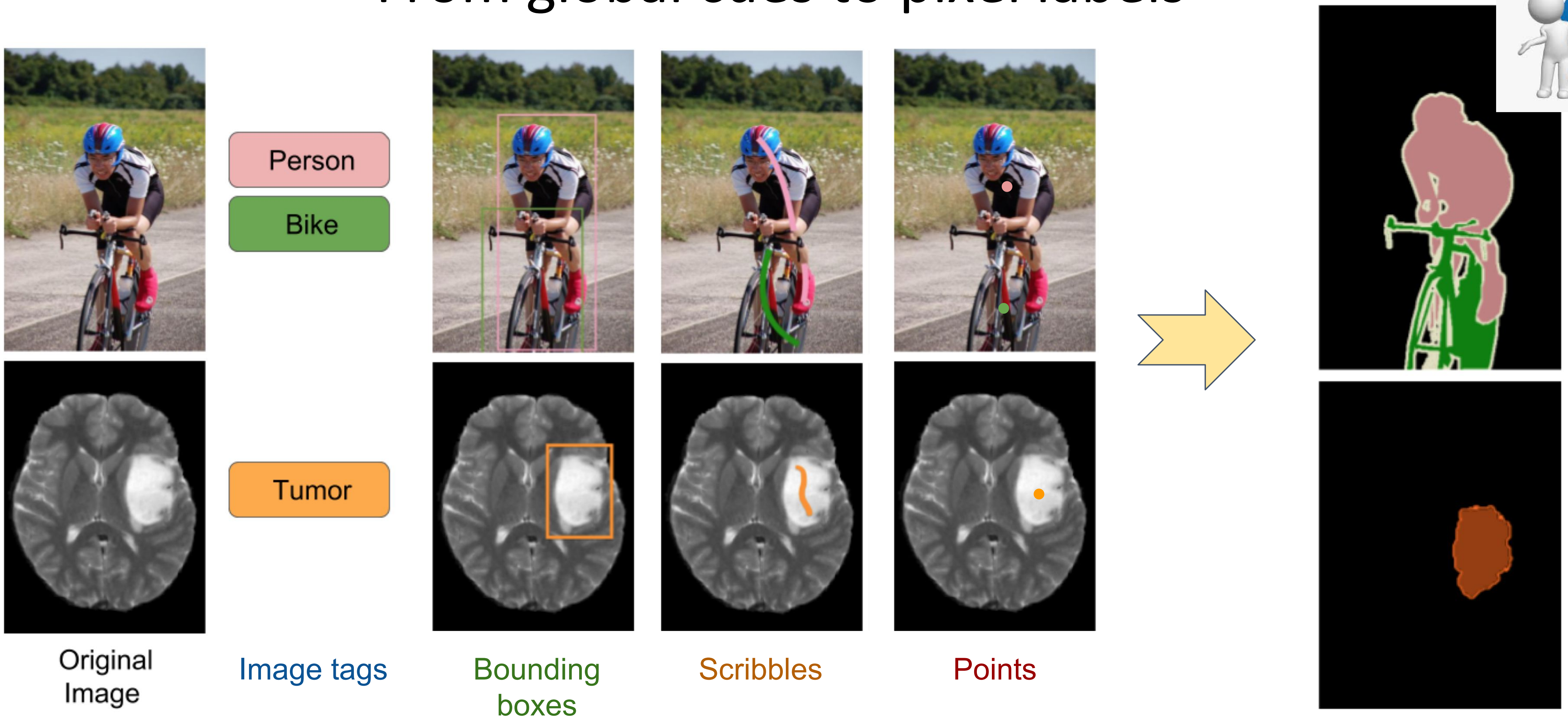


Image from Maninis et al, CVPR'18

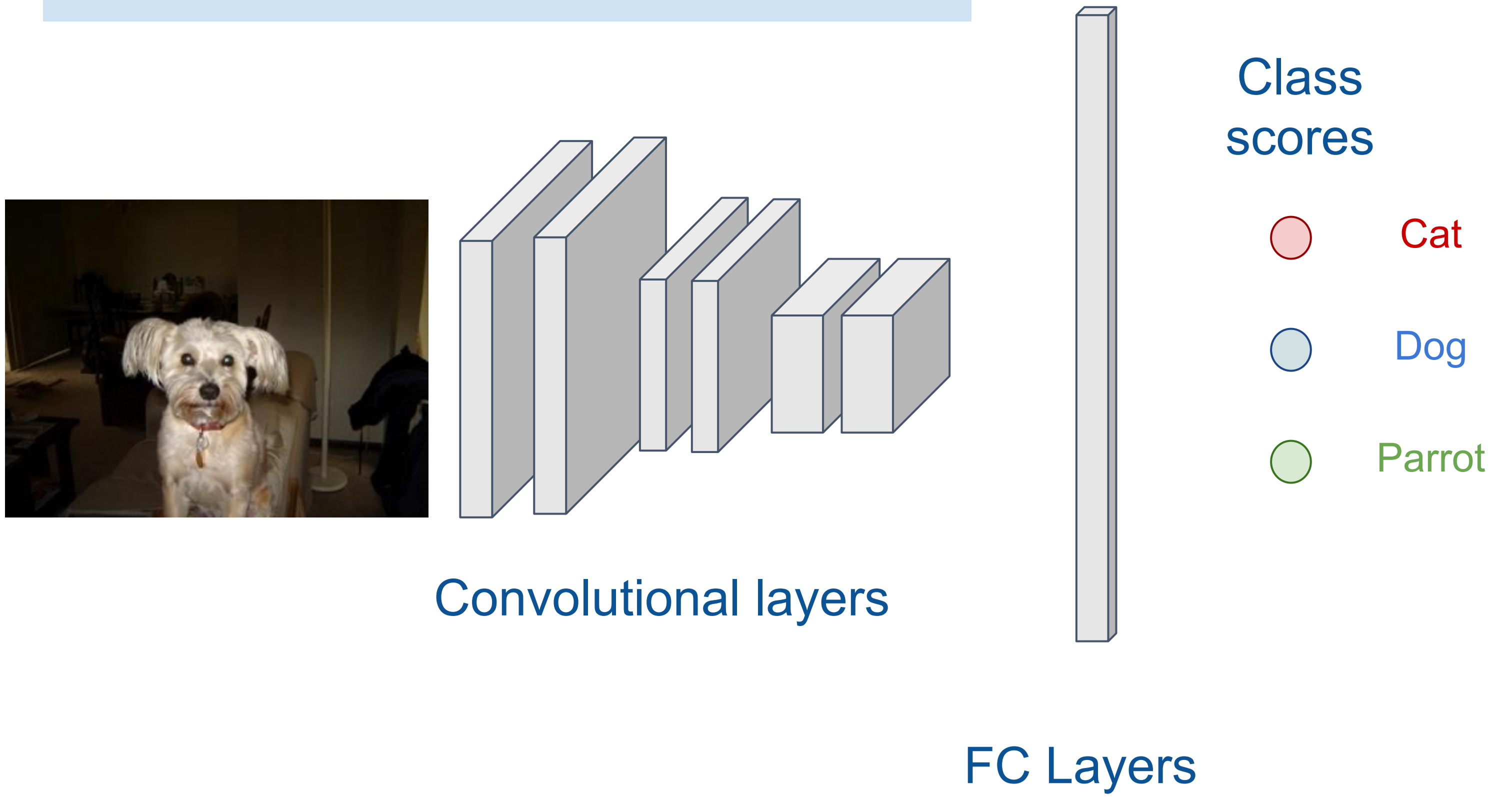
- Maninis et al. Deep extreme cut: From extreme points to object segmentation. CVPR 2018

