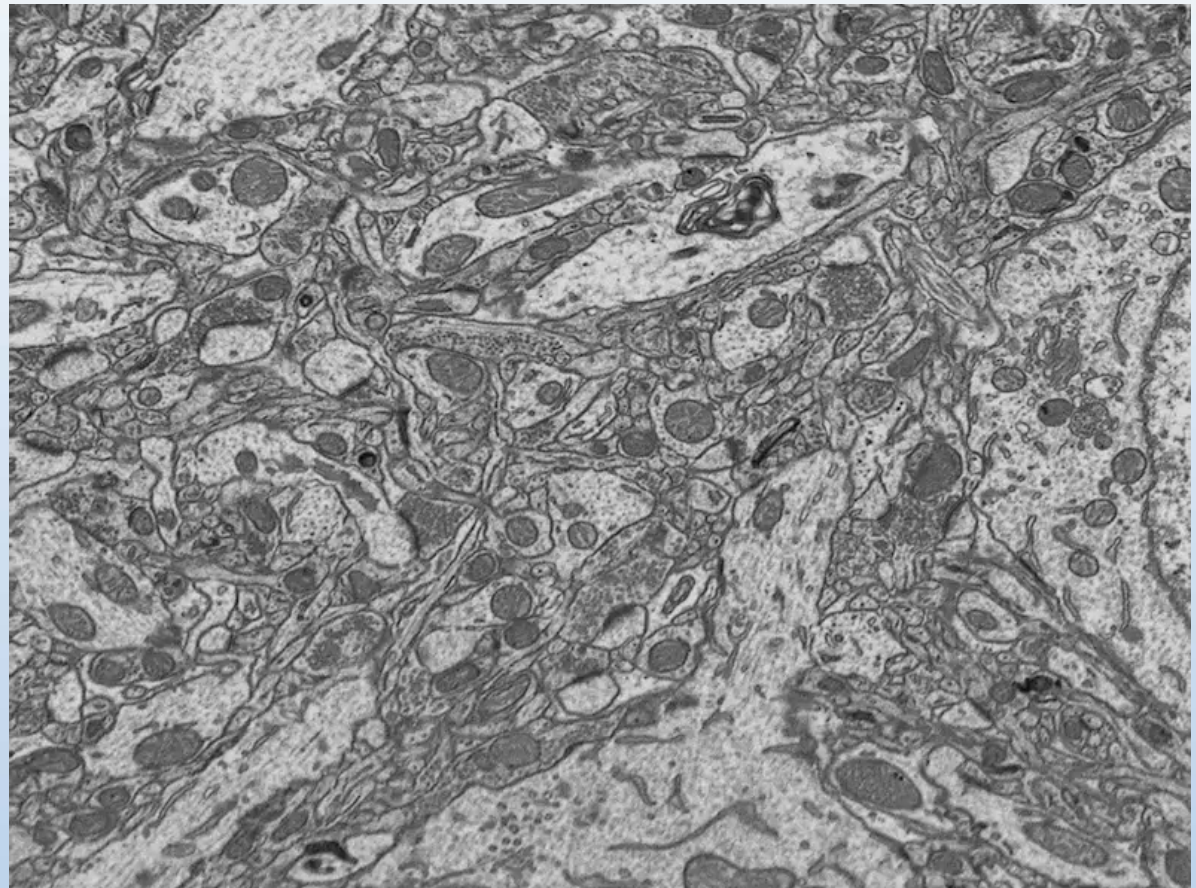
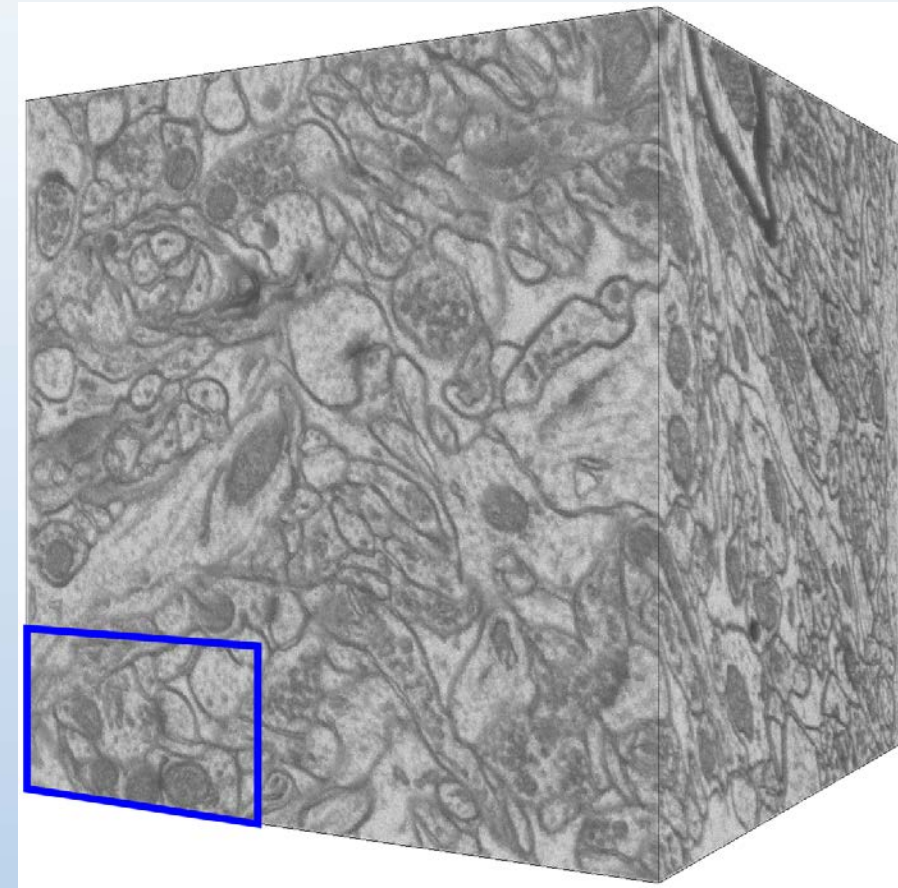


Domain Adaptation and Active Learning for Microscopy

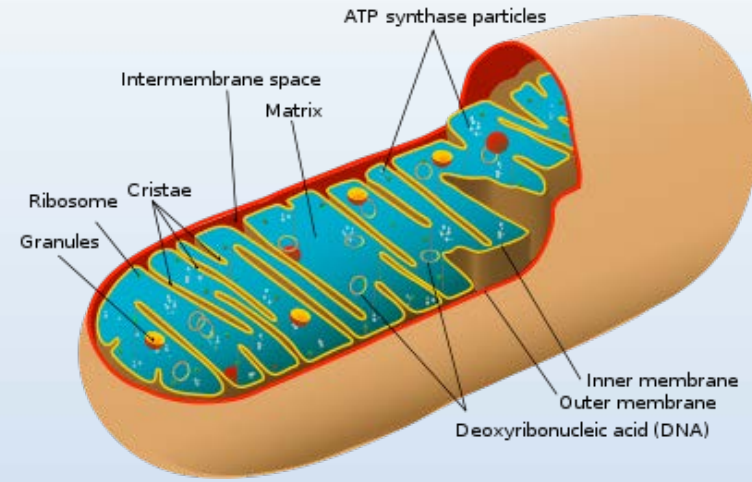
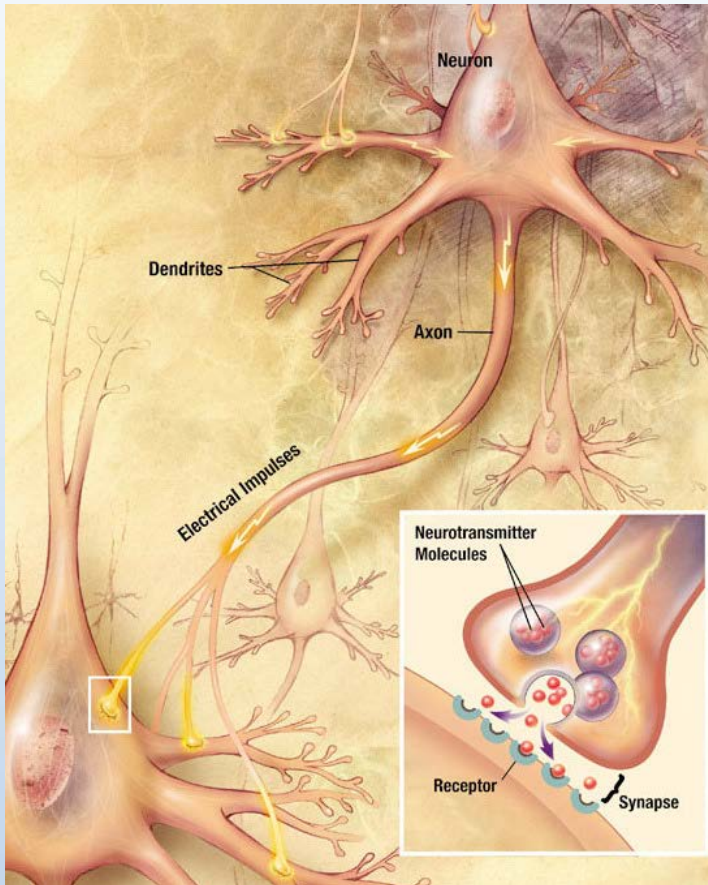
P. Fua
IC-CVLab
EPFL

Electron Microscopy



FIBSEM stack at 5 nm resolution.

Synapses and Mitochondria



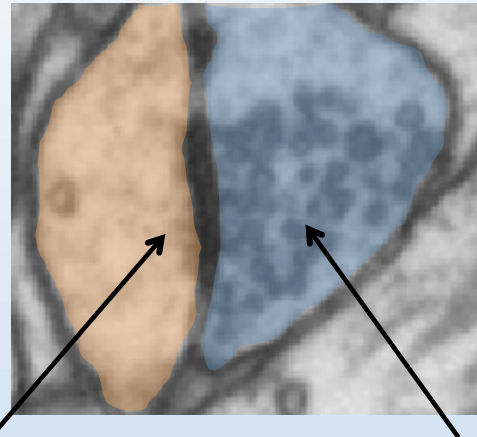
- Synapses transmit signals from neurons to other cells.
- Mitochondria provide cells with the energy they need.
- Their shape is affected by neuro-degenerative diseases.

