Topology-Preserving Deep Image Segmentation

Xiaoling Hu

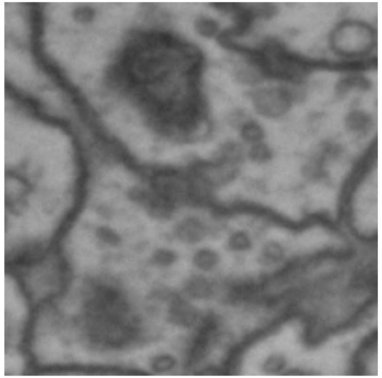
Stony Brook University

NeuIPS'19, Dec. 8, 2019

Joint work with Fuxin Li, Dimitris Samaras & Chao Chen

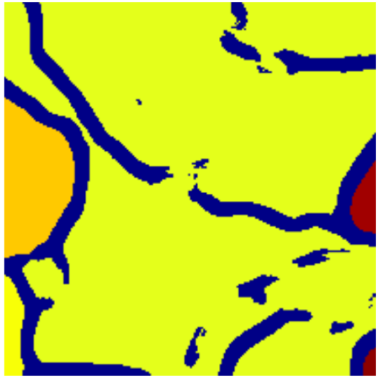
When Deep Learning is Not Enough

- Existing methods optimize w.r.t. per-pixel accuracy
- Topological errors:
  - broken connection, missing components
- Broken membranes:
  - small per-pixel error
  - topological error – large error in downstream analysis



Input Image

Ground Truth



With Topo Errors

Our Method

Segmentation with Convolutional Neural Network (CNN)

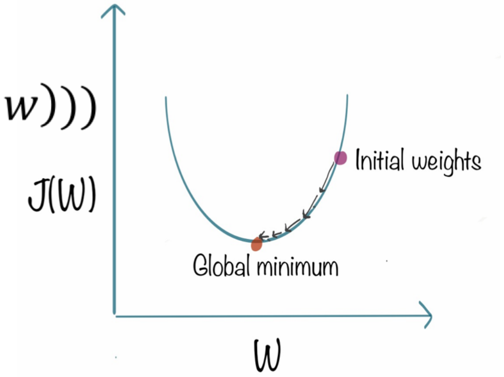
- Likelihood: pixel x belonging to membrane

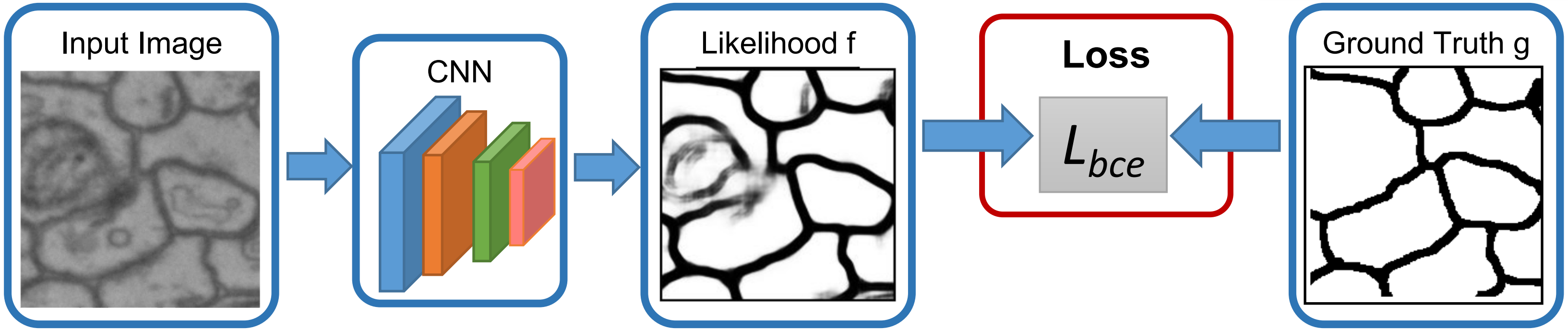
$$f(x,w) = p(y=1|x,w)$$

- Training: optimize cross-entropy loss (t_x true label of x)

$$L_{ce}(w) = -\sum_x ([t_x = 1]\log f(x,w) + [t_x = 0]\log(1 - f(x,w)))$$

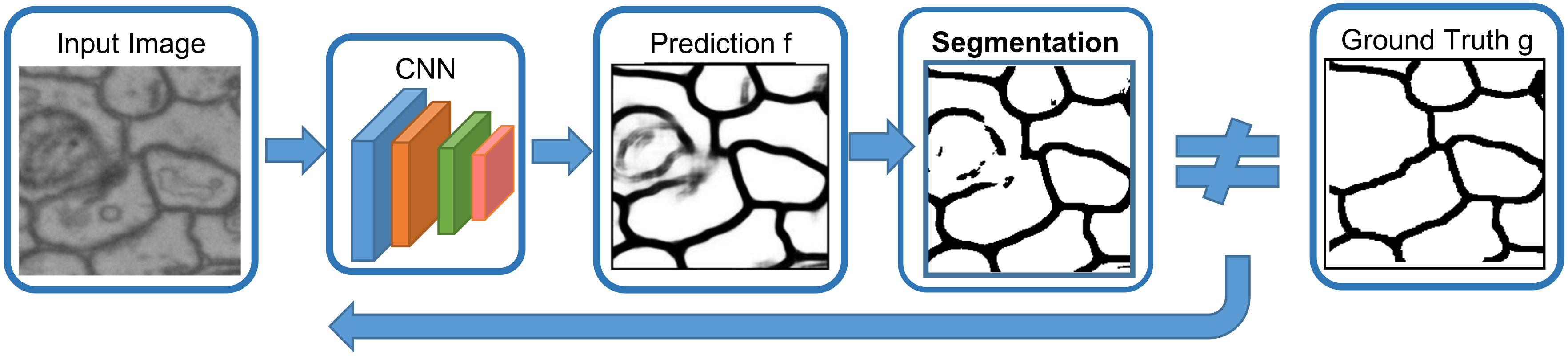
- Gradient descent
- Inference: test img → predict f, threshold at 0.5





Solution: Topological Loss

- loss function – train the model to be topology-preserving
  - Repeatedly evaluate the accuracy in topology and force the model to correct mistakes



- Betti Number Error – comparing segmentation and ground truth
- Challenge: not differentiable
- Solution: persistent homology