From global cues to pixel labels



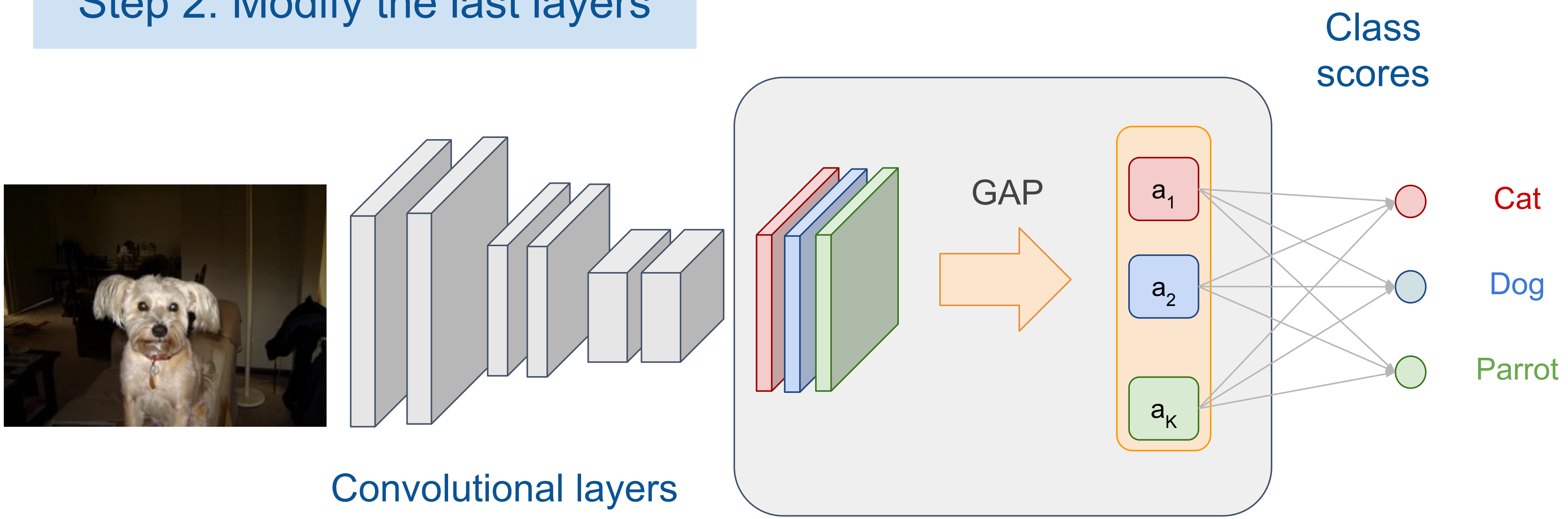
From global cues to pixel labels

Step 1: Get a classification CNN



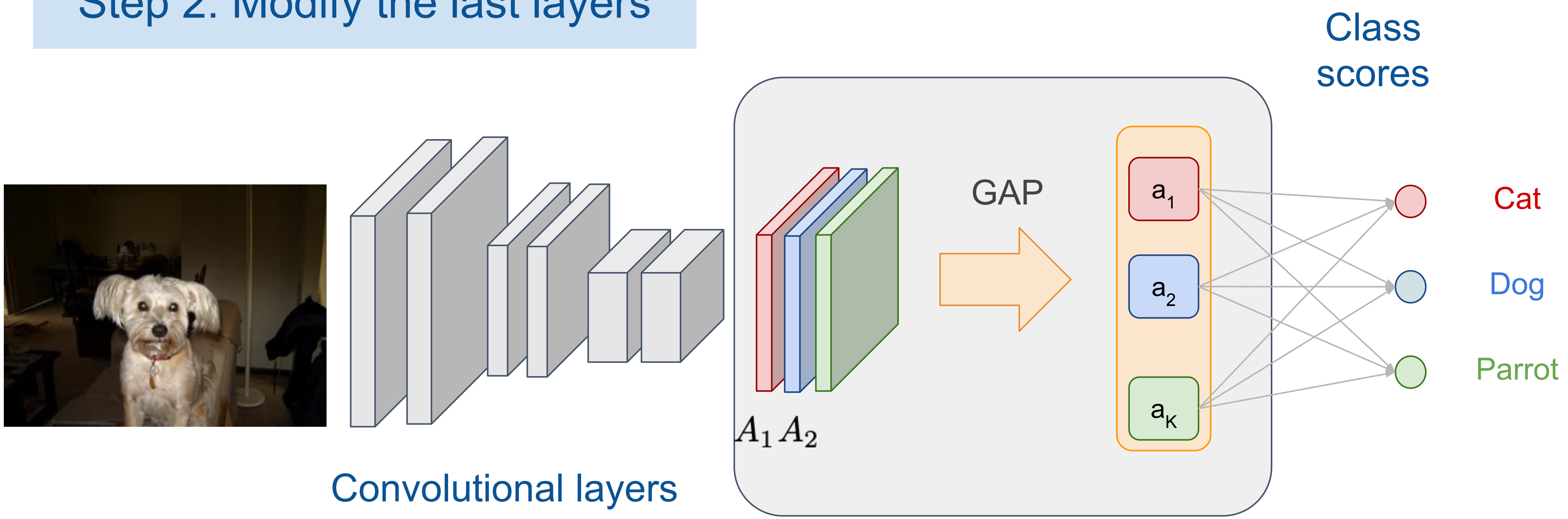
From global cues to pixel labels

Step 2: Modify the last layers



From global cues to pixel labels

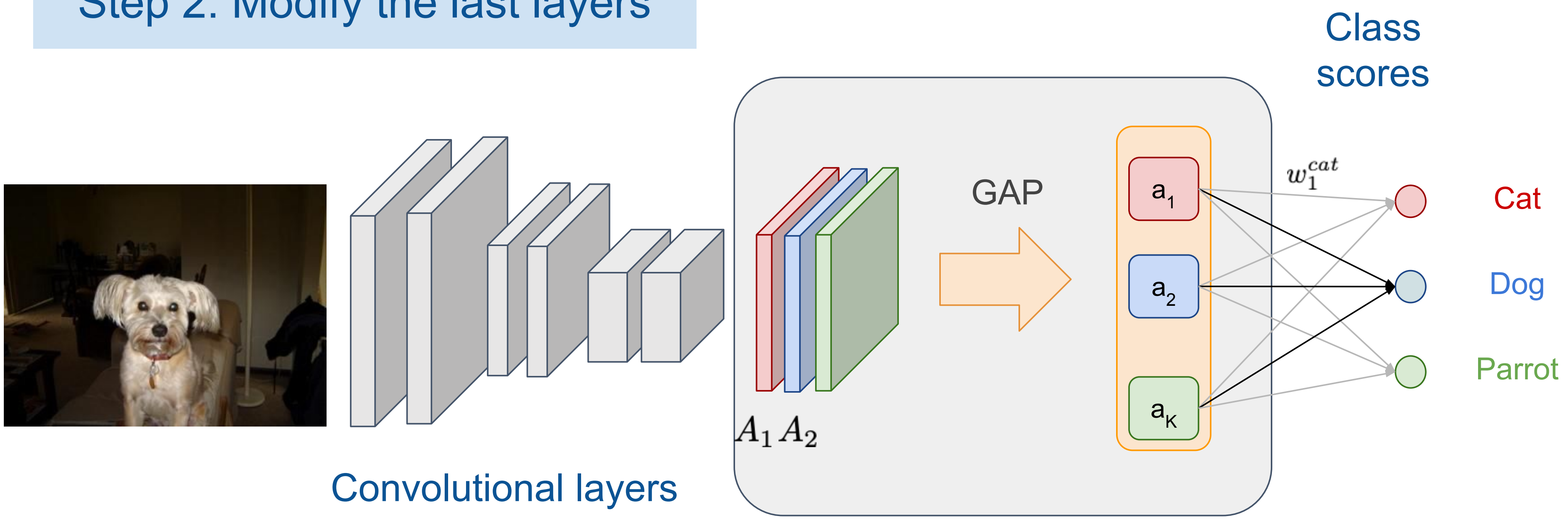
Step 2: Modify the last layers



$$GAP(A_k) = a_k = \frac{1}{|N|}\sum_{x,y} A_k(x,y)$$