Talk Outline

1. Segmentation given sufficient amounts or training data.
2. Domain adaptation
3. Active Learning

Context-Based Features

Synapse:

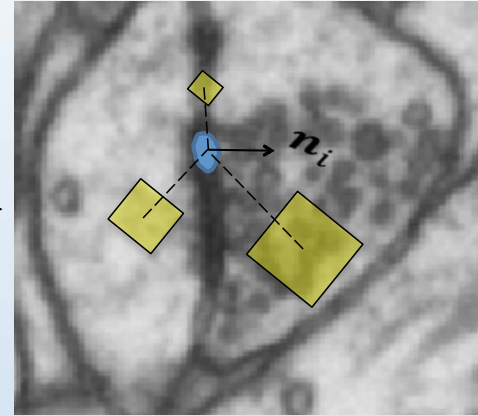
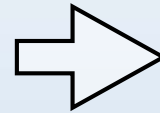


Post-synaptic region

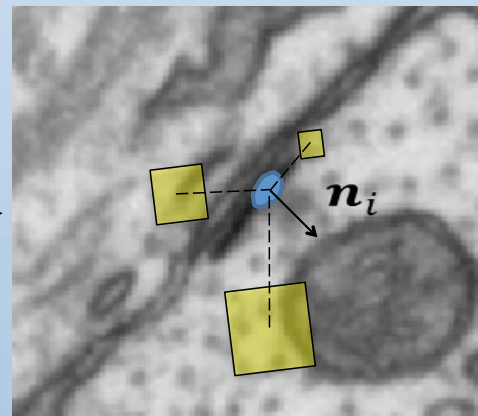
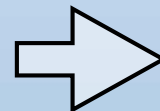
- Dendrite
- No vesicles

Pre-synaptic region

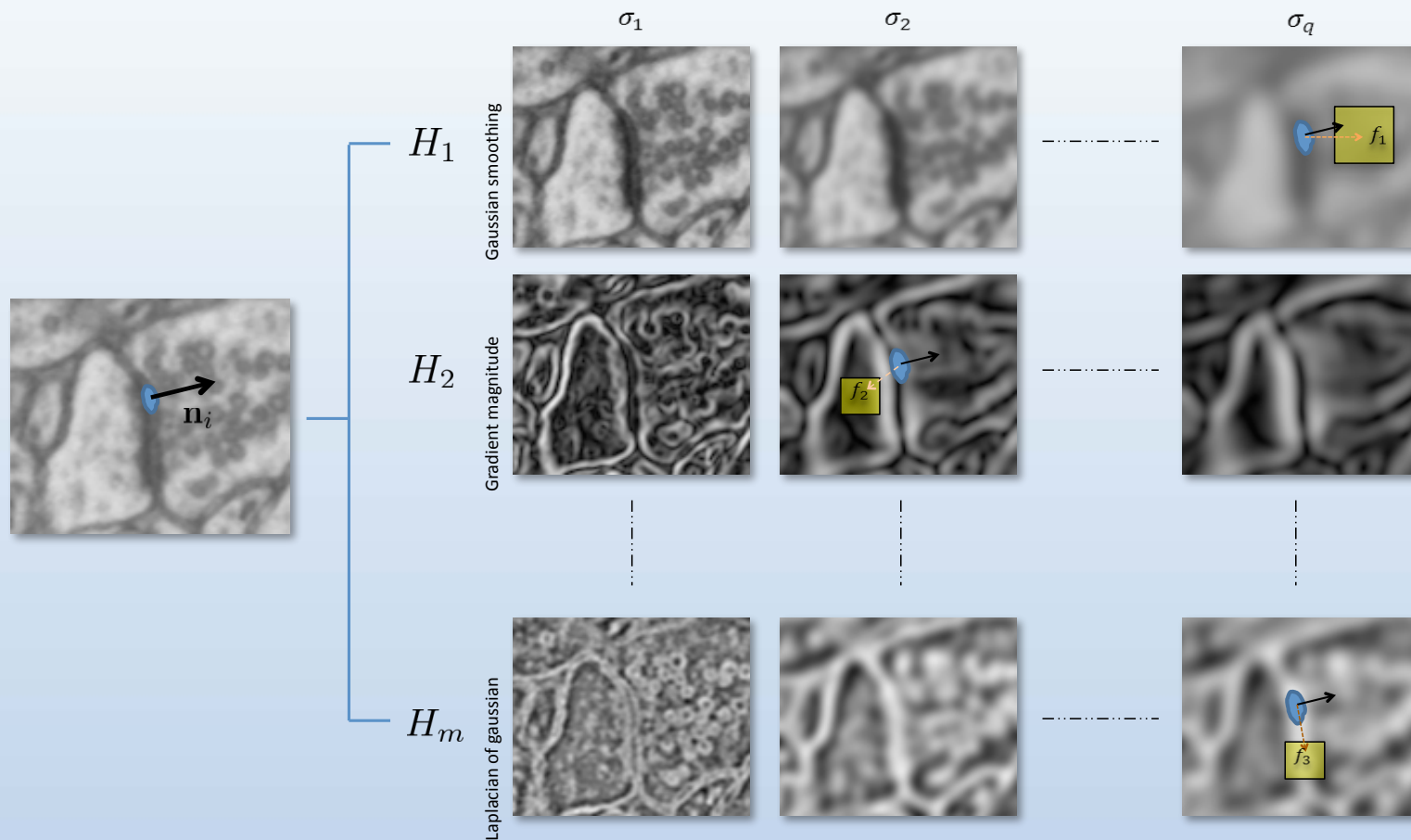
- Axon terminal
- Many vesicles



Non-Synapse:



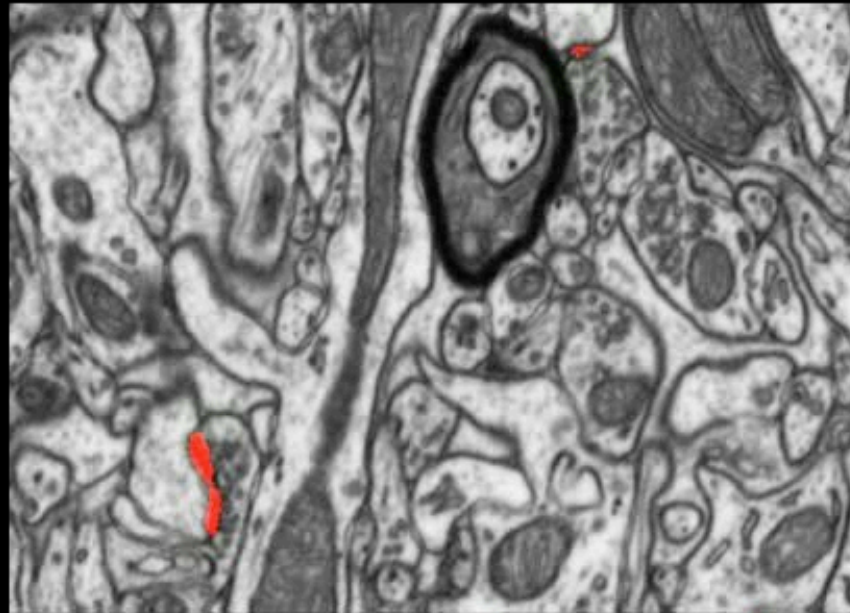
Contextual Features



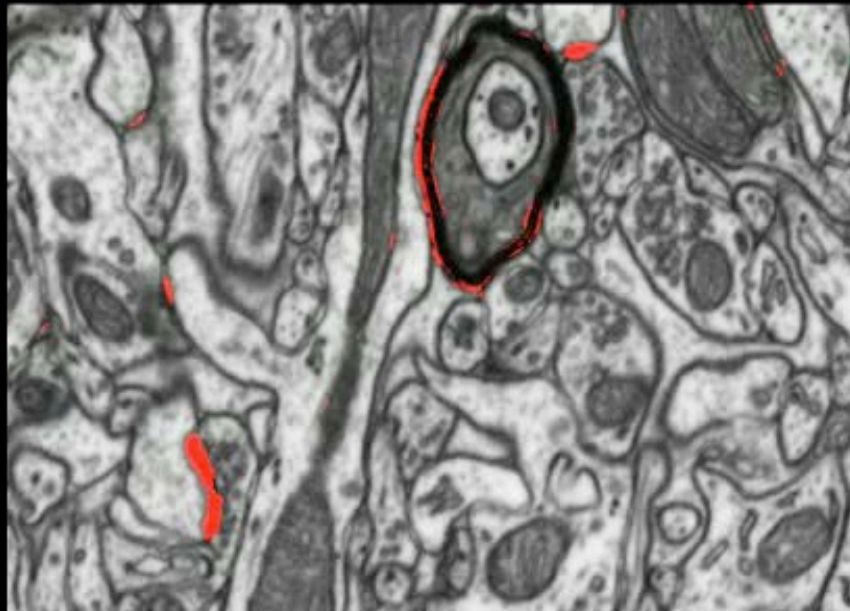
- Context features are based on image statistics inside a cubic neighborhood centered around a location relative to the voxel to be classified.
- Many different channels are computed and AdaBoost selects the location, channel and cubic neighborhood size.

Synaptic vs Local Features

Our method

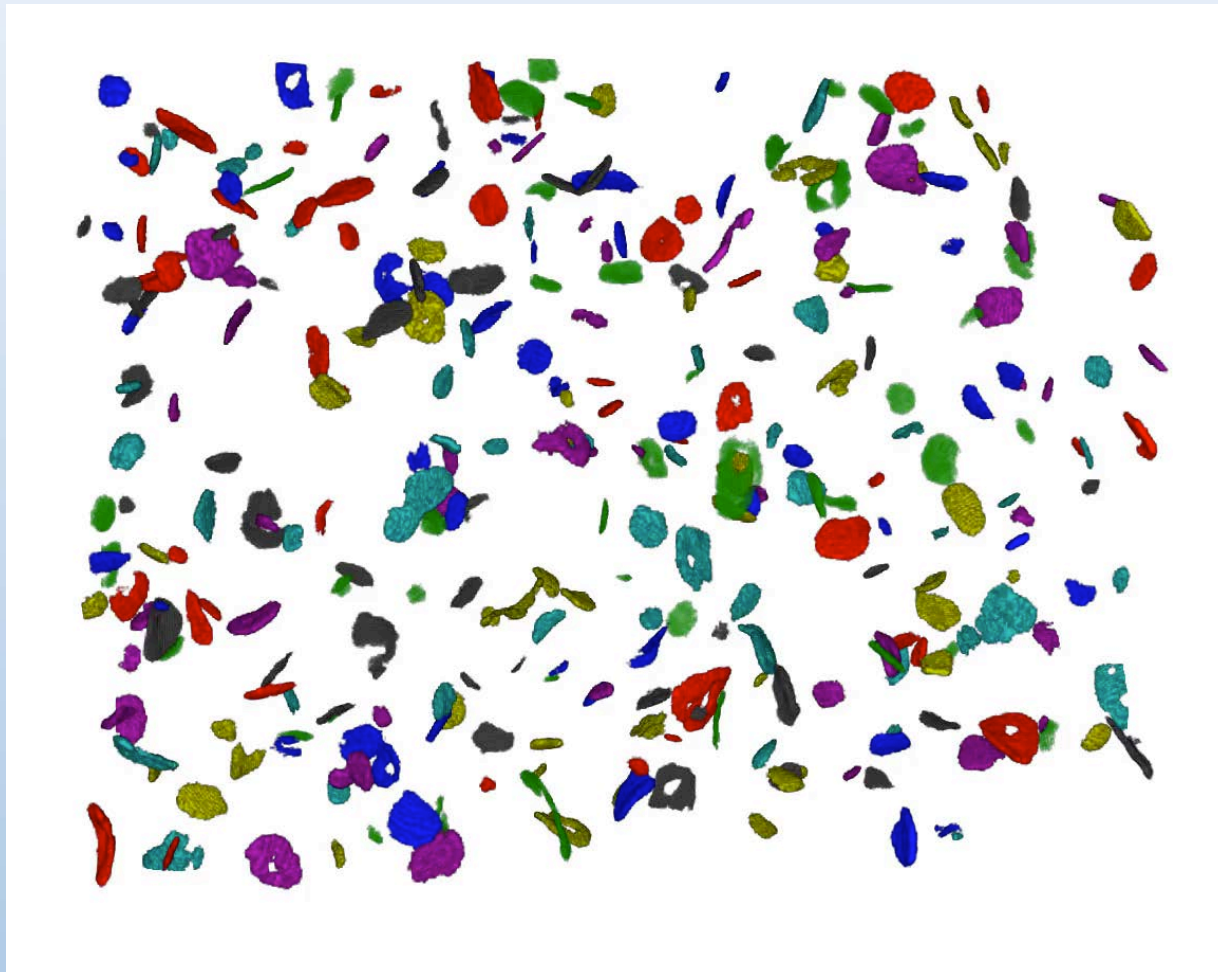
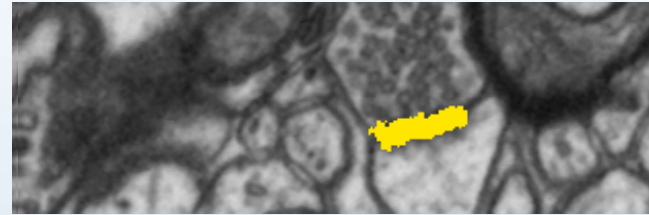
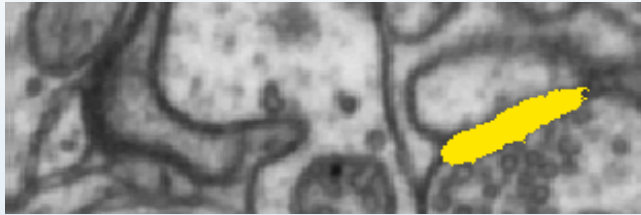


Kreshuk et al.