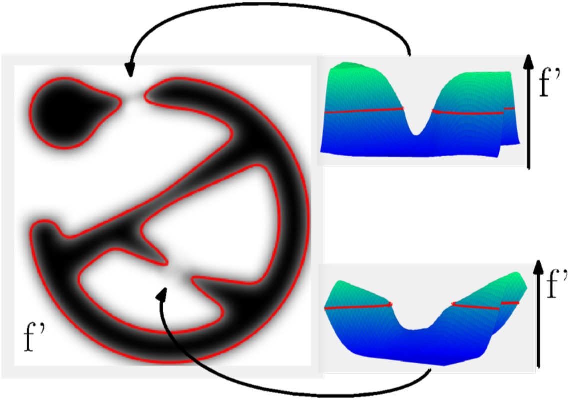
Summary of Contributions

- Optimizing a PH-based objective function:
  - [Poulenard et al. SGP'18] – shape matching
  - [Chen et al. AISTATS'19] – a regularizer for classifiers

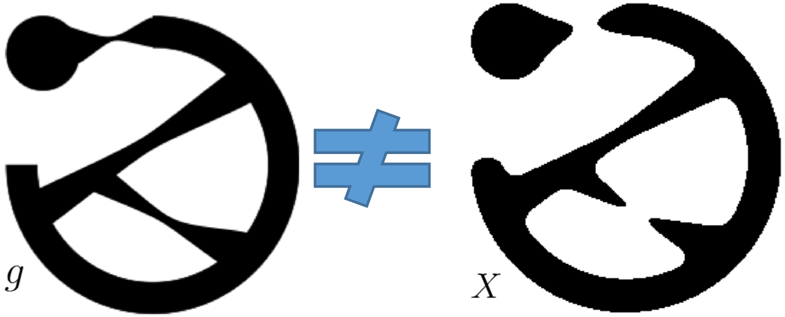
- Our contributions:
  - PH-based **loss function** for end-to-end training of neural networks
  - **Topological loss:** capturing the topology of the training data
  - Adaptive to different local topology – relative homology
  - Empirically: fine structure segmentation

Intuition: focus on likelihood f

- how far f is from a generating a segmentation X with correct topology?
- what is the best way to fix X by changing f?
- Compare with a worse likelihood f’ (same segmentation X)

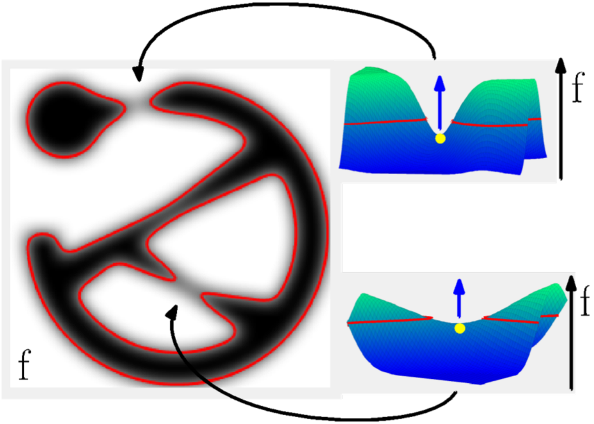


Likelihood f’



Ground Truth

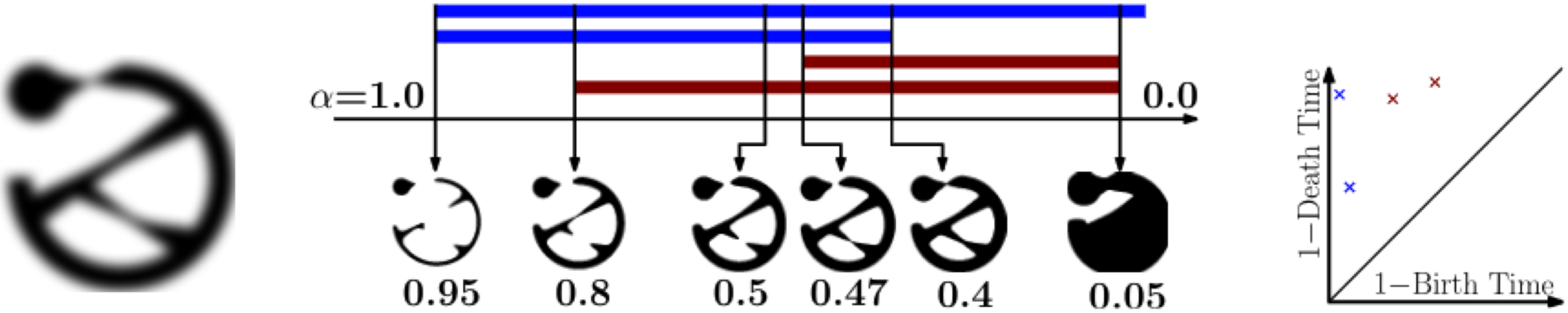
Segmentation



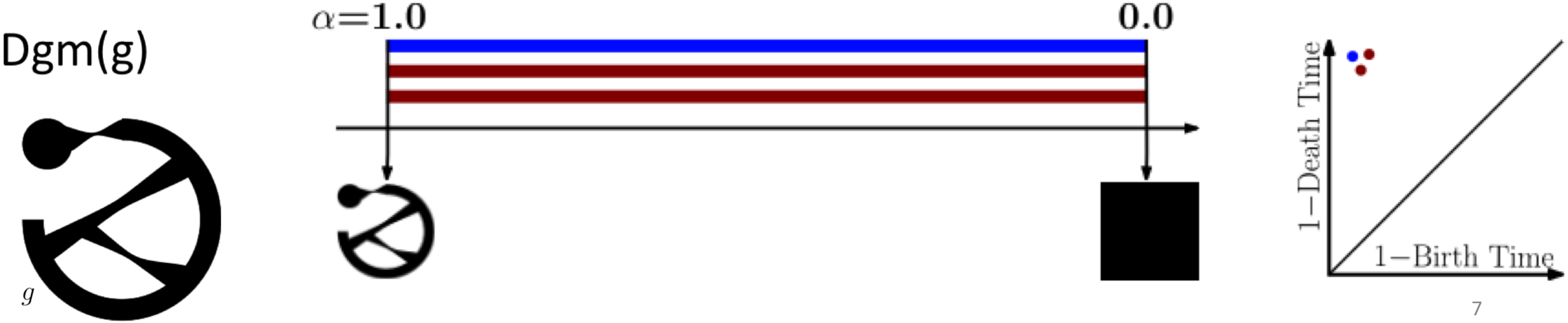
Likelihood f

Persistent Homology

- Superlevel set: $f^{\alpha} = \{ x \mid f(x) \geq \alpha \}$,
- Dgm(f) = {(death(p), birth(p))}, 0-dim, 1-dim