From global cues to pixel labels

Step 2: Modify the last layers



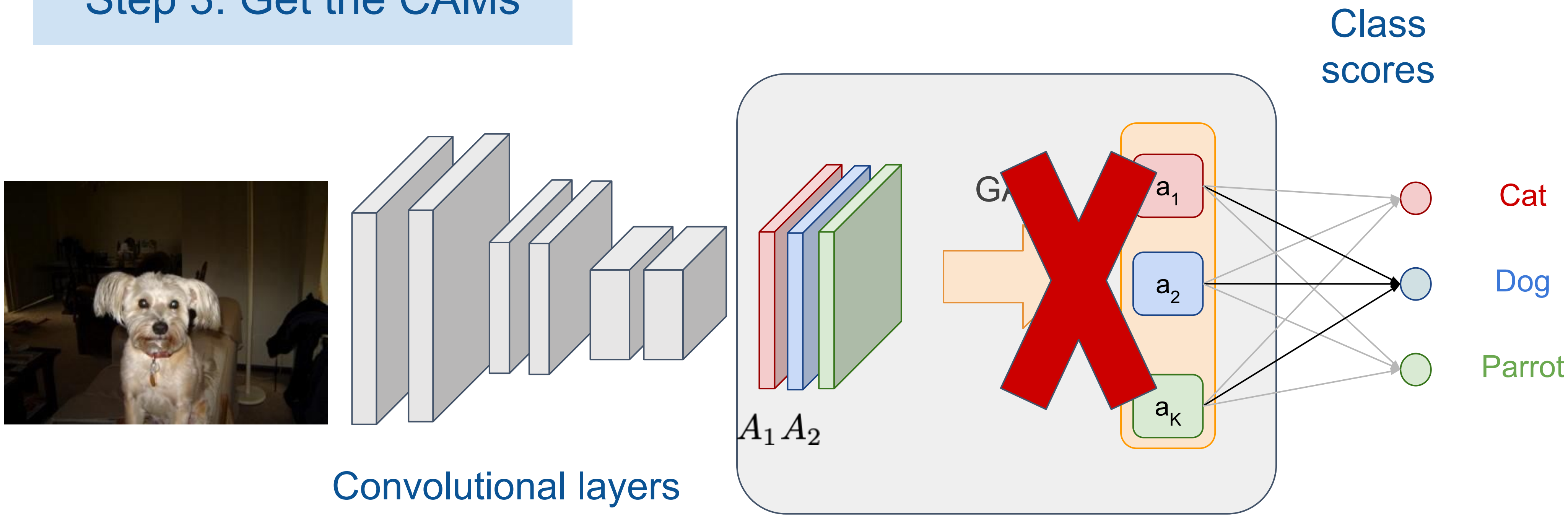
$$GAP(A_k) = a_k = \frac{1}{|N|}\sum_{x,y} A_k(x,y)$$

Class score (logits)

$$S_c = \sum_k w_k^c a_k = \frac{1}{N}\sum_k w_k^c \sum_{x,y} a_k(x,y)$$

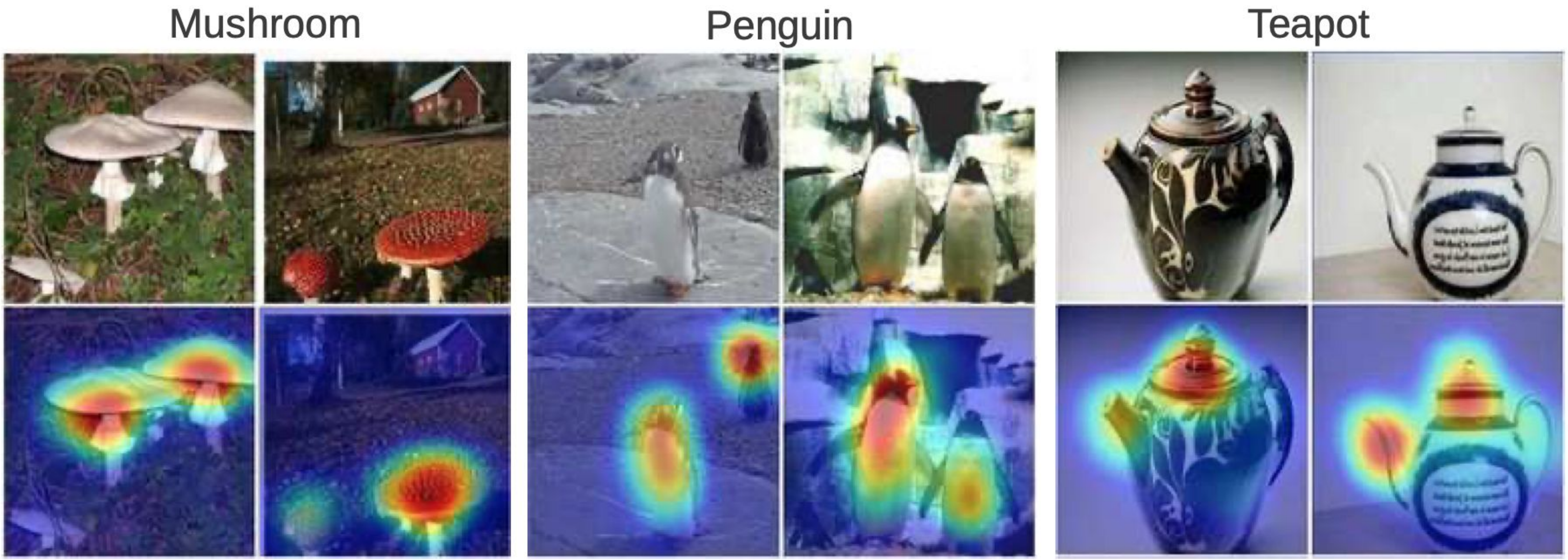
From global cues to pixel labels

Step 3: Get the CAMs



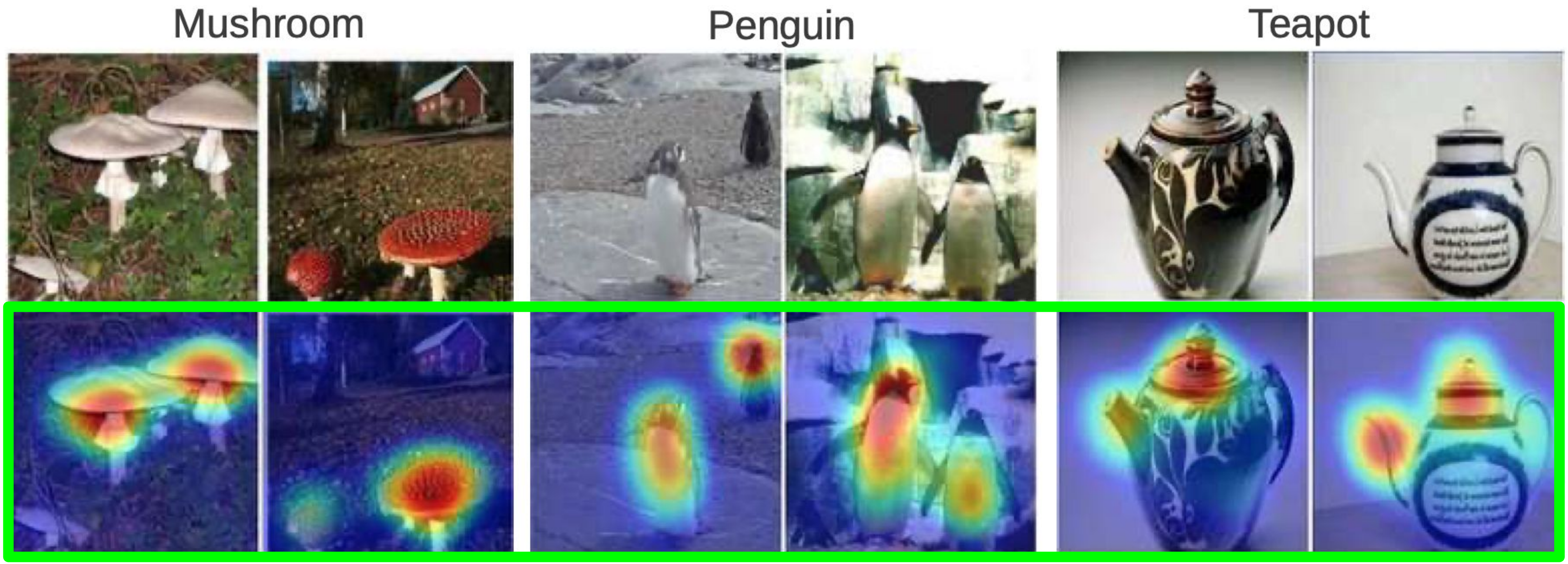
$$CAM_{Dog}(x,y) = \sum_k w_k^{Dog} A_k(x,y) =$$