Threshold set at 90% TPR
for both methods

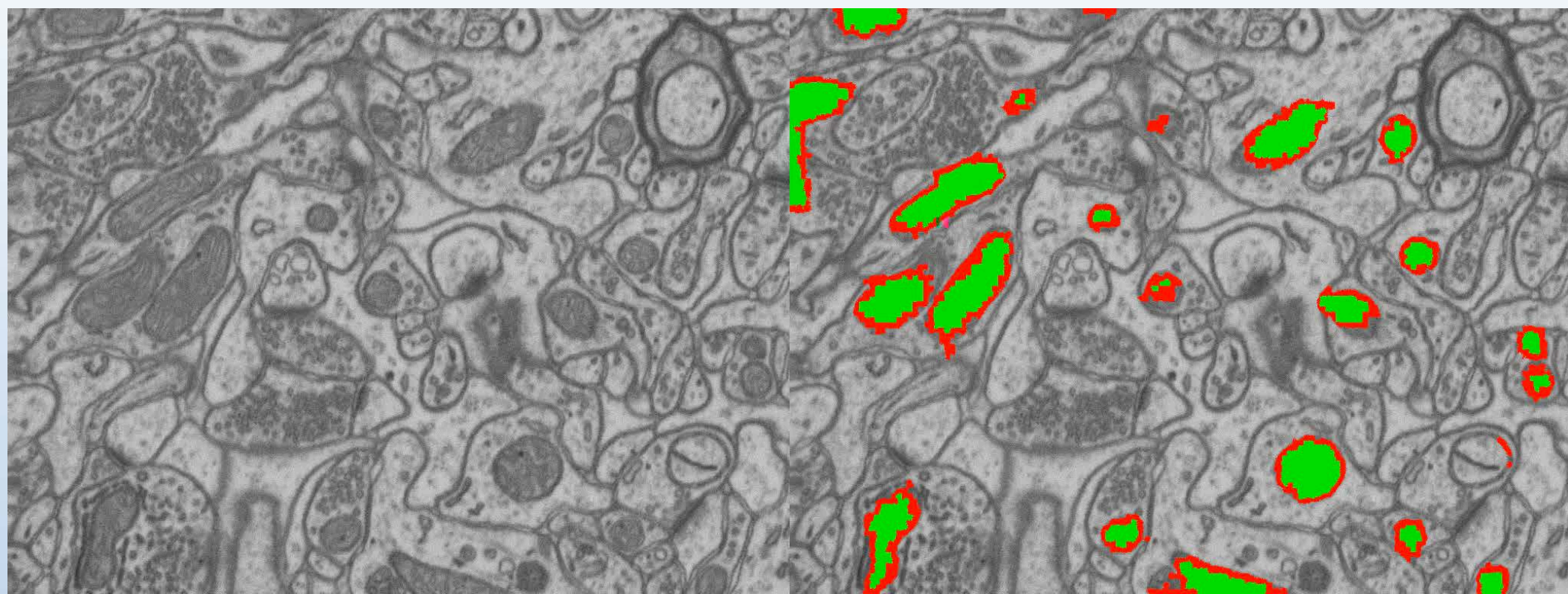
3D Detection



Quantitative Evaluation

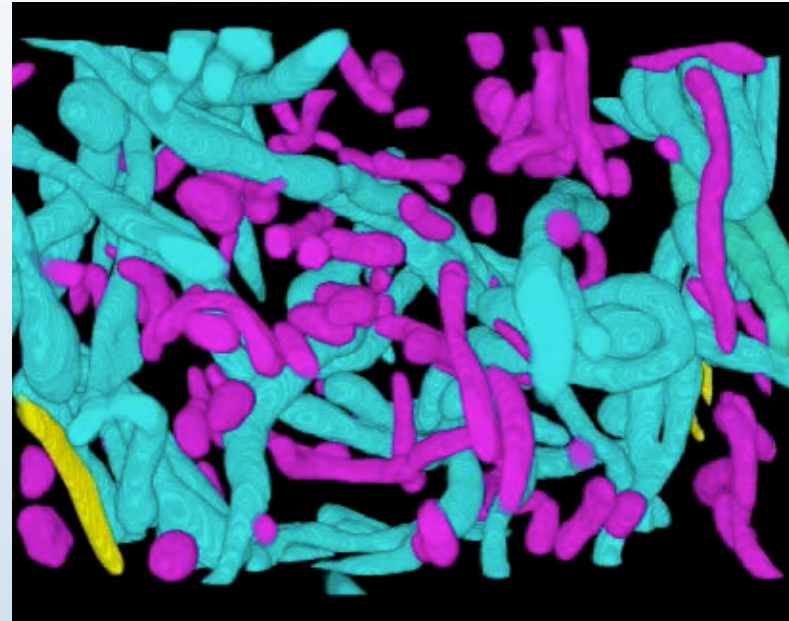
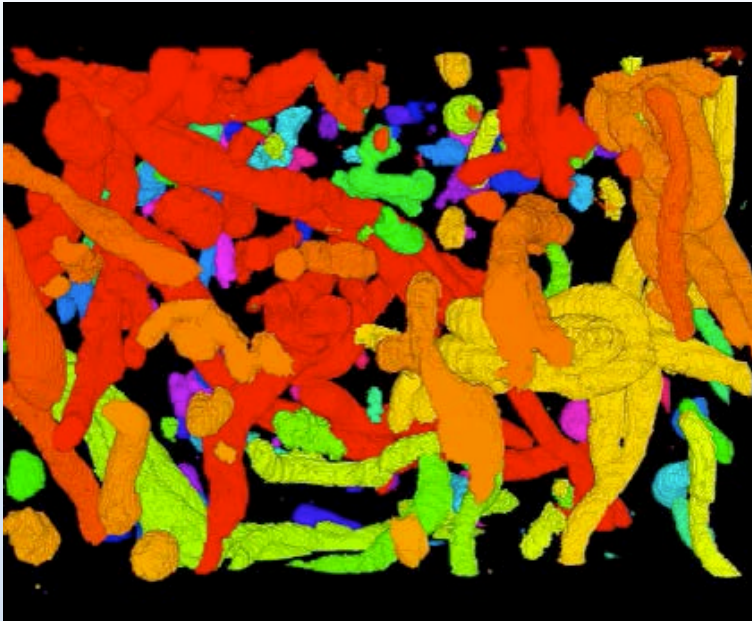
Dataset	Voxel Size	Train		Test	
		Size (voxels)	Labeled synapses	Size (voxels)	Labeled synapses
(A) Som. cortex	6.8 nm	$750 \times 564 \times 750$	9 (some)	$655 \times 429 \times 250$	28 (all)
(B) Hippocampus	5 nm	$1024 \times 653 \times 165$	20 (all)	$1024 \times 883 \times 165$ $1024 \times 1536 \times 200$	79 (all)
(C) Cerebellum	5 nm	$1398 \times 1343 \times 299$	7 (some)	$1966 \times 1343 \times 200$	56 (all)

Membranes and Contextual Features



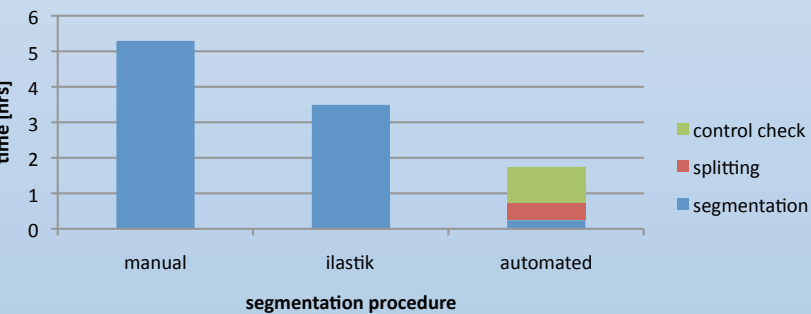
- Explicitly model membranes as separate regions and exploit the fact that the inside is enclosed within them to retain the graph cut formulation.
- Use the contextual features to compute the unary terms.
- ➔ Improved performance and faster training.