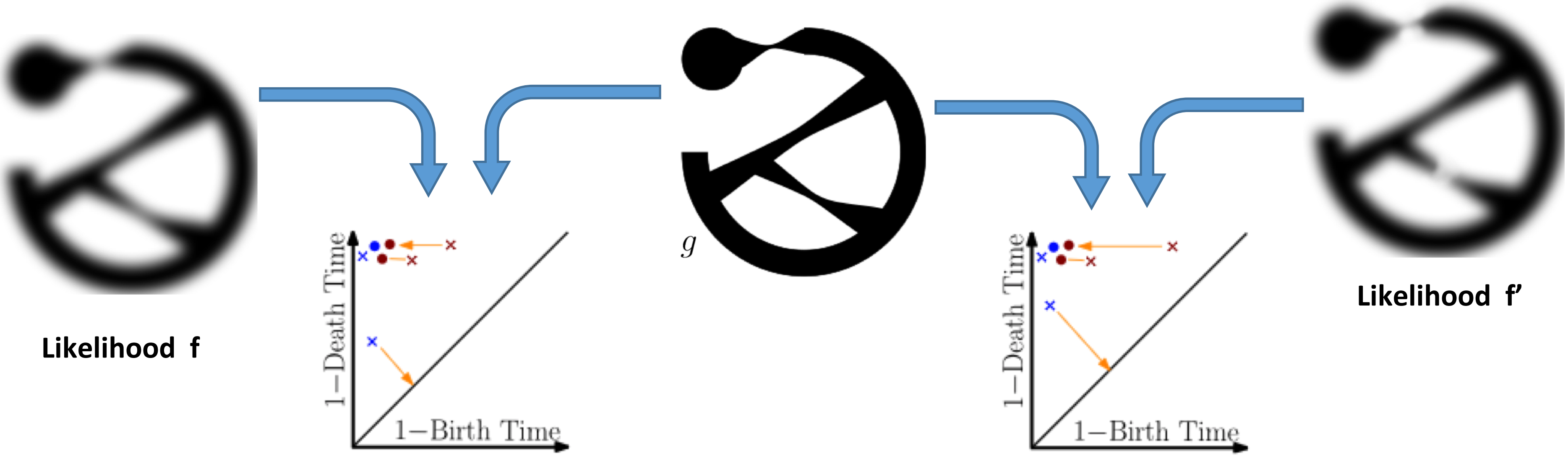
- Dgm(g)



Topological Loss (L_{topo}) = Distance Between Diagrams

$$\min_{\gamma \in \Gamma} \sum_{p \in \mathrm{Dgm}(f)} ||p - \gamma(p)||^2 = \sum_{p \in \mathrm{Dgm}(f)} [\mathrm{birth}(p) - \mathrm{birth}(\gamma^*(p))]^2 + [\mathrm{death}(p) - \mathrm{death}(\gamma^*(p))]^2$$

γ^* -- the optimal matching



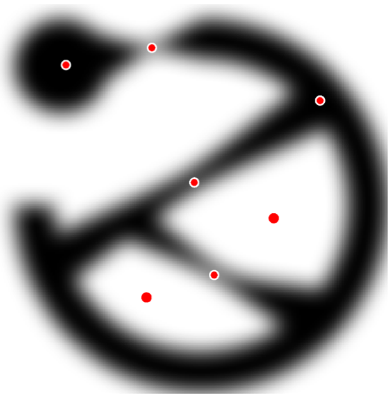
Gradient of L_{topo}

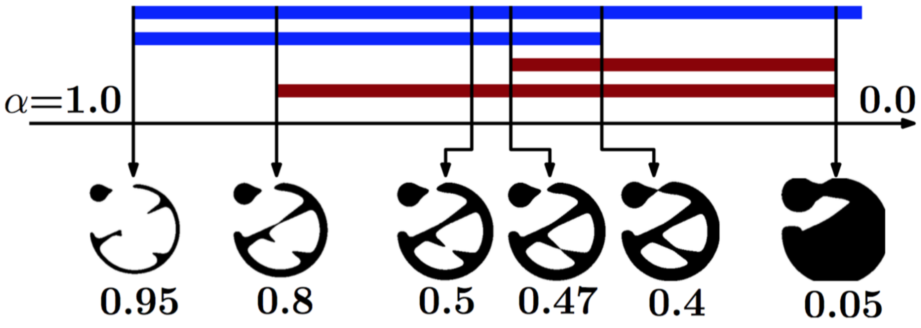
$$\min_{\gamma\in\Gamma} \sum_{p\in \mathrm{Dgm}(f)} ||p-\gamma(p)||^2 = \sum_{p\in \mathrm{Dgm}(f)} [\mathrm{birth}(p) - \mathrm{birth}(\gamma^*(p))]^2 + [\mathrm{death}(p) - \mathrm{death}(\gamma^*(p))]^2$$

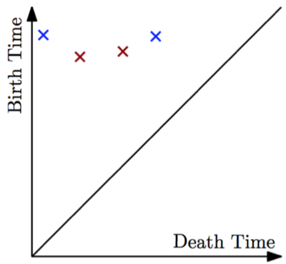
- Recall cross entropy loss:

$$\nabla_w L_{ce} = -\sum_x \left([t_x = 1]\frac{1}{f(x,w)}\frac{\partial f(x,w)}{\partial w} - [t_x = 0]\frac{1}{f(x,w)}\frac{\partial f(x,w)}{\partial w}\right)$$

- Missing link: critical values (birth/death) vs. specific pixels (f(x,w))
- Critical points: locations where birth/death happens