From global cues to pixel labels



- Zhou et al., Learning deep features for discriminative localization. CVPR 2016

From global cues to pixel labels



These activations maps can be used as **pseudo-masks**

- Zhou et al., Learning deep features for discriminative localization. CVPR 2016

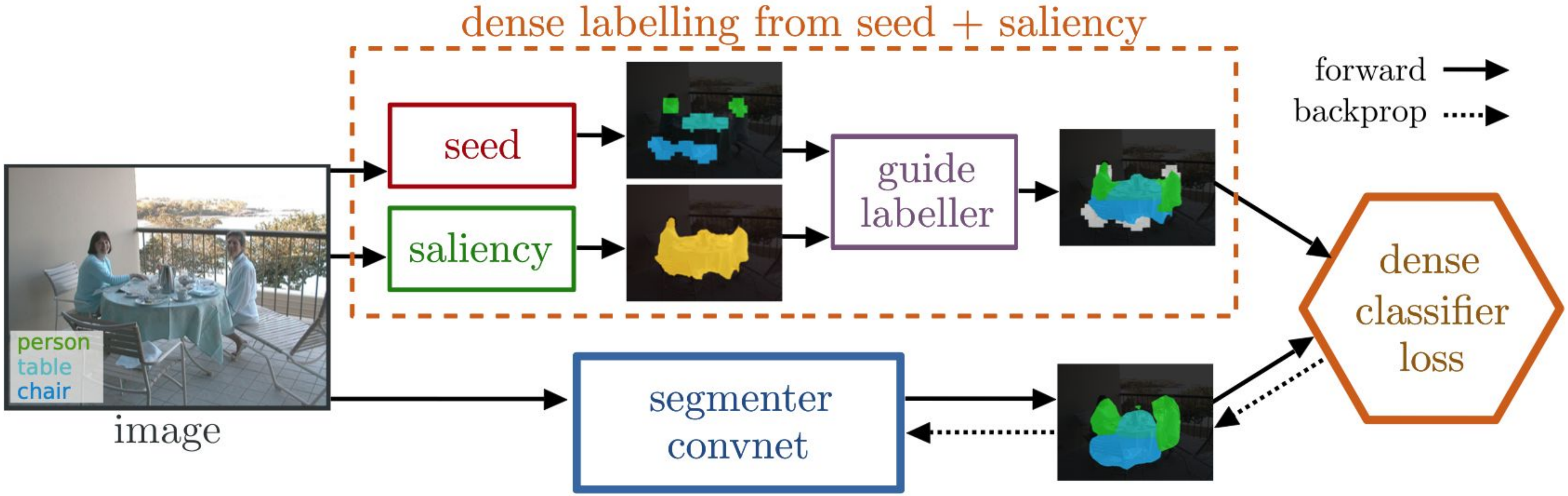
From global cues to pixel labels

Problem: they focus only on highly discriminative regions

From global cues to pixel labels

Problem: they focus only on highly discriminative regions

Incorporate saliency maps

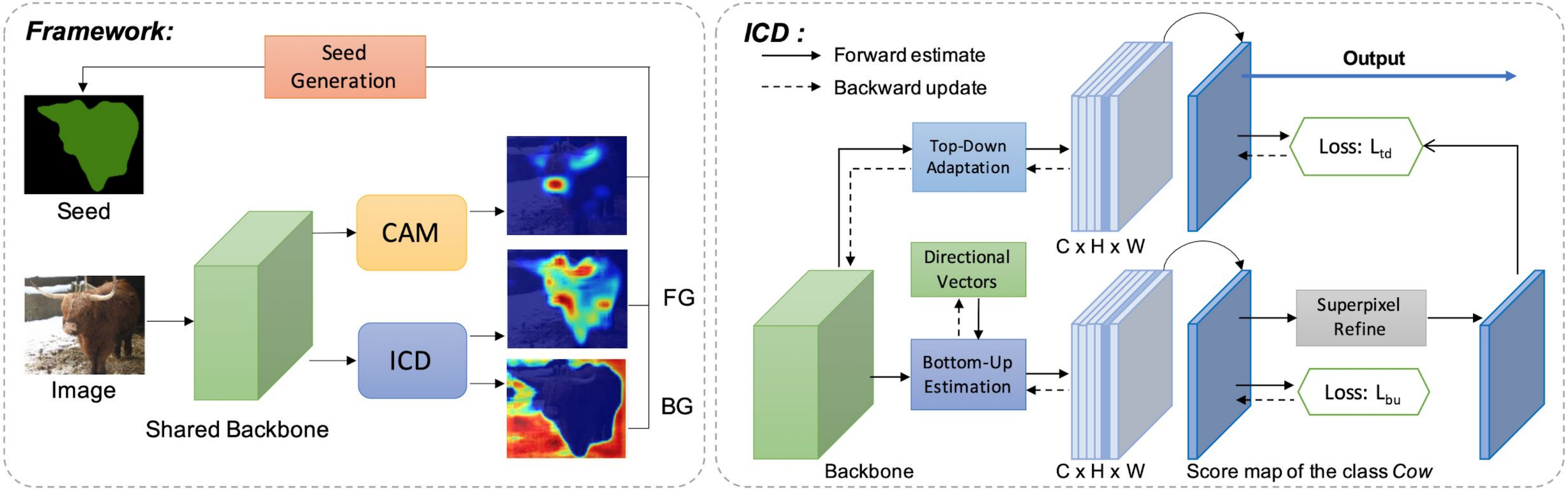


- Oh et al. Exploiting Saliency for Object Segmentation from Image Level Labels. CVPR 2017

From global cues to pixel labels

Problem: they focus only on highly discriminative regions

Incorporate saliency maps

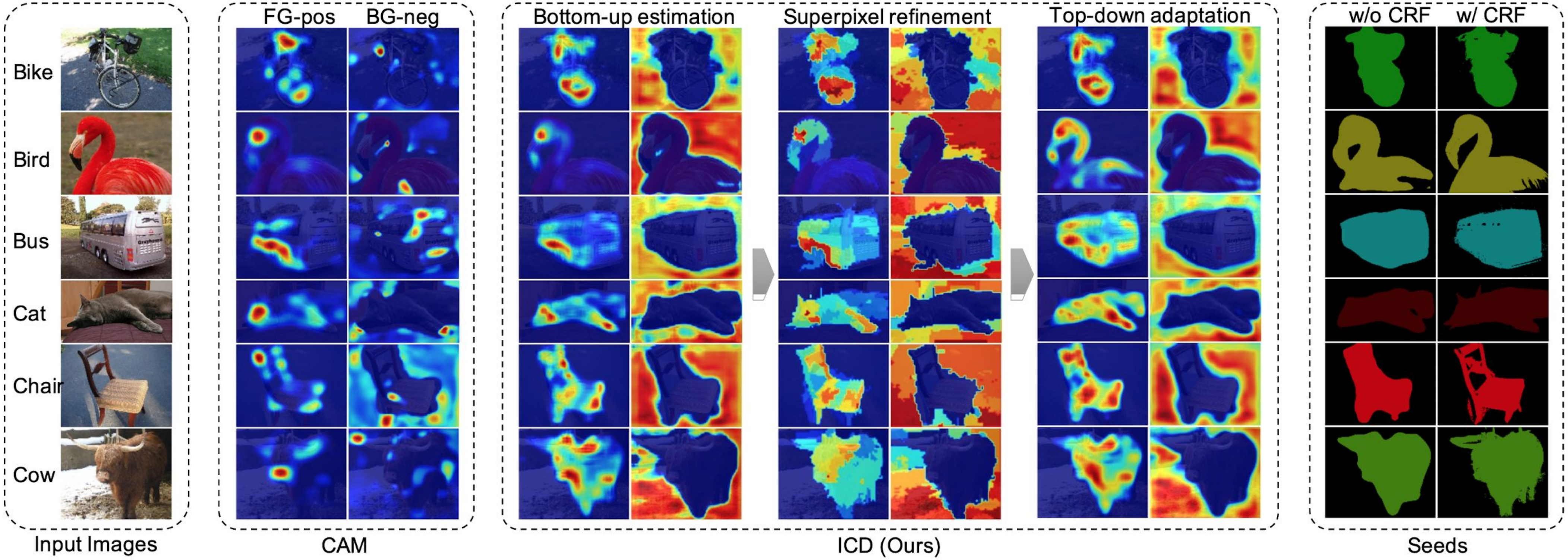


- Fan et al. Learning Integral Objects With Intra-Class Discriminator for Weakly-Supervised Semantic Segmentation. CVPR 2020