Time and Accuracy

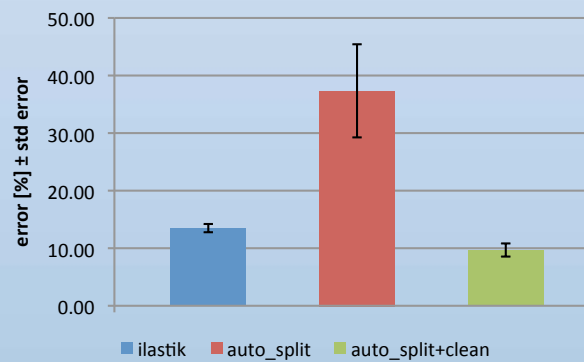


3.21 μm \times 3.21 μm \times 1.08 μm : 53 mitochondria

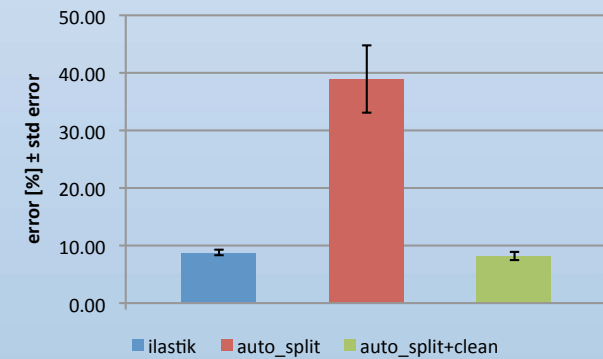
Mitochondria detection



Volume average error



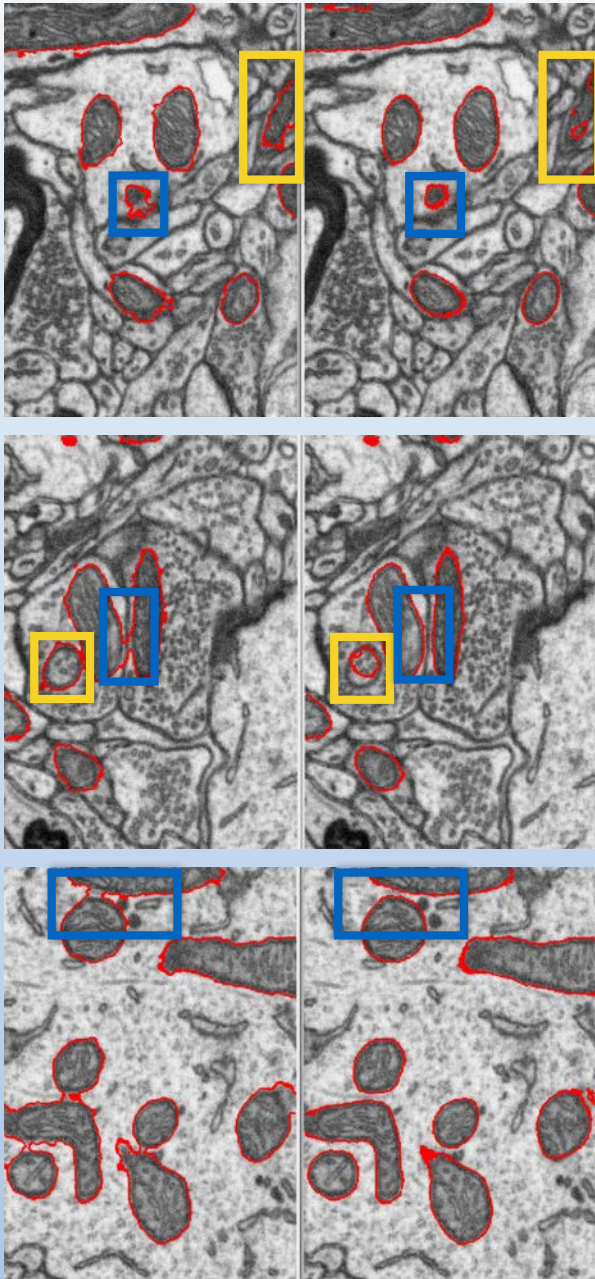
Surface average error



From Contextual Features to U-Net

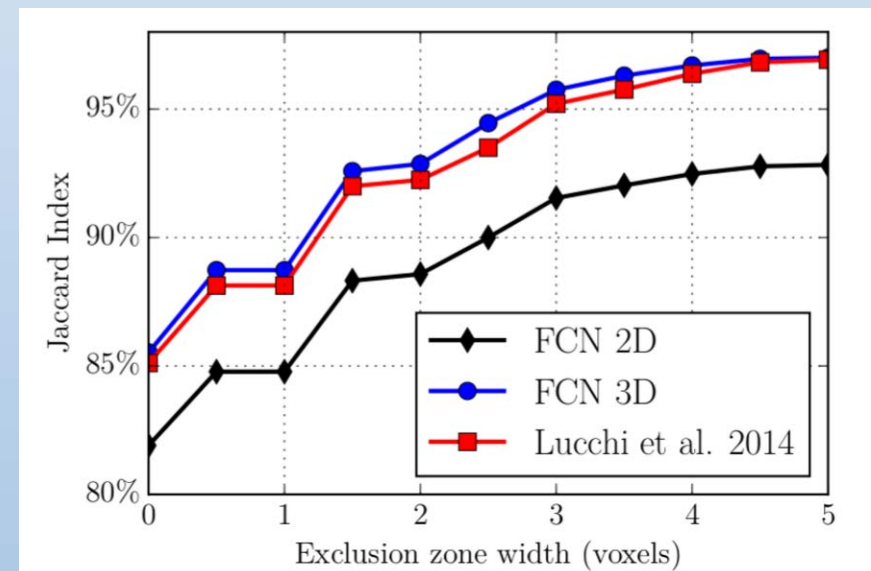
Context Features + CRF

U-Net 3D



Striatum Mitochondria

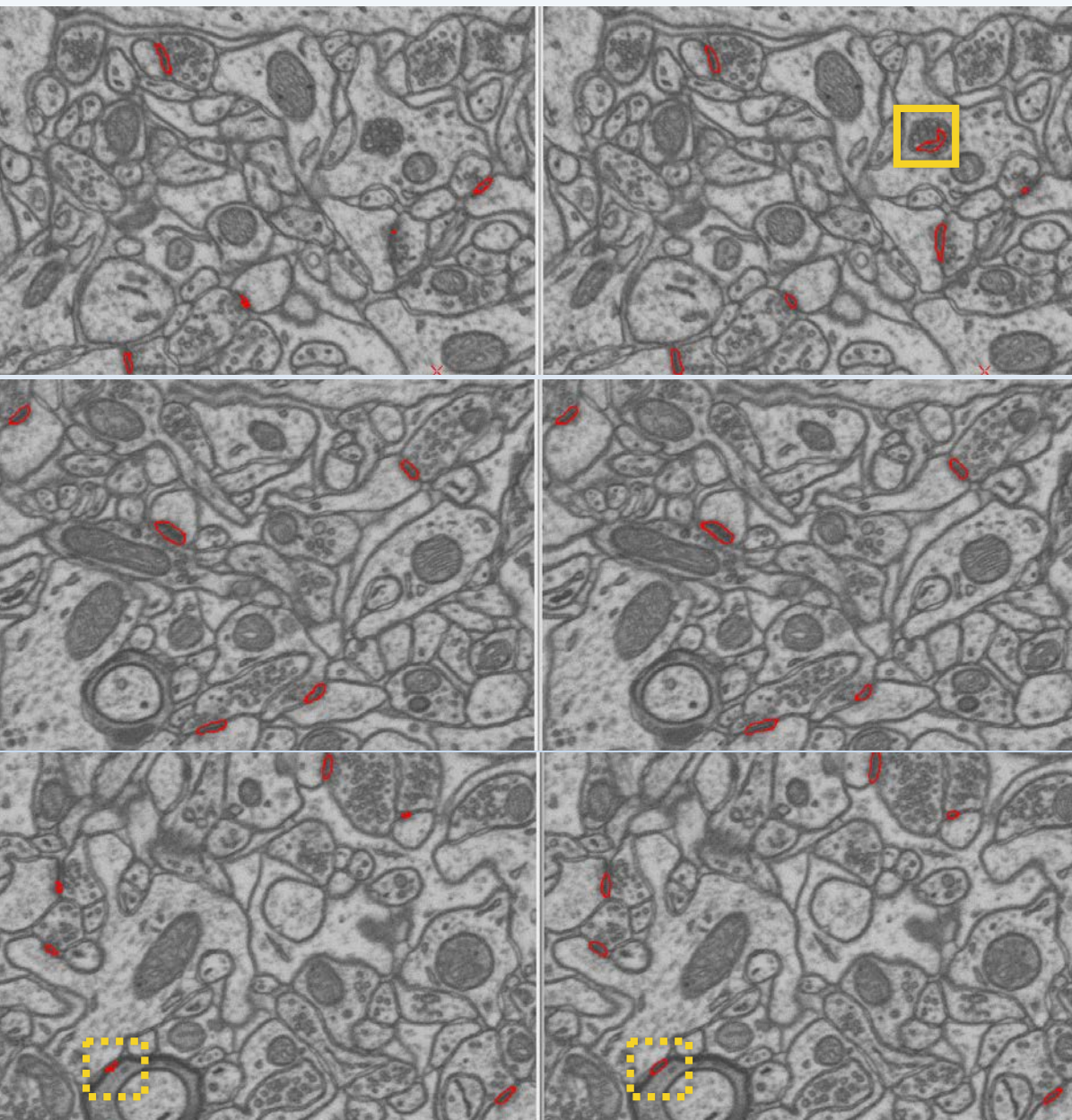
Method	Jaccard Index
Context F. + CRF	84.6%
U-Net 2D	82.4%
U-Net 3D	86.1%



Is Resistance Futile?

Context Features 3D CRF

U-Net 3D



Method	Jaccard Index
Context Features 2D	66.8%
Context Features 3D	85.2%
U-Net 2D	73.5%
U-Net 3D	77.0%

.. but U-Net does not take orientation into account.



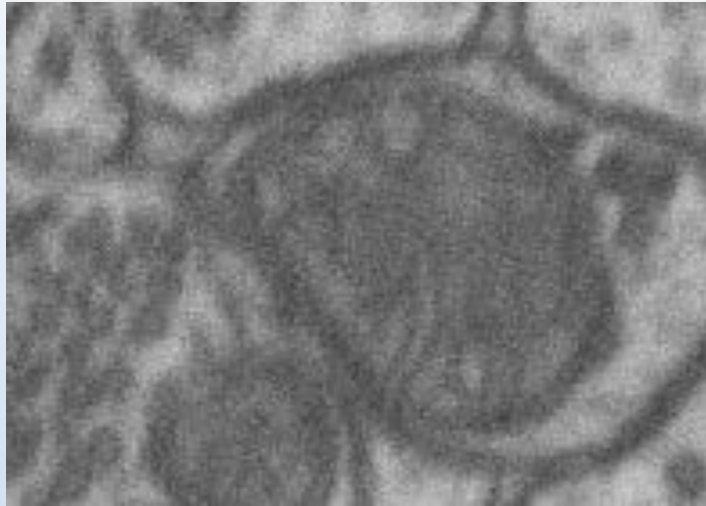
VS



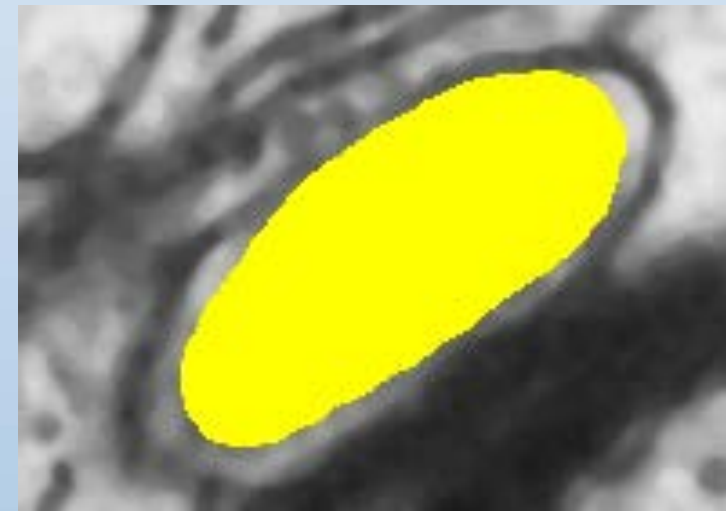
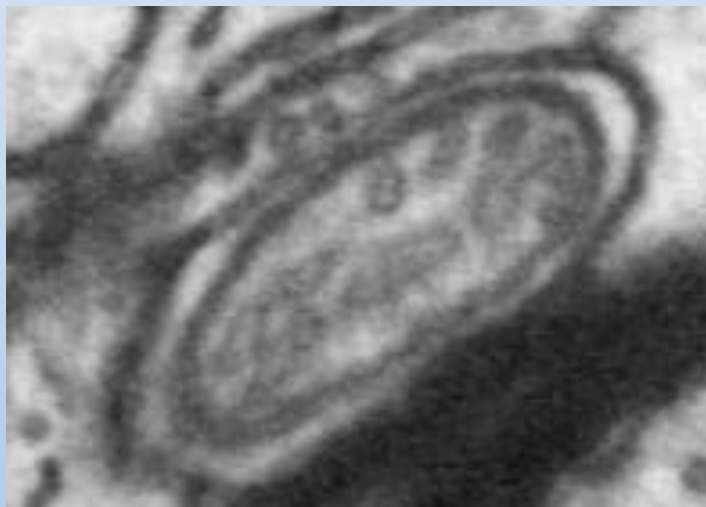
So, probably, but high-level knowledge still has its place.