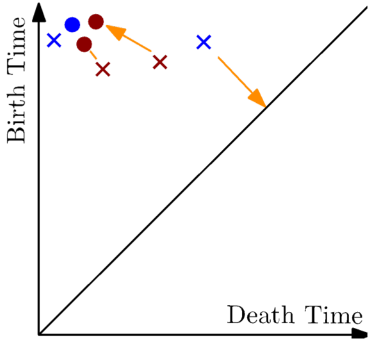


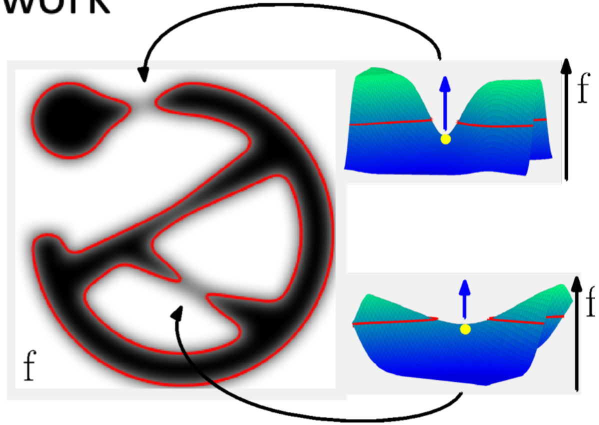




Gradient Cont'd

- $c_b(p)$ – birth critical point of p; $c_d(p)$ – death critical point of p
- Assume γ^* fixed, and $c_b(p)$ $c_d(p)$ are fixed.
- $\nabla_w L_{topo} = \sum_{p \in \mathrm{Dgm}(f)} 2[f(c_b(p)) - \mathrm{birth}(\gamma^*(p))] \boxed{\frac{\partial f(c_b(p))}{\partial \omega}} + 2[f(c_d(p)) - \mathrm{death}(\gamma^*(p))] \boxed{\frac{\partial f(c_d(p))}{\partial \omega}}$
- $\frac{\partial f(x,w)}{\partial w}$ backpropagate through the network

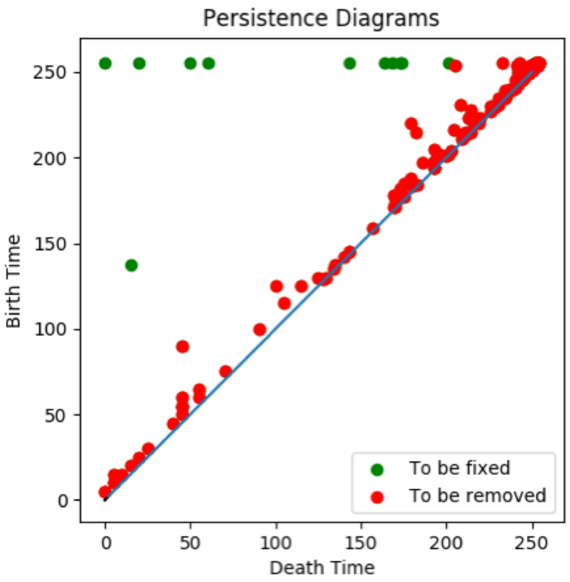




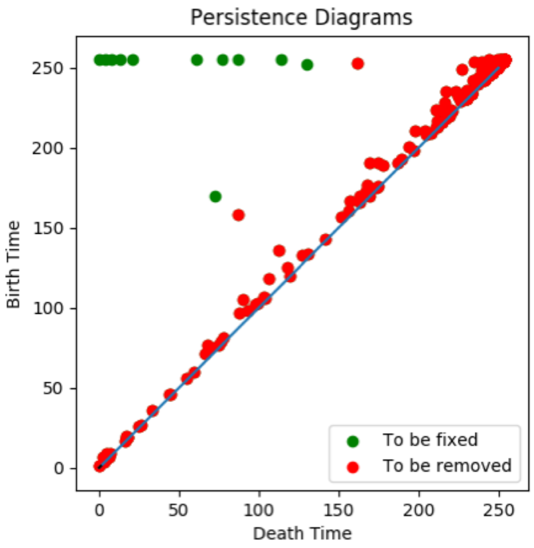
How it looks like in training?

- Diagrams over time

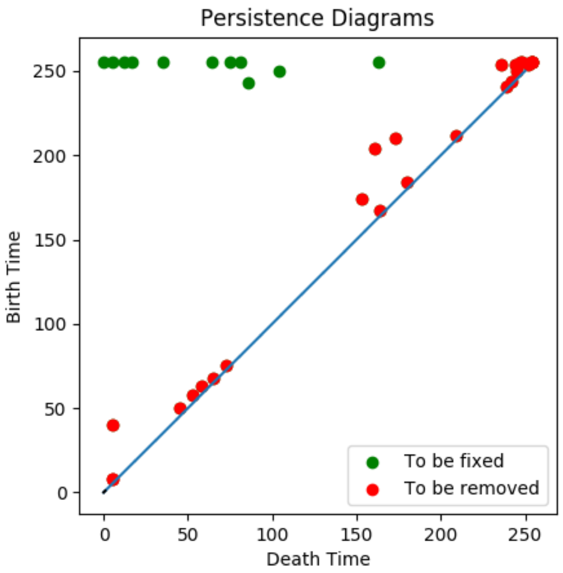
Epoch 10



Epoch 30



Epoch 50

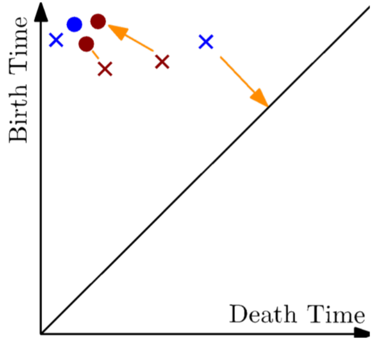


Understanding of diagrams over time

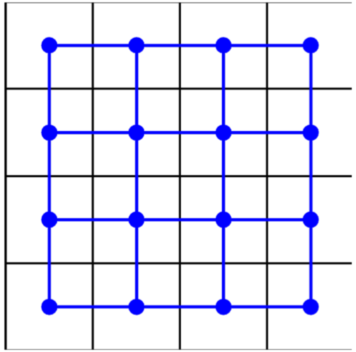
- Green dots in the diagram correspond to critical points where topological structures are incorrect, and to be fixed
- Red dots correspond to noise and to be removed
- As time goes on, there will be fewer red dots as lots of noises were removed
- As time goes on, green dots will be more compact and close to upper left corner

Differentiability?

$$\min_{\gamma\in\Gamma} \sum_{p\in \mathrm{Dgm}(f)} ||p-\gamma(p)||^2 = \sum_{p\in \mathrm{Dgm}(f)} [\mathrm{birth}(p) - \mathrm{birth}(\gamma^*(p))]^2 + [\mathrm{death}(p) - \mathrm{death}(\gamma^*(p))]^2$$