From global cues to pixel labels

Problem: they focus only on highly discriminative regions

Incorporate saliency maps

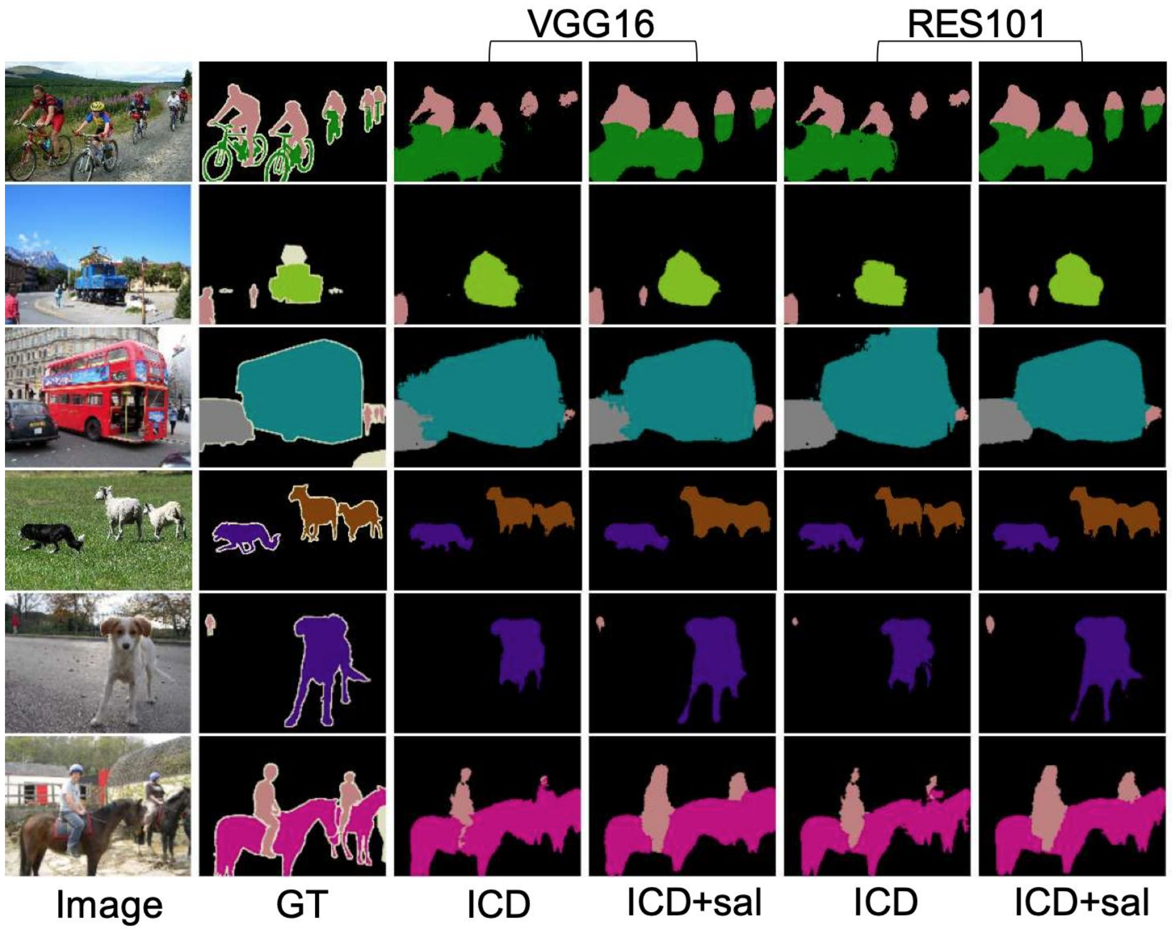


- Fan et al. Learning Integral Objects With Intra-Class Discriminator for Weakly-Supervised Semantic Segmentation. CVPR 2020

From global cues to pixel labels

Problem: they focus only on highly discriminative regions

Incorporate saliency maps

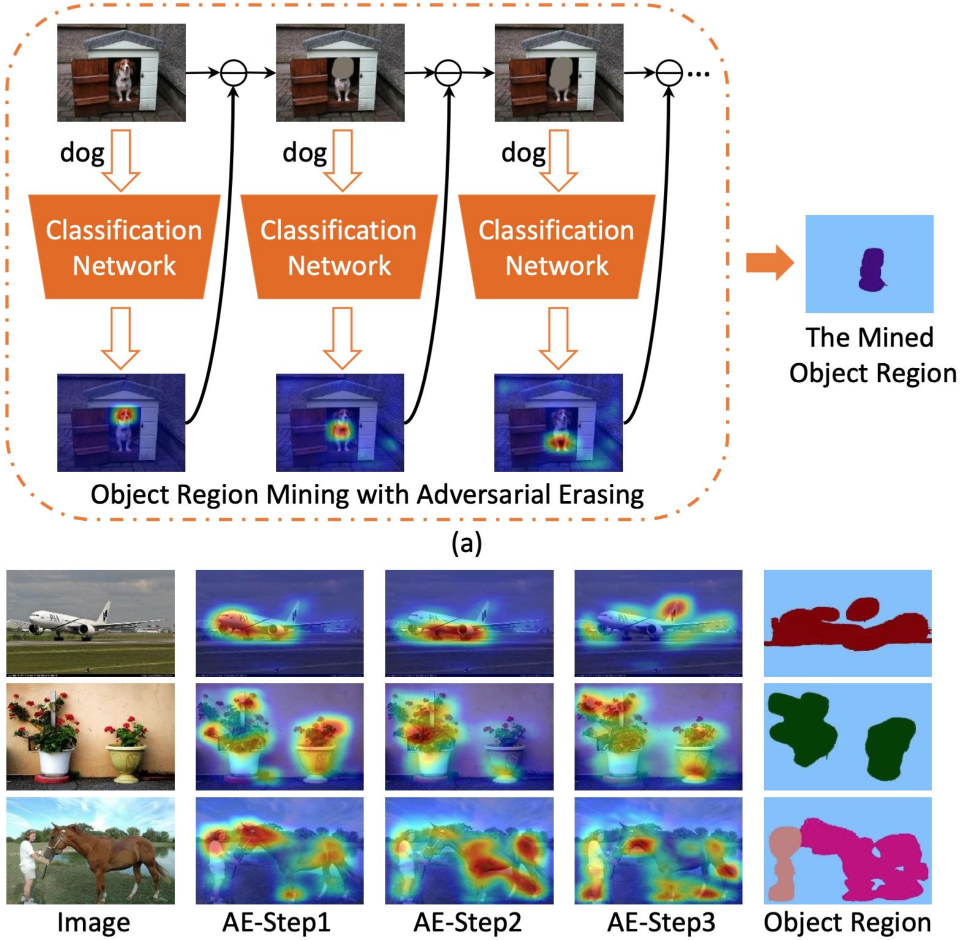


- Fan et al. Learning Integral Objects With Intra-Class Discriminator for Weakly-Supervised Semantic Segmentation. CVPR 2020