A Closer Look at Mitochondria

Hippocampus



Striatum

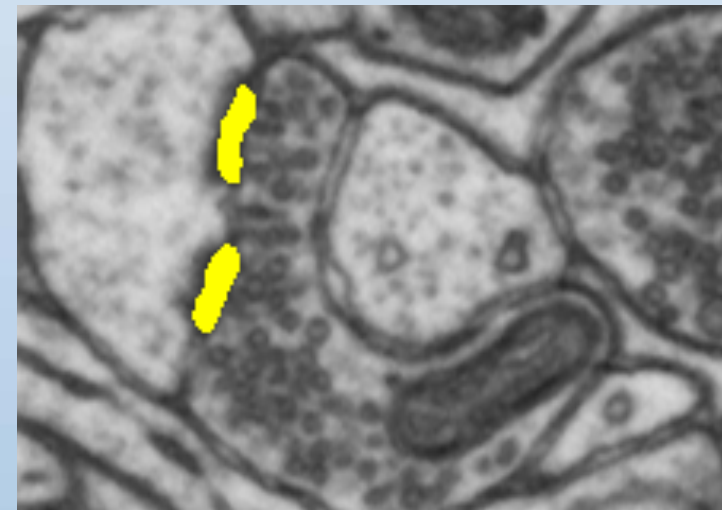
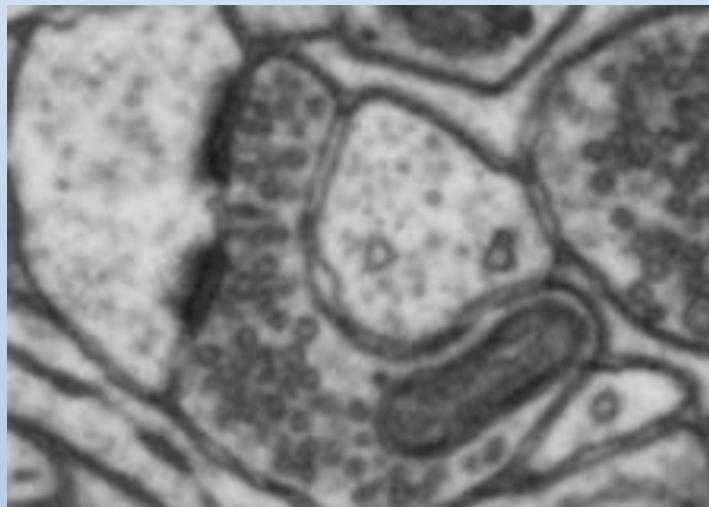


A Closer Look at Synapses

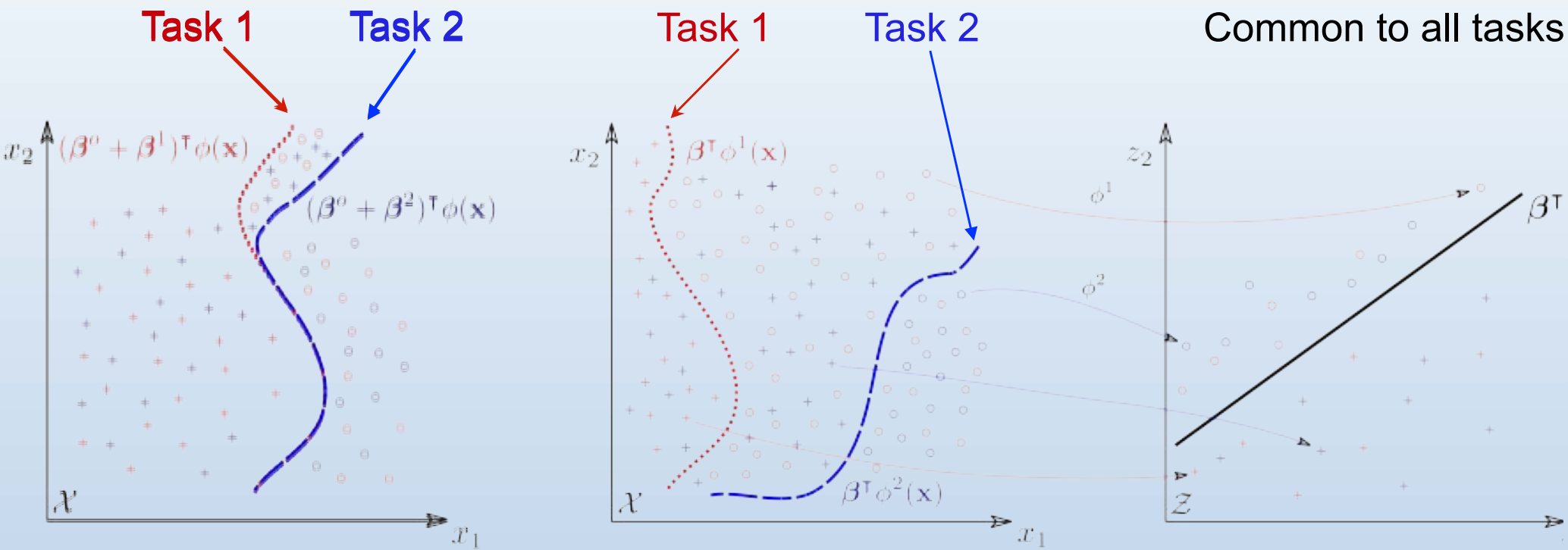
Cerebellum



Somatosensory
Cortex



Supervised Domain Adaptation



Standard
Multi-Task Learning

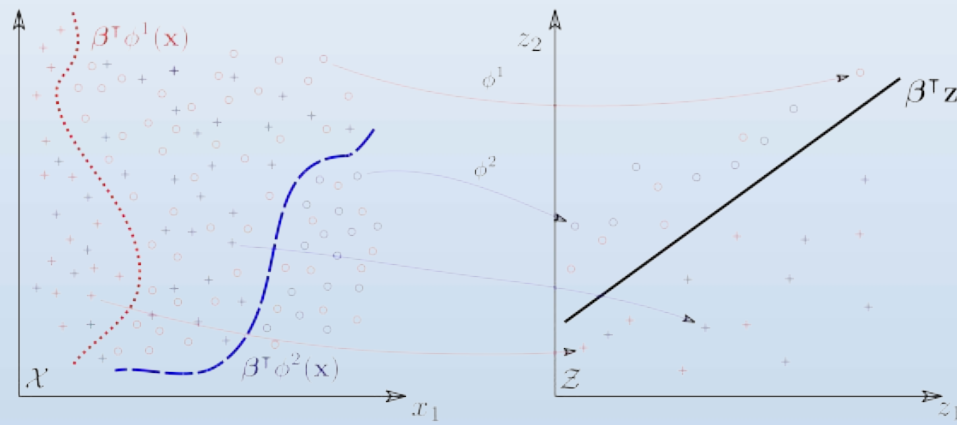
Our
Approach

Our Approach

Training samples $X^t = \{\mathbf{x}_i^t, y_i^t\}_{i=1}^{N^t}$, $t = 1, \dots, T$

Goal: Learn a mapping $\phi^t(\mathbf{x}^t) : \mathcal{X}^t \rightarrow \mathcal{Z}$

such that $f_t(\mathbf{x}) = \boldsymbol{\beta}^T \phi^t(\mathbf{x})$



How: Set of task-specific non-linear functions

$$\phi^t(\mathbf{x}^t) = [h_1^t(\mathbf{x}^t), \dots, h_M^t(\mathbf{x}^t)]^T$$

We propose $h_j^t(\mathbf{x}^t) = h_j(\mathbf{x}^t - \boldsymbol{\tau}_j^t)$, $h_j(\cdot) \in \mathcal{H}$

$$f_t(\mathbf{x}) = \boldsymbol{\beta}^T \phi^t(\mathbf{x}^t) = \sum_{j=1}^M \beta_j h_j(\mathbf{x}^t - \boldsymbol{\tau}_j^t)$$