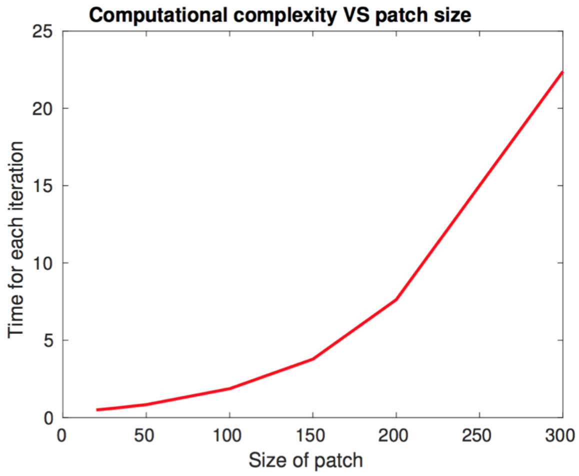
- Assumptions: γ^* fixed, and $c_b(p)$ $c_d(p)$ are fixed.
- Are they true? No in general. In image setting, Yes.
- Differentiability proof sketch:
  - f is piecewise linear, controlled by f values at all pixels
  - Assume all pixel values are different; matching unambiguous
  - Within a small nbd of f, critical pixels and γ^* are constant
  - Degenerate cases: $f(x) = f(y)$ for some $x \neq y$, ambiguous matching --- measure zero set
  - Differentiable almost everywhere.

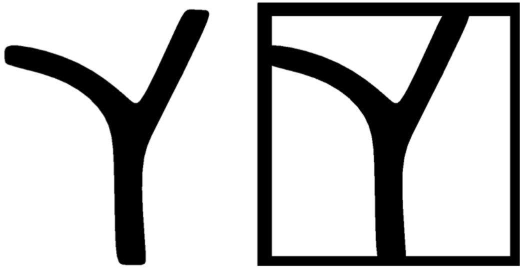


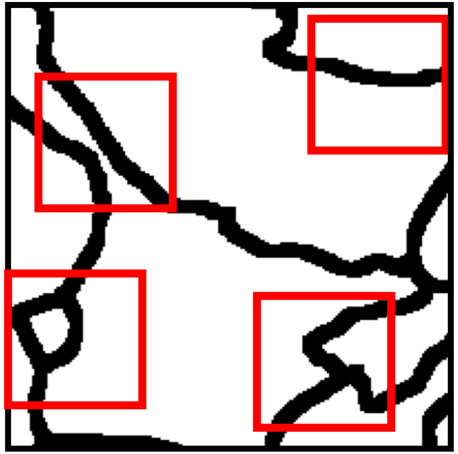
- Guaranteed correctness: zero loss = correct topology

Localized topological loss

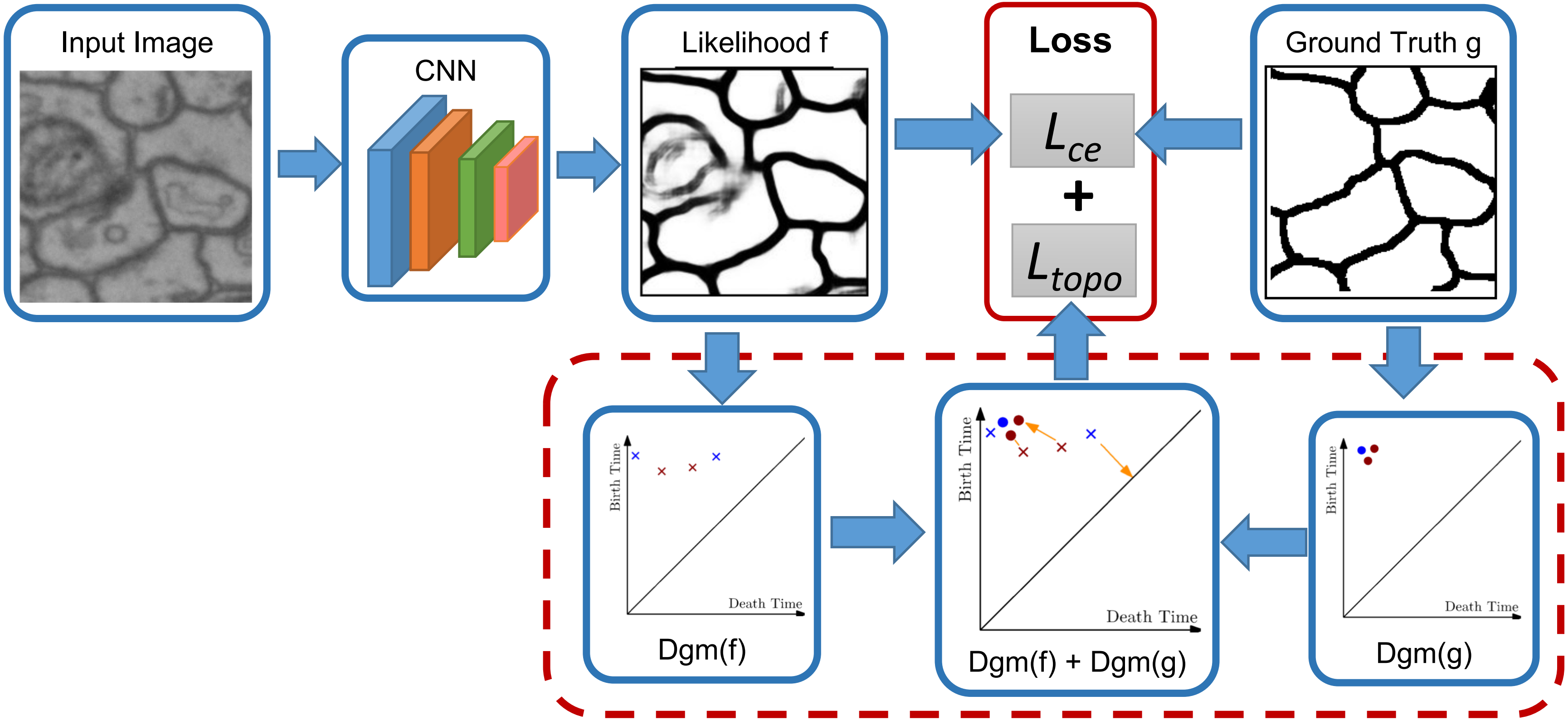
- Evaluate topology in whole image
  Expensive; Too many dots: matching is difficult.
- Localized topological measure: relative homology
- Sample patches over images and evaluate topology loss