From global cues to pixel labels

Problem: they focus only on highly discriminative regions

Region mining

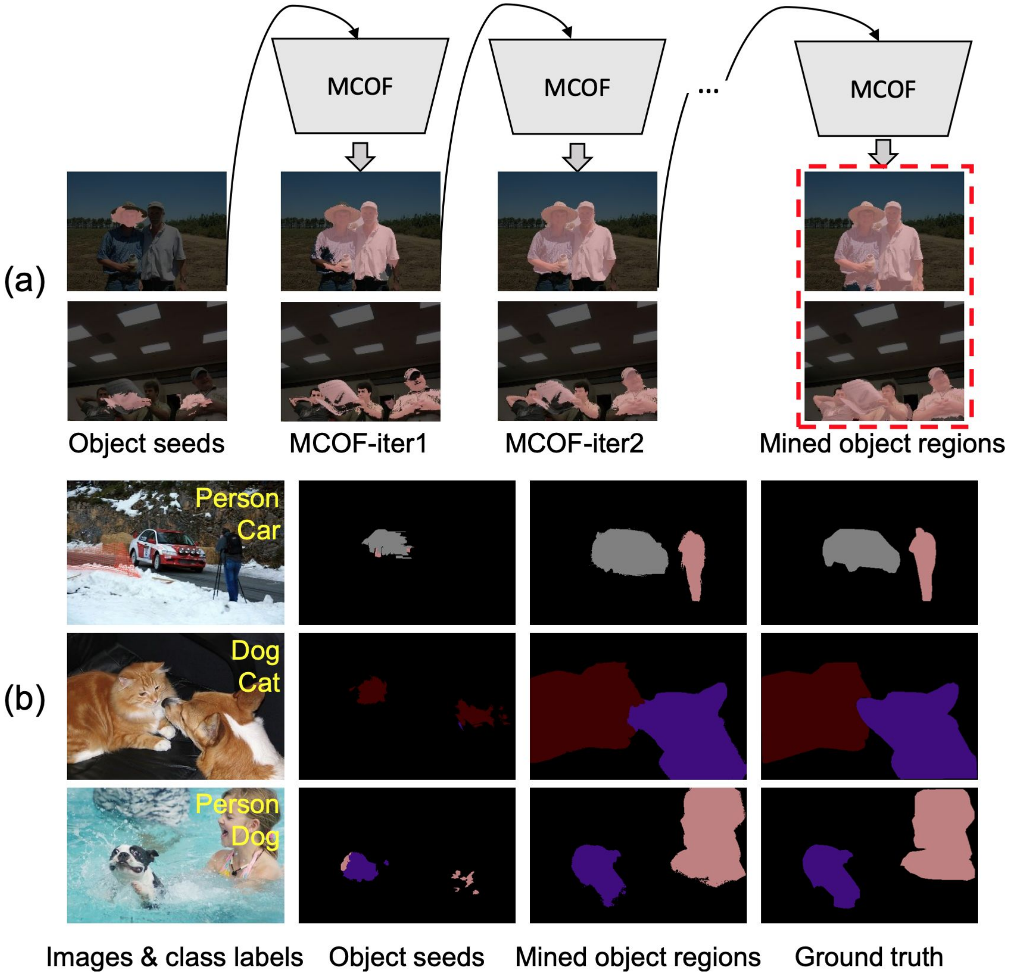


- Wei et al. Object Region Mining with Adversarial Erasing: A Simple Classification to Semantic Segmentation Approach. CVPR 2017

From global cues to pixel labels

Problem: they focus only on highly discriminative regions

Region mining

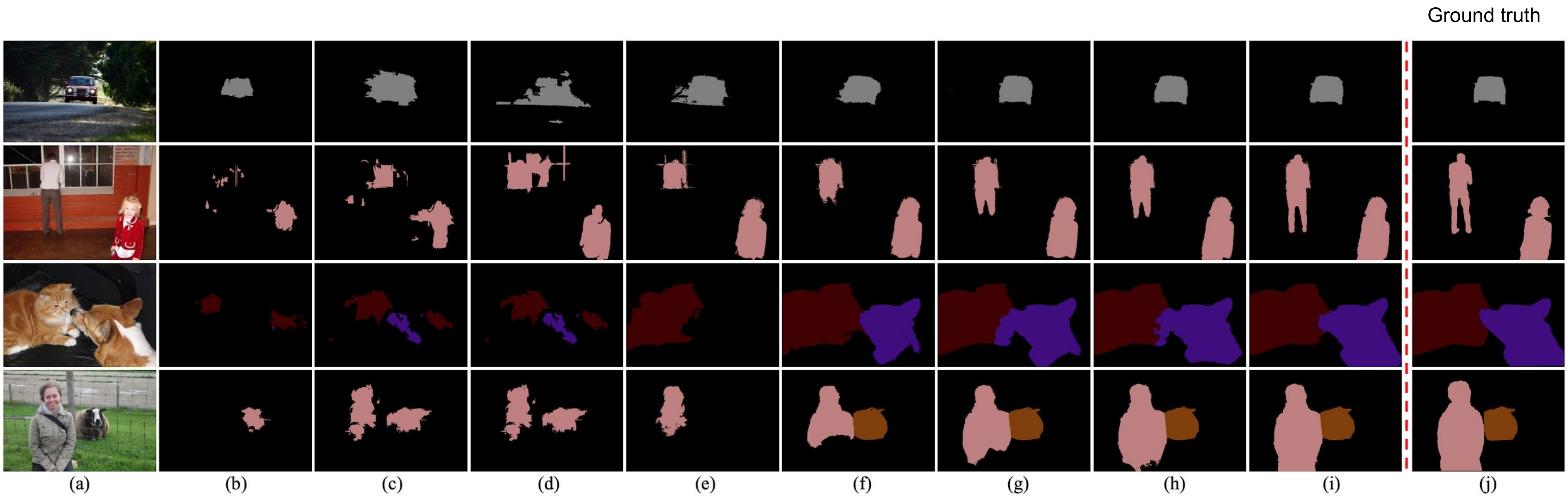


- Wang et al. Weakly-Supervised Semantic Segmentation by Iteratively Mining Common Object Features. CVPR 2018

From global cues to pixel labels

Problem: they focus only on highly discriminative regions

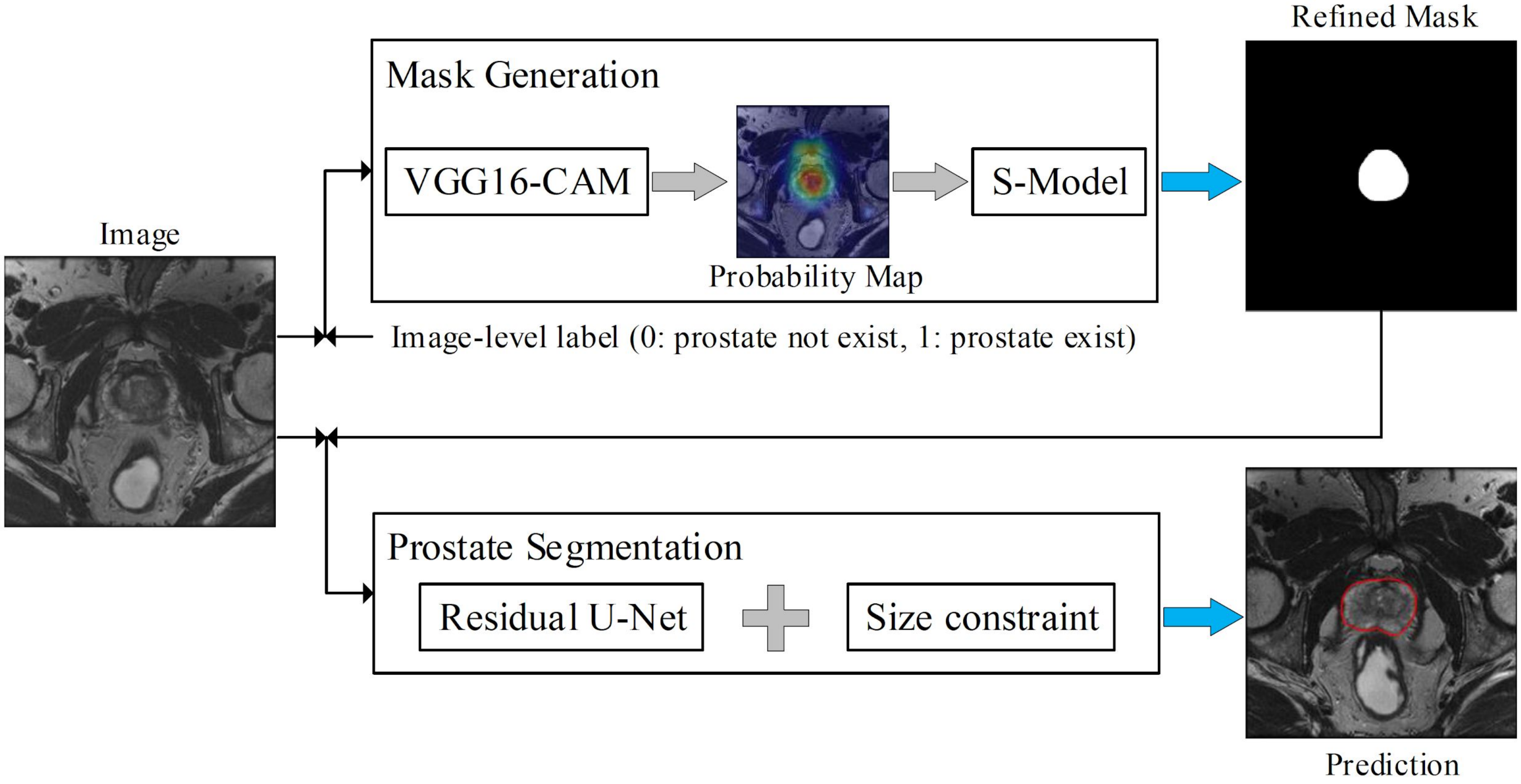
Region mining



- Wang et al. Weakly-Supervised Semantic Segmentation by Iteratively Mining Common Object Features. CVPR 2018