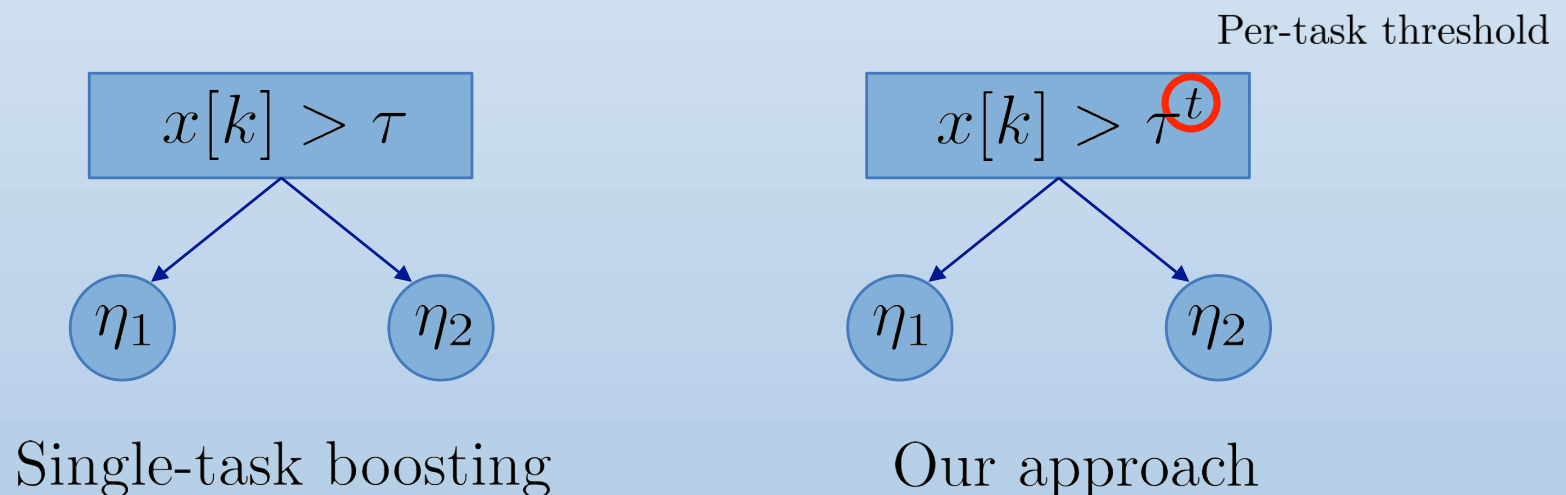
Gradient Boosting

Weak Learner Search

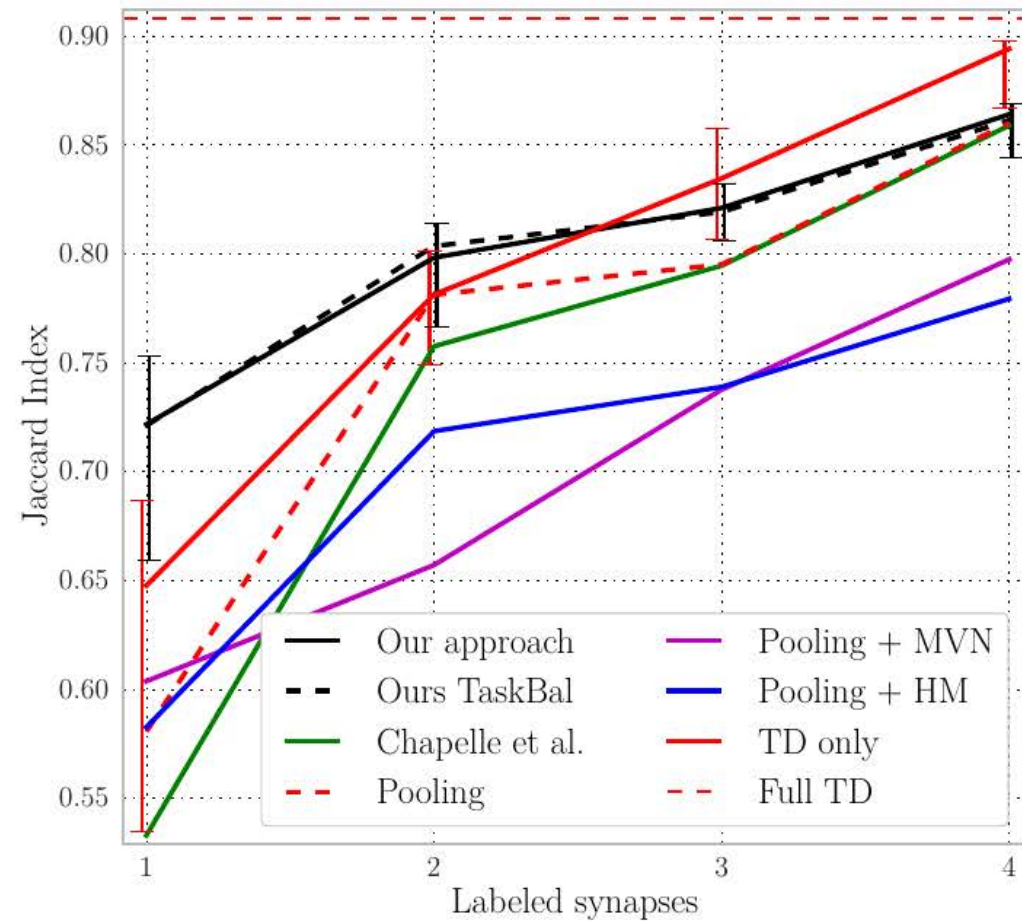
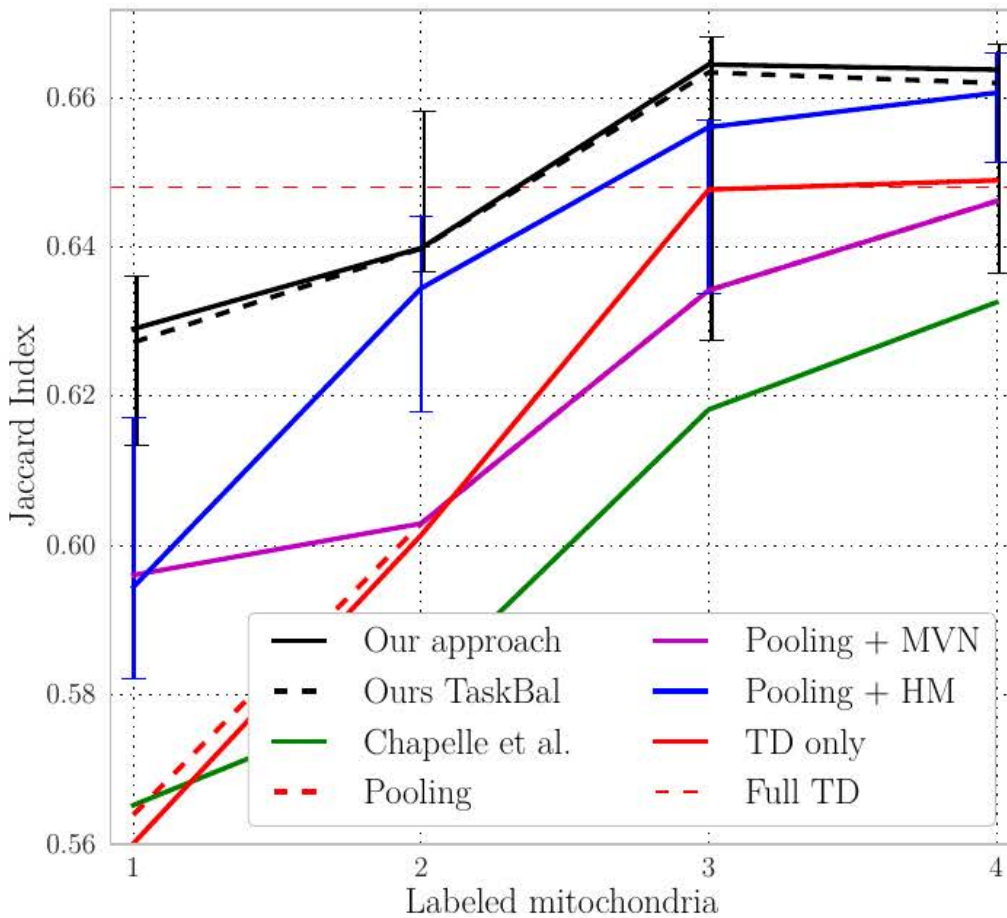
At each boosting iteration, the optimal weak learner \tilde{h} minimizes

$$\sum_{t=1}^T \left(\sum_{i=1}^{N^t} w_{ik}^t \left[\tilde{h}(\mathbf{x}^t - \tilde{\boldsymbol{\tau}}^t) - y_i^t \right]^2 \right)$$

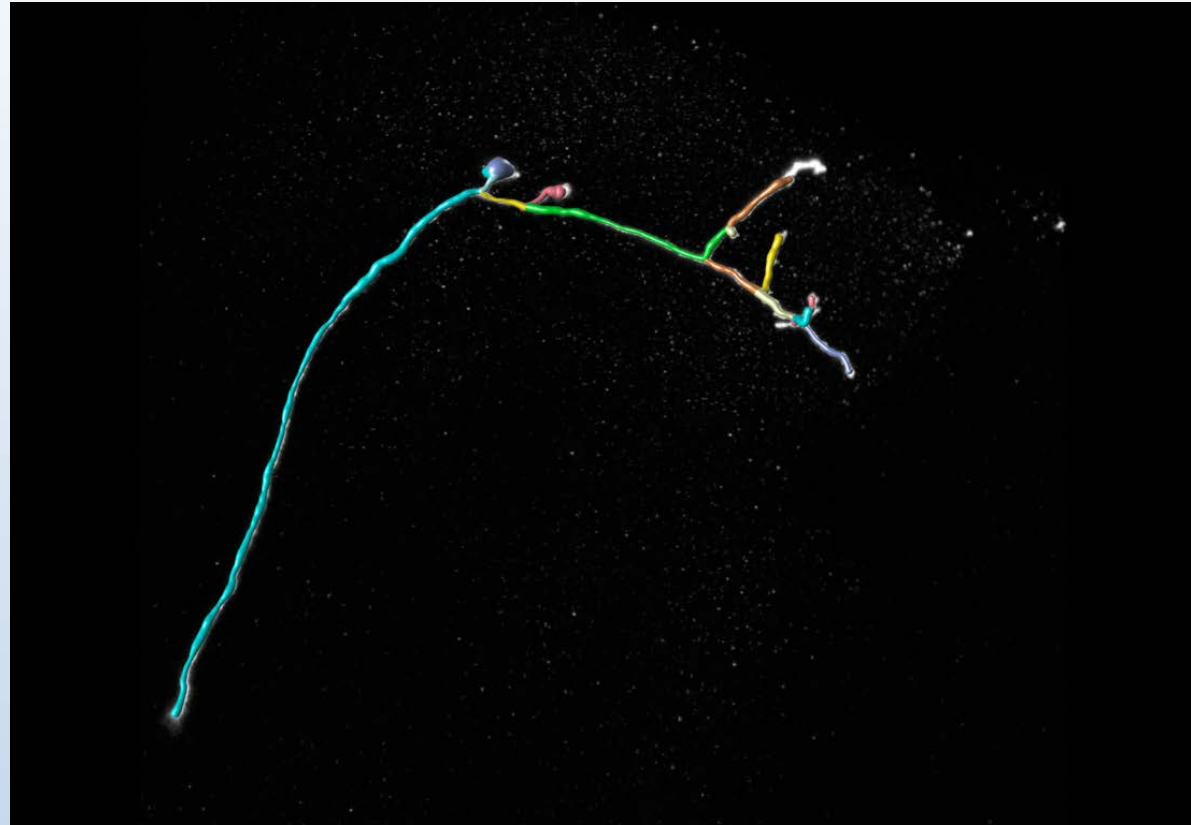
In practice, $h(\cdot)$ are **regression trees** or **decision stumps**



Quantitative Evaluation



Delineating Dendritic Trees ...



Processing steps:

1. Compute the *tubularity* of individual voxels.
2. Select voxels that locally maximize this measure.
3. Build a *tubular* graph by linking these voxels.
4. Find an optimal tree within the graph.

... and Roads

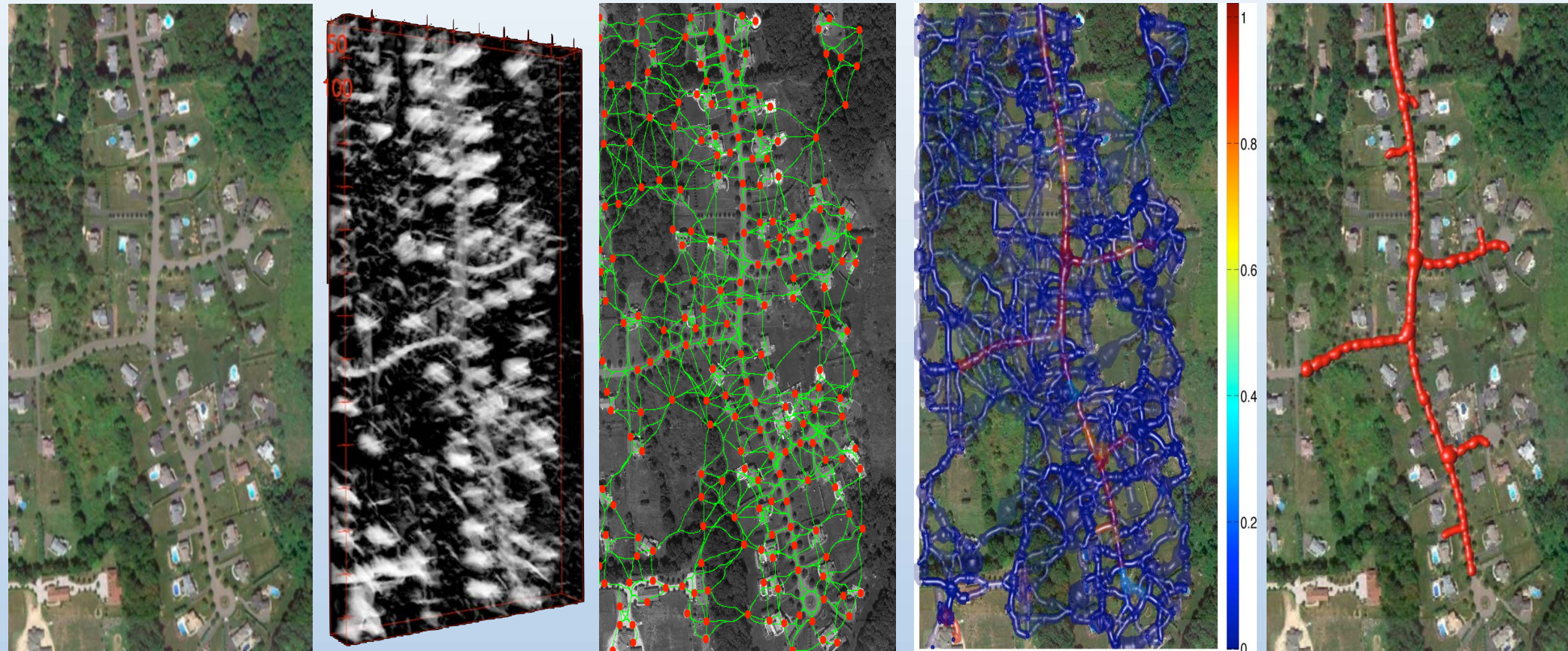
Image

Tubularity Volume

Graph

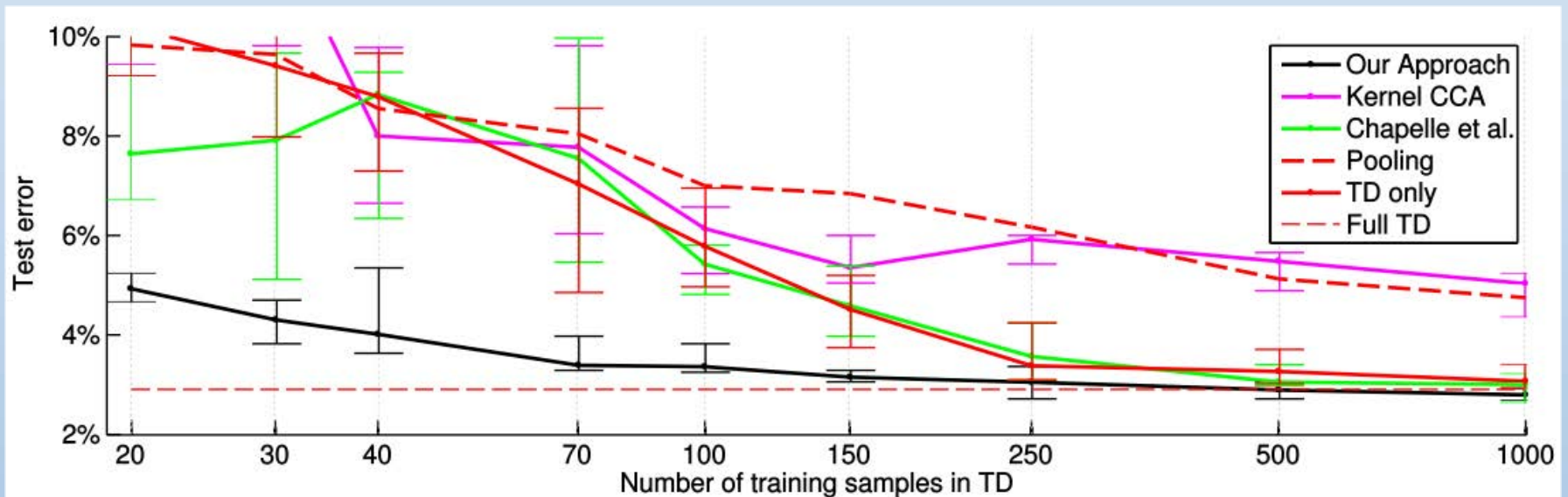
Weighted Graph

Reconstruction



—> The only difference between the two versions comes from the data used to train the path classifier!