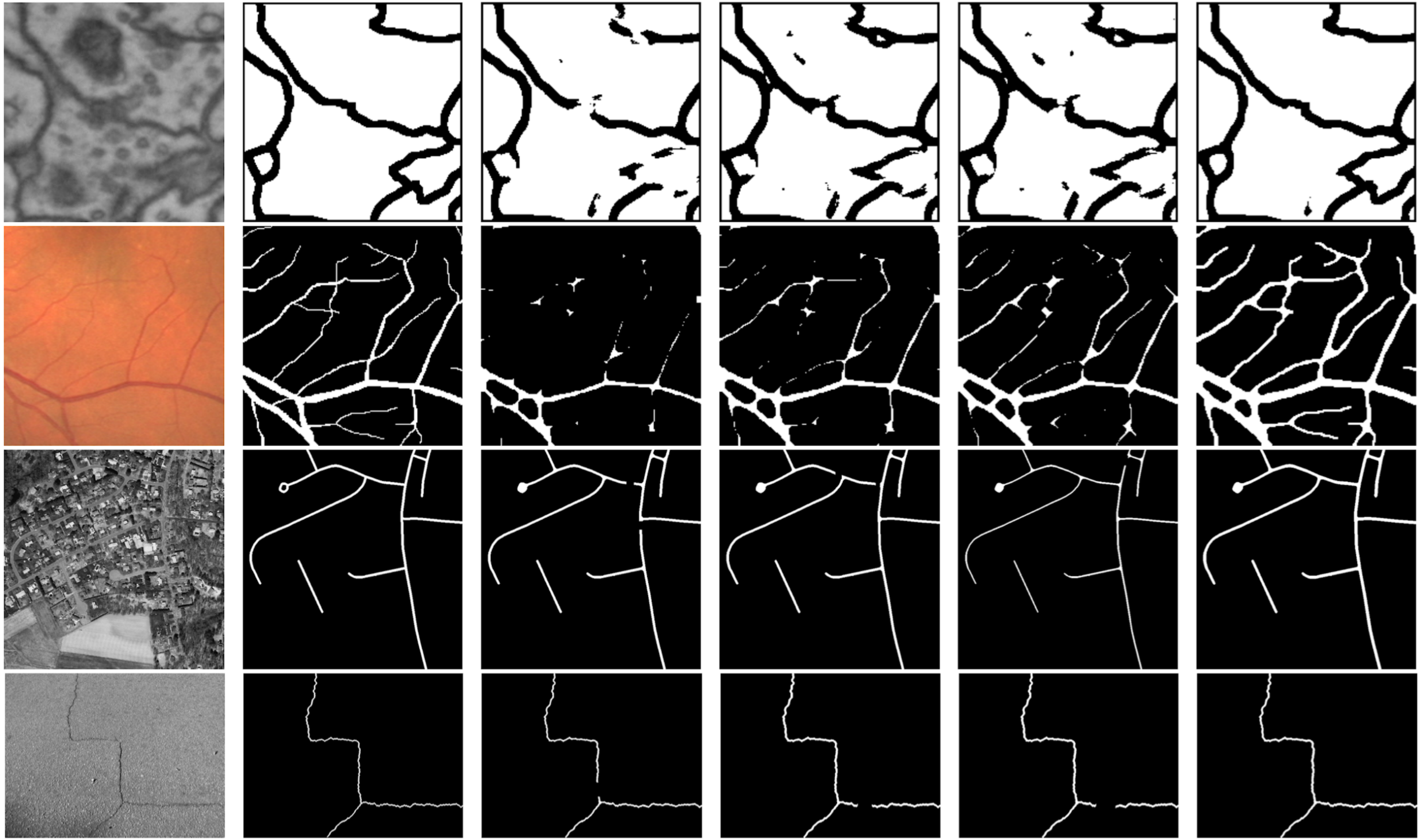


Overview of the System



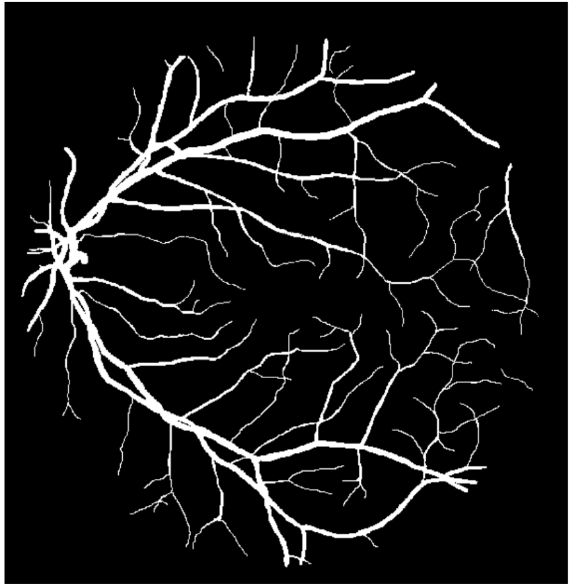
Input Image **Ground Truth** **DIVE** **U-Net** **Mosin.** **TopoNet**

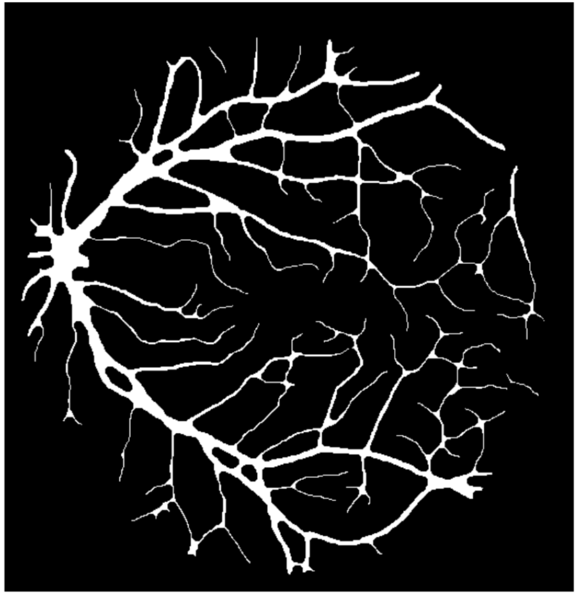
[Bioinformatics’16] [MICCAI’15] [CVPR’18]