From global cues to pixel labels

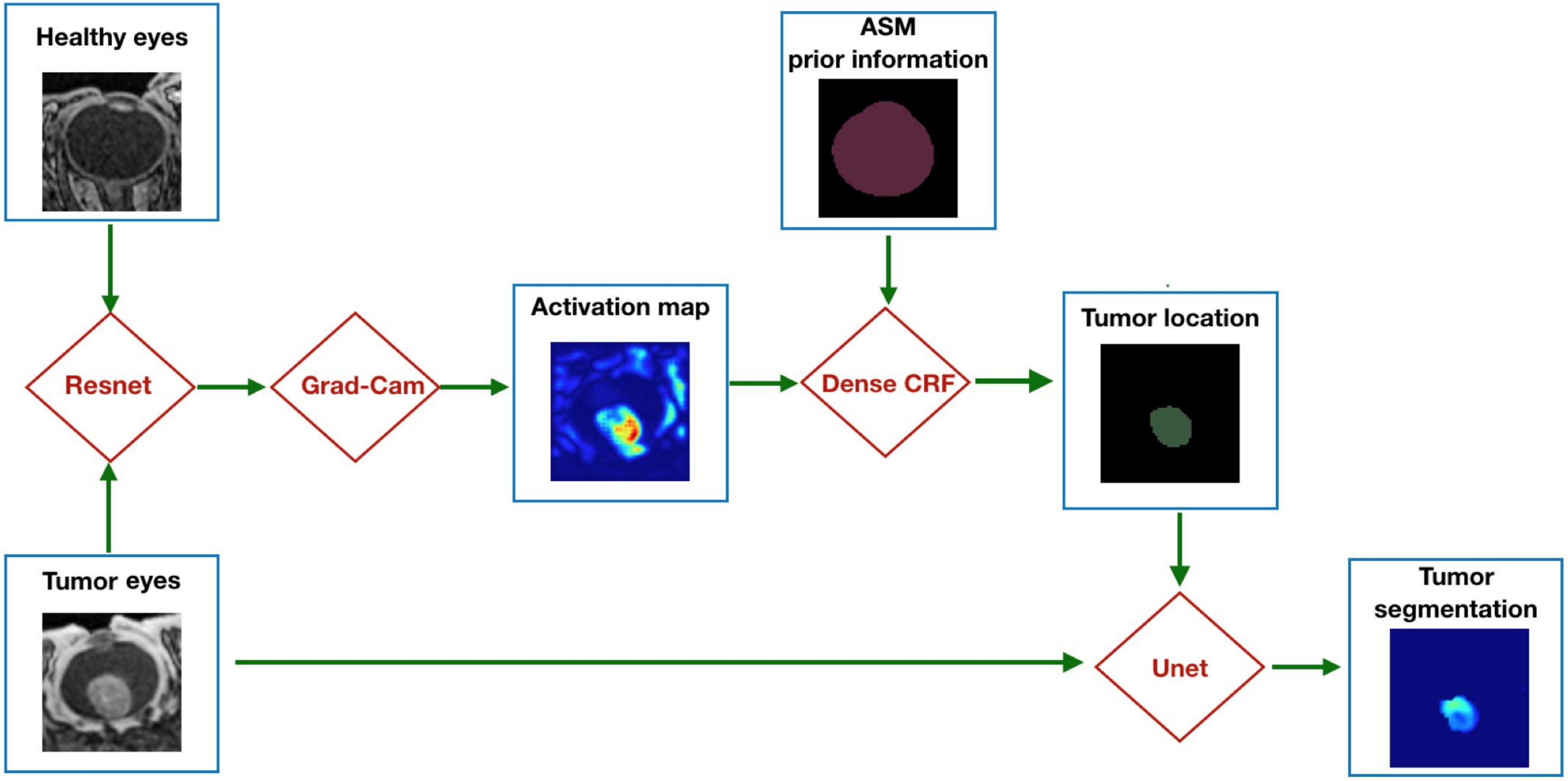
CAMs in the medical domain



- Chen et al. Exploiting confident information for weakly supervised prostate segmentation based on image-level labels. SPIE Medical Imaging 2020

From global cues to pixel labels

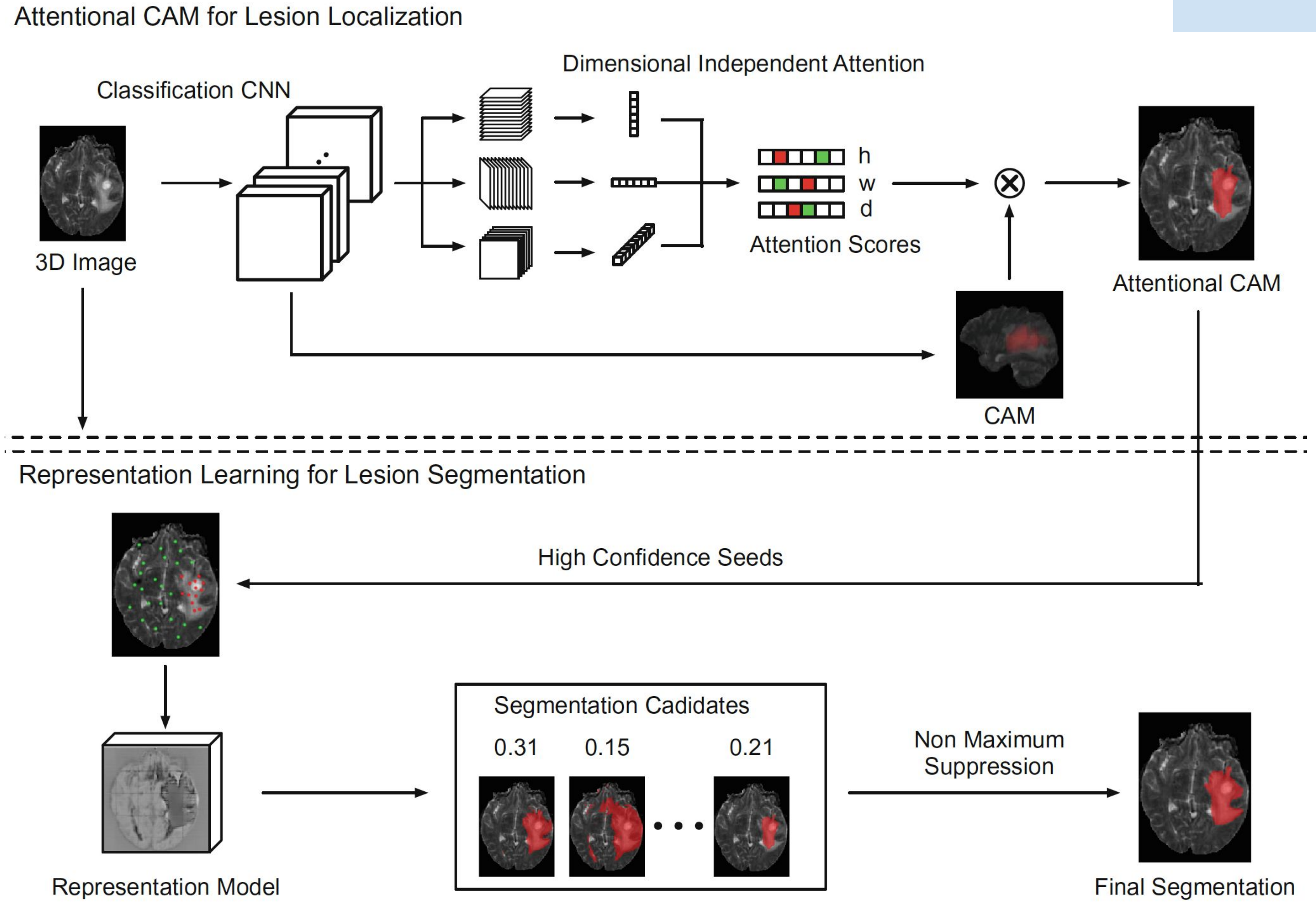
CAMs in the medical domain



- Nguyen et al. A novel segmentation framework for uveal melanoma based on magnetic resonance imaging and class activation maps. MIDL 2019.