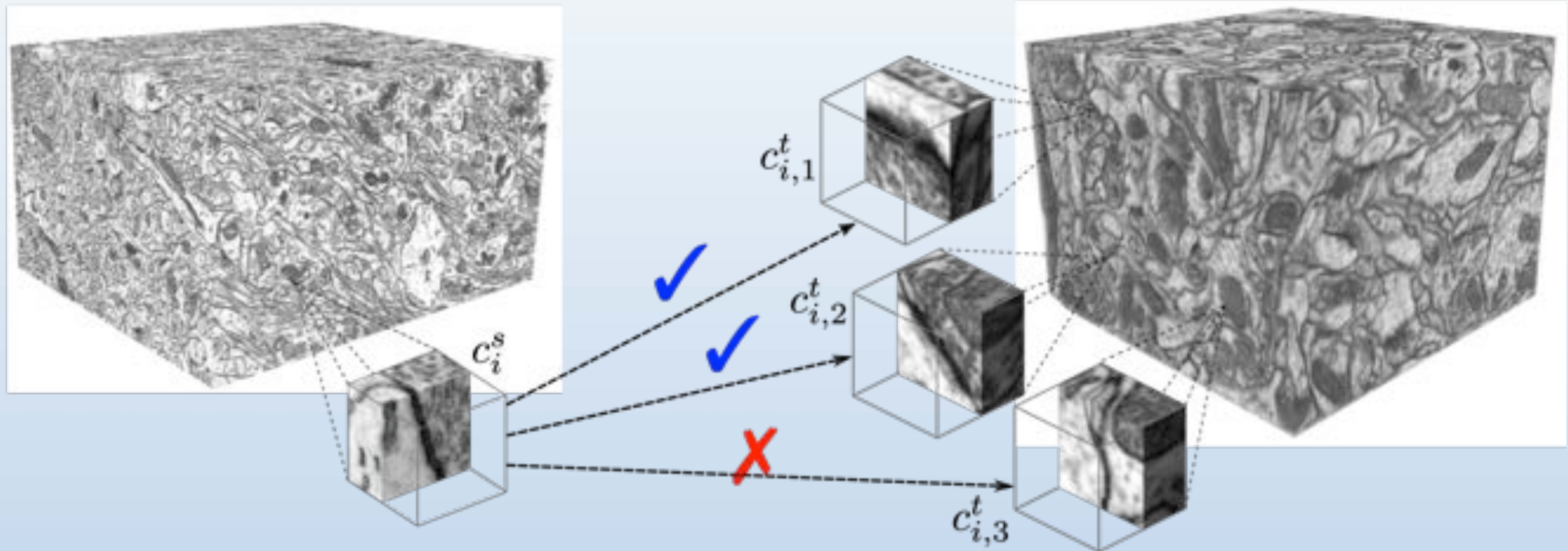
Path Classification



Unsupervised Domain Adaptation

Source Domain

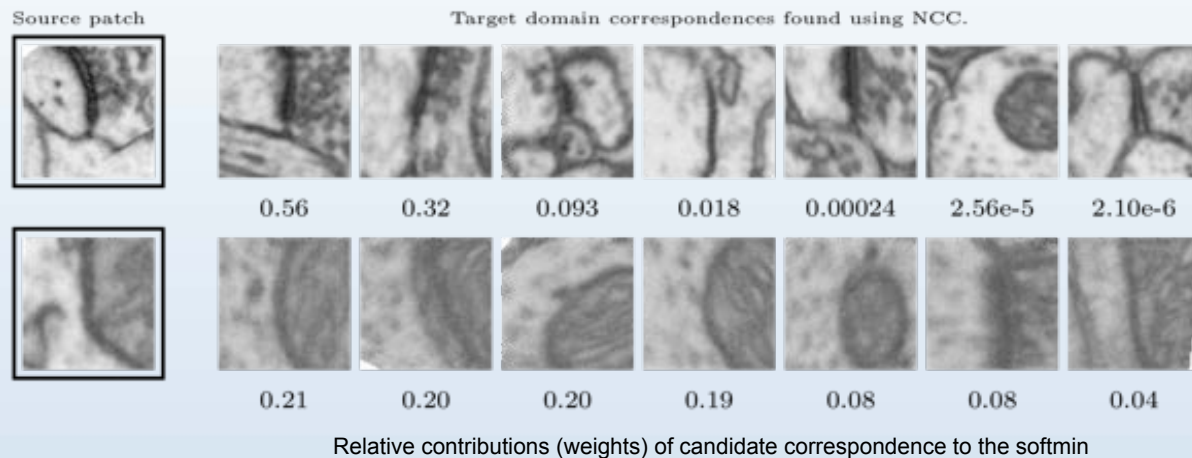
Target Domain



Given annotations in a source stack:

- Establish correspondences in the target stack, some correct, other not.
- Formulate a multiple instance learning problem.

Multiple Instance Learning



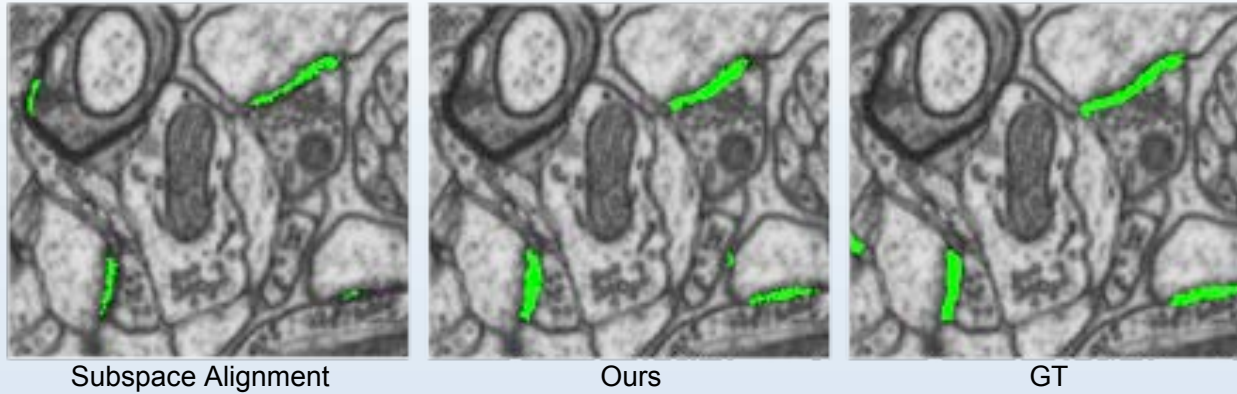
$$\hat{\Gamma}^t = \mathit{arg\,min}_{\Gamma^t} \frac{1}{|\mathcal{C}^s|} \sum_{c_i^s \in \mathcal{C}^s} \mathit{softmin} [\ell_{i1}, \ell_{i2}, \dots, \ell_{ik}],$$

where

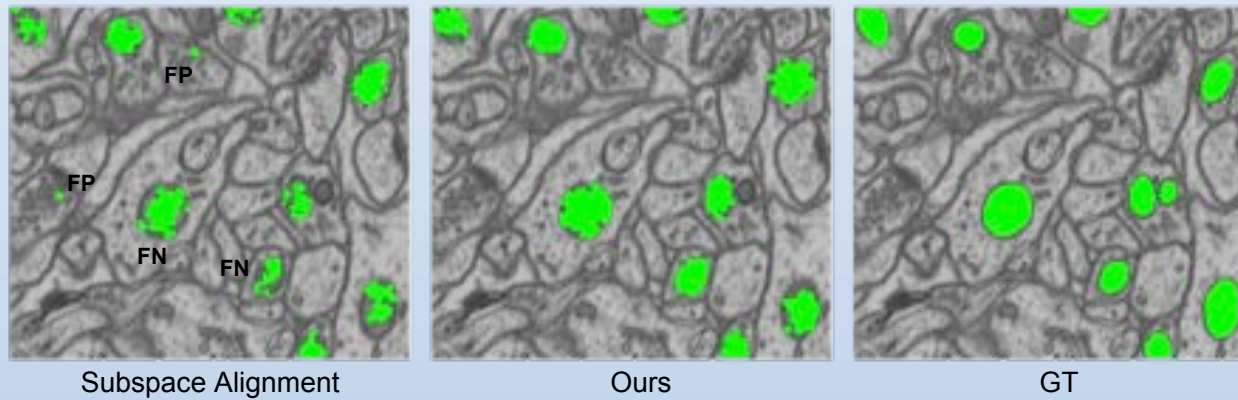
- $\ell_{ij} = L_\delta (f_{\Gamma^s}(c_i^s) - f_{\Gamma^t}(c_{i,j}^t))$
- Γ^s and Γ^t are model parameters, that is boosted tree thresholds.
- L_δ is the Huber loss
- $\mathit{softmin} [\ell_1, \dots, \ell_k] = -\frac{1}{r} \ln \frac{1}{k} \sum_{j=1}^k \exp(-r\ell_j)$