More Results

- tend to connects well;
- overconfident/mistakes with with small holes/junctions





Input Image

Ground Truth

TopoNet

Quantitative Results

- Per-pixel error, Betti number error, Adjusted Rand Index, Variation of Information
- EM (neuron) image segmentation datasets

Dataset	Method	Accuracy	ARI	VOI	Betti Error
ISBI12	**DIVE**	0.9640	0.9434	1.235	3.187
	U-Net	**0.9678**	0.9338	1.367	2.785
	Mosin.	0.9532	0.9312	0.983	1.238
	TopoNet	0.9626	**0.9444**	**0.782**	**0.429**
ISBI13	**DIVE**	**0.9642**	0.6923	2.790	3.875
	U-Net	0.9631	0.7031	2.583	3.463
	Mosin.	0.9578	0.7483	1.534	2.952
	TopoNet	0.9569	**0.8064**	**1.436**	**1.253**
CREMI	**DIVE**	**0.9498**	0.6532	2.513	4.378
	U-Net	0.9468	0.6723	2.346	3.016
	Mosin.	0.9467	0.7853	1.623	1.973
	TopoNet	0.9456	**0.8083**	**1.462**	**1.113**

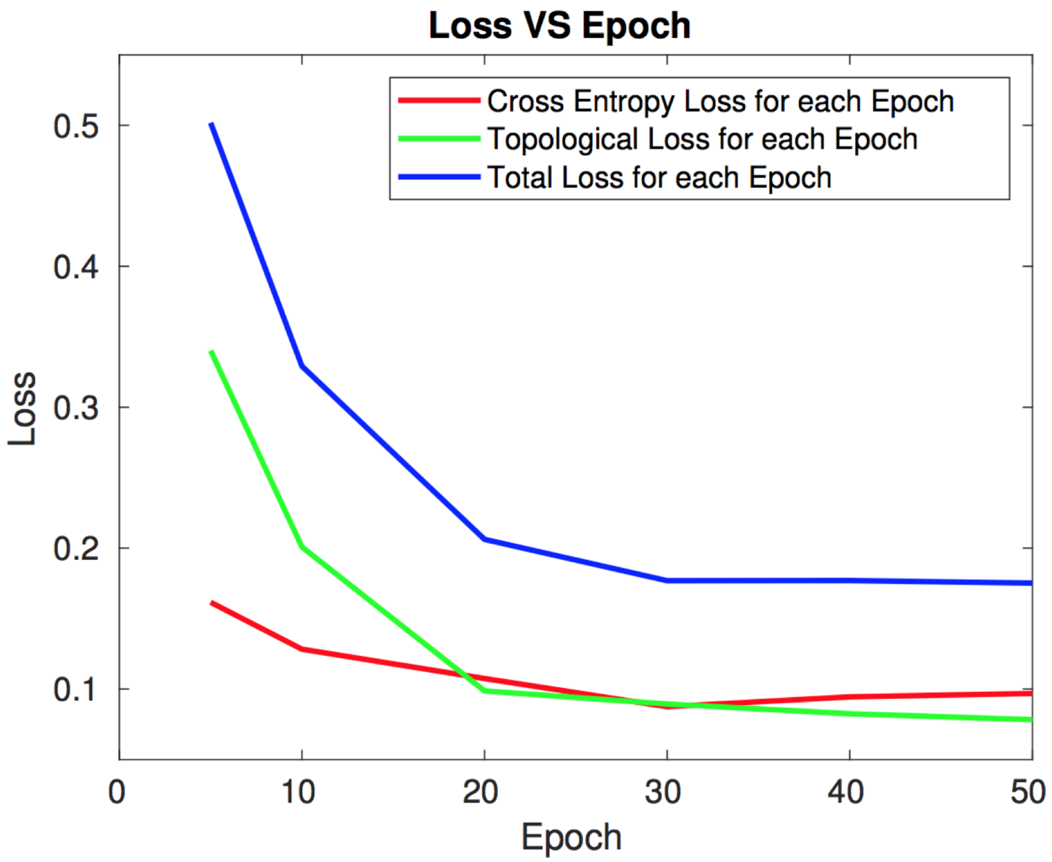
Quantitative Results

- Per-pixel error, Betti number error, Adjusted Rand Index, Variation of Information
- Retinal vessel, crack detection from pavement images, roads in aerial images

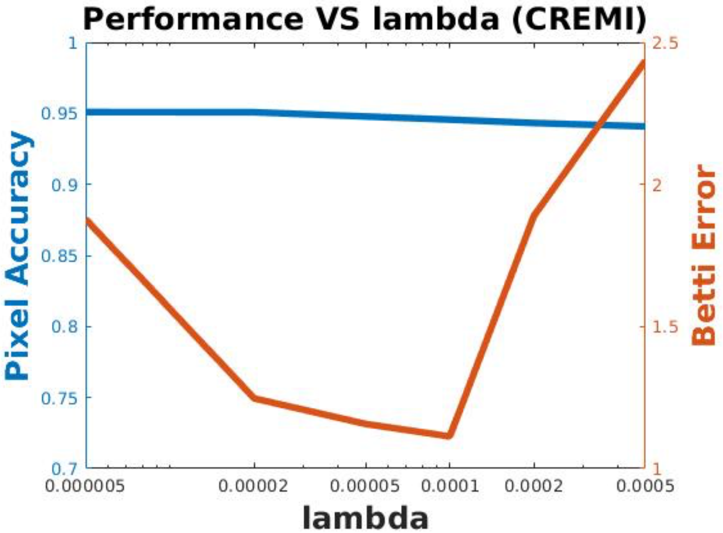
Dataset	Method	Accuracy	ARI	VOI	Betti Error
DRIVE	**DIVE**	**0.9549**	0.8407	1.936	3.276
	U-Net	0.9452	0.8343	1.975	3.643
	Mosin.	0.9543	0.8870	1.167	2.784
	TopoNet	0.9521	**0.9024**	**1.083**	**1.076**
CrackTree	**DIVE**	**0.9854**	0.8634	1.570	1.576
	U-Net	0.9821	0.8749	1.625	1.785
	Mosin.	0.9833	0.8897	1.113	1.045
	TopoNet	0.9826	**0.9291**	**0.997**	**0.672**
Road	**DIVE**	0.9734	0.8201	2.368	3.598
	U-Net	**0.9786**	0.8189	2.249	3.439
	Mosin.	0.9754	0.8456	1.457	2.781
	TopoNet	0.9728	**0.8671**	**1.234**	**1.275**