From global cues to pixel labels

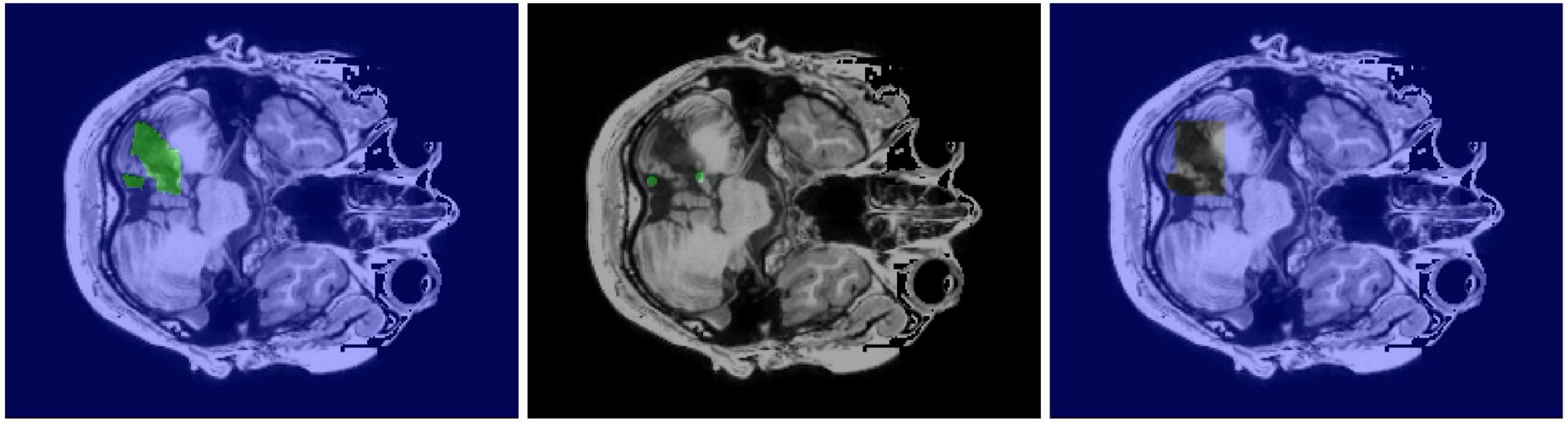
CAMs in the medical domain



- Wu et al. Weakly Supervised Brain Lesion Segmentation via Attentional Representation Learning. MICCAI 2019.

From global cues to pixel labels

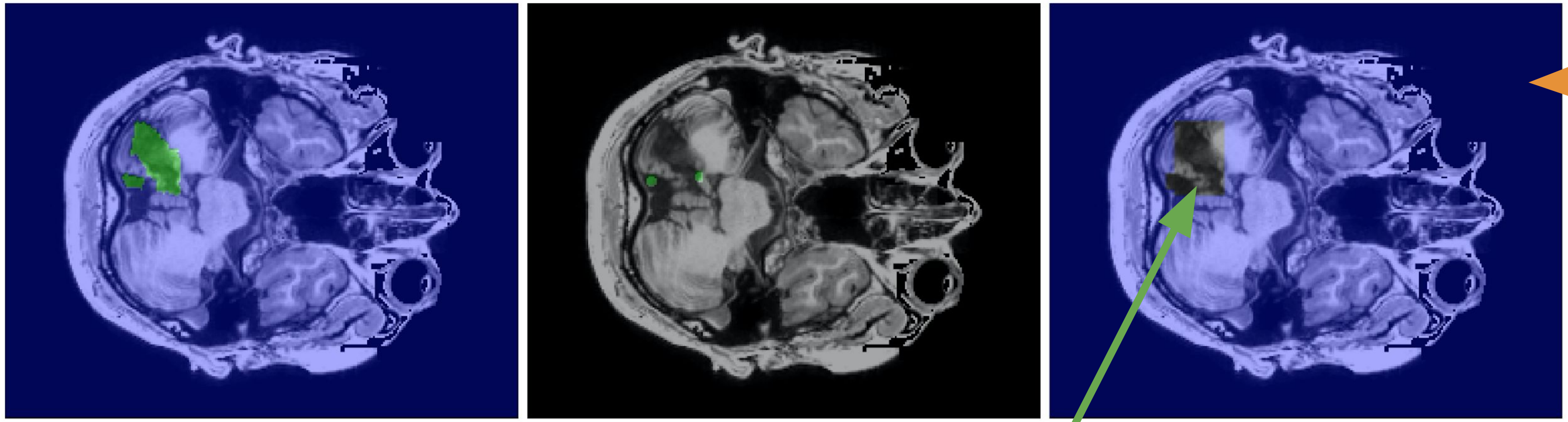
Bounding boxes



- Kervadec et al. Bounding boxes for weakly supervised segmentation: Global constraints get close to full supervision.MIDL 2020.

From global cues to pixel labels

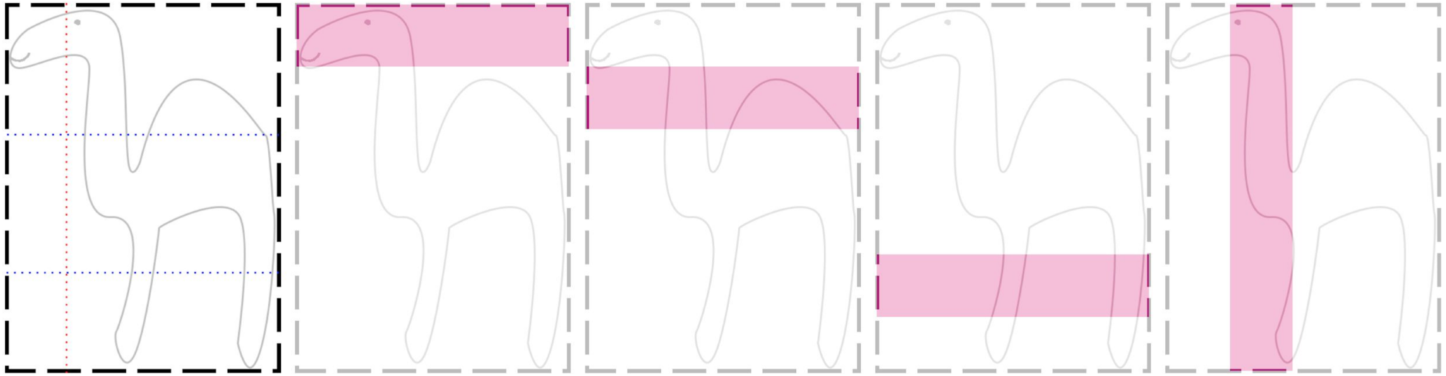
Bounding boxes



$$\sum_{p \in \Omega_O} s_{\boldsymbol{\theta}}(p) \leq 0$$

Outside the bounding box
(Certainty)

Inside the bounding box
(Uncertainty)



Include a tightness prior term

$$\sum_{p \in s_l} y_p \geq w \qquad \forall s_l \in \mathcal{S}_L$$

- Kervadec et al. Bounding boxes for weakly supervised segmentation: Global constraints get close to full supervision.MIDL 2020.