Segmentation

Synapses

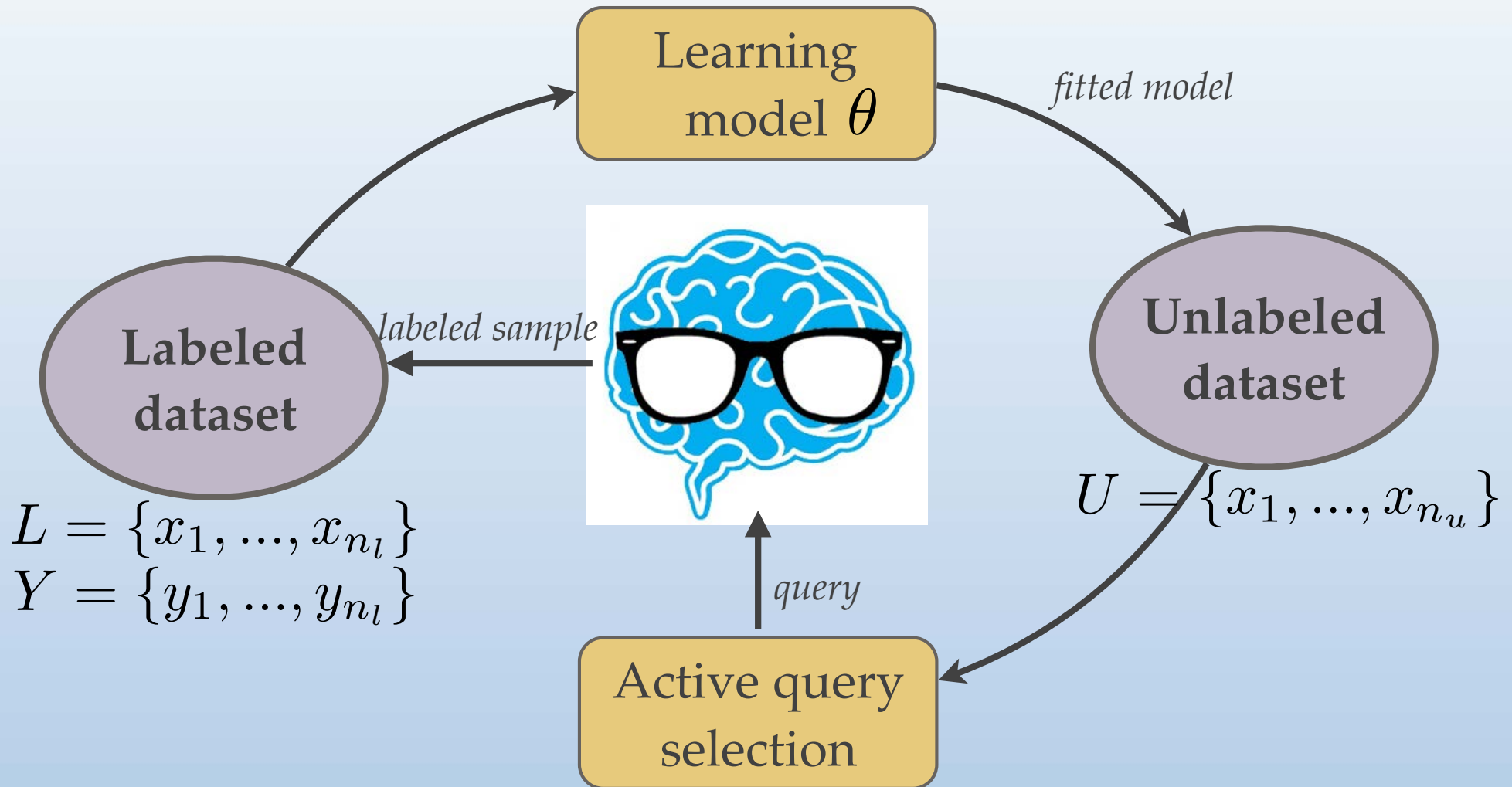


Mitochondria

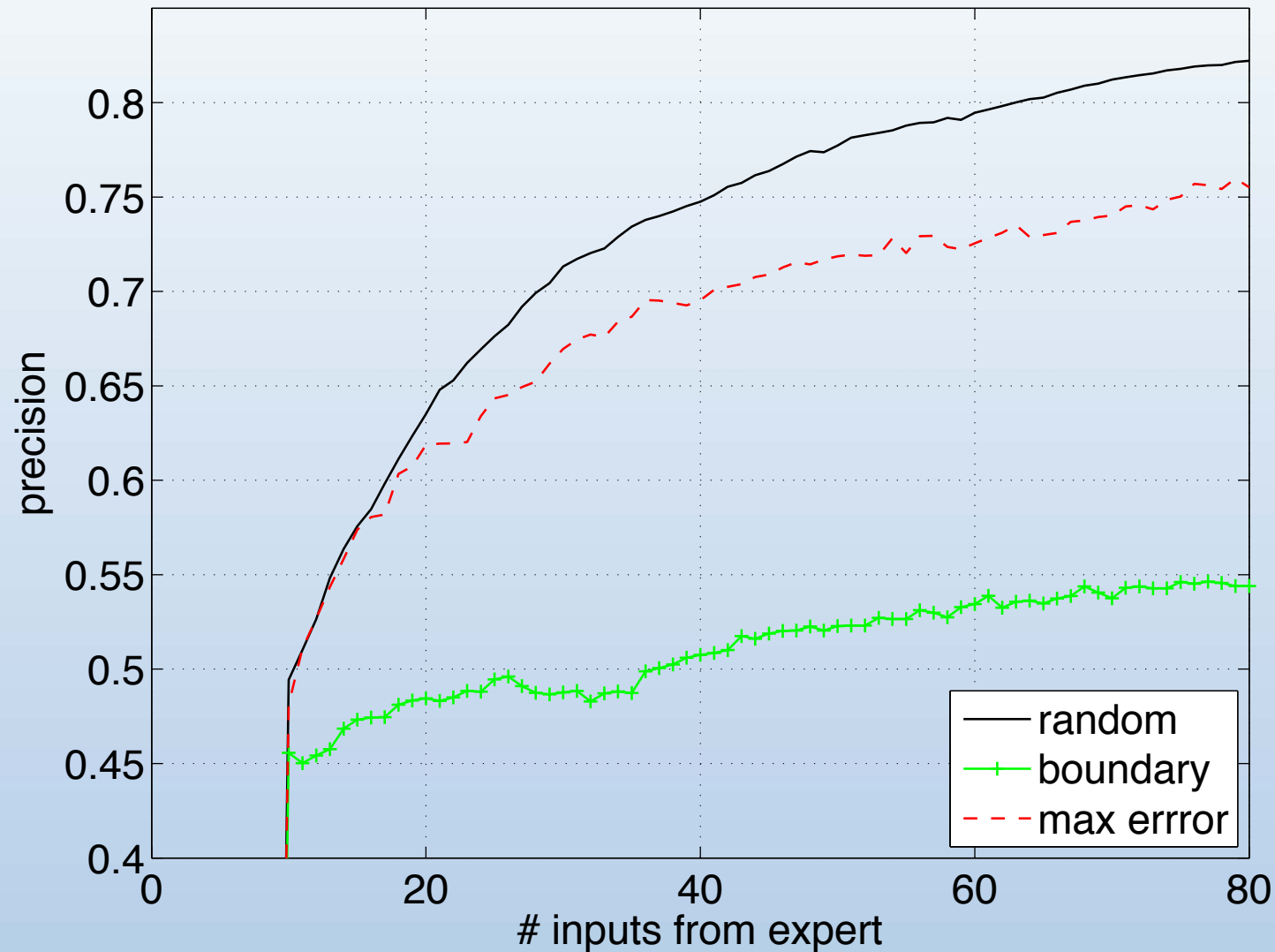


	No adaptation	Histogram Matching	TD only	SA [7] + Lin. SVM	SA [7] + AdaBoost	Ours ($k = 8$)
Synapse	0.22	0.32	0.39	0.13	0.39	0.57
Mitochondria	0.50	0.39	0.57	0.24	0.59	0.62

Active Learning



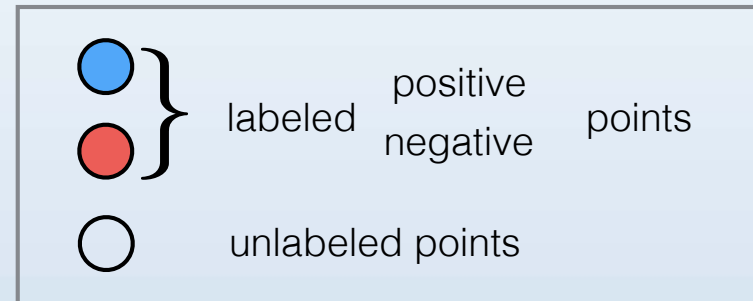
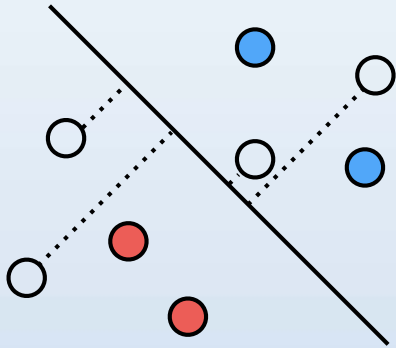
Obvious Strategies



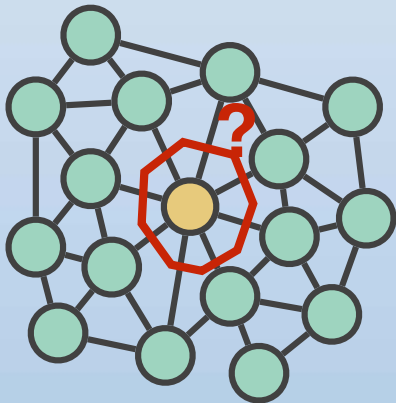
... are not necessarily best.

Uncertainty Sampling

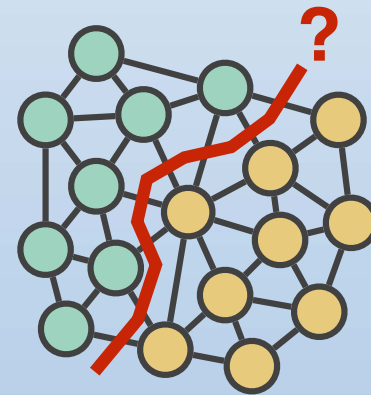
Conventional uncertainty sampling:



Introducing geometry:

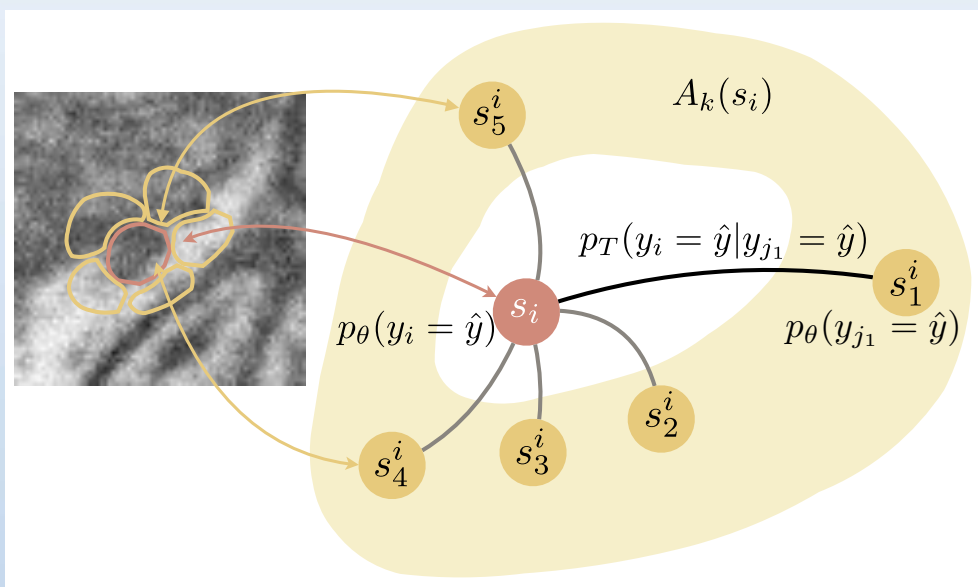


Potentially mislabeled points.



Favoring smooth boundaries.

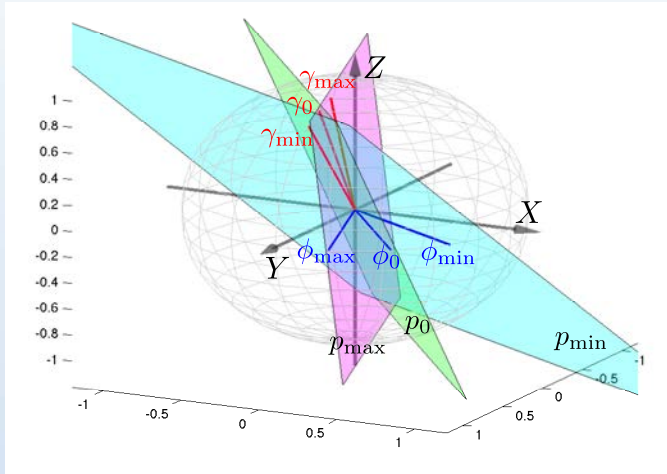
Geometry-Based Active Learning