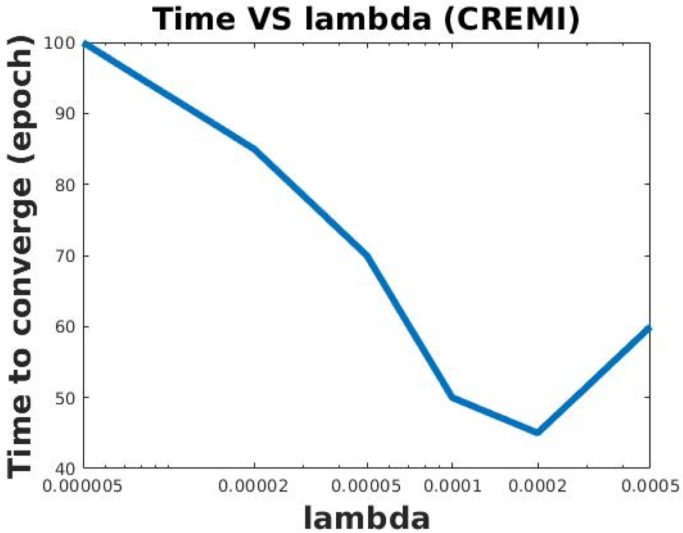
Final Note

- Topology vs per-pixel
  - $L = L_{ce} + \lambda L_{topo}$
- $\lambda \approx 10^{-4}$
- Patch size dependent
- Dataset dependent



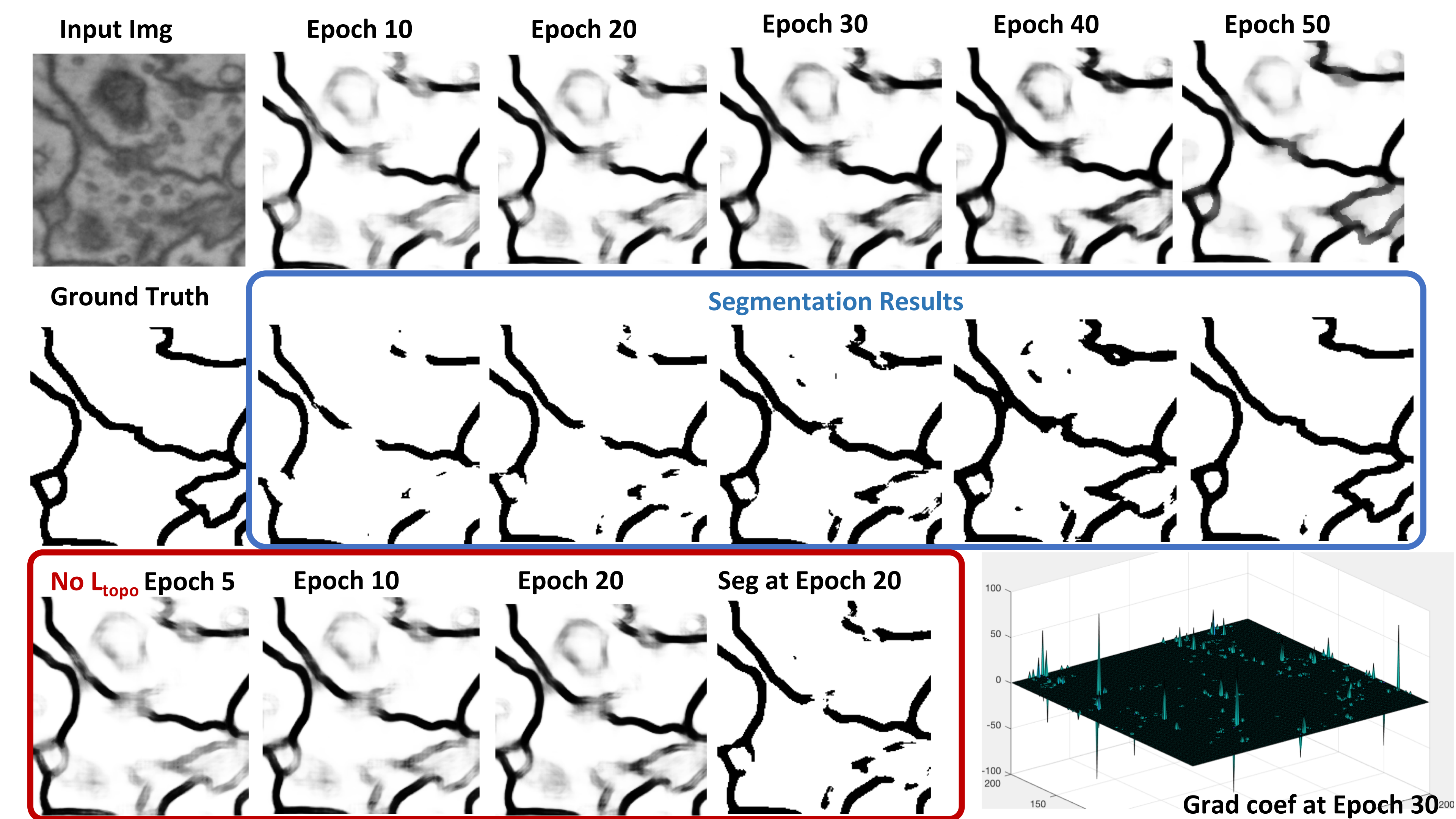
Ablation study of loss weights





Rationale of topological loss

- Within short period, the likelihood map and the segmentation are stabilized globally, mostly because of cross-entropy loss
- After several epochs, topological errors are gradually fixed by the topological loss. Notice the change of the likelihood map is only at specific topology-relevant locations
- Topological loss compliments cross-entropy loss by combating sampling bias
- Without cross-entropy loss, inferring topology from a completely random likelihood map is meaningless.


Input Img
Epoch 10
Epoch 20
Epoch 30
Epoch 40
Epoch 50
Ground Truth
Segmentation Results
No L_{topo} Epoch 5
Epoch 10
Epoch 20
Seg at Epoch 20
100
50
0
-50
-100
200
150
200
Grad coef at Epoch 30

Thanks!

Q&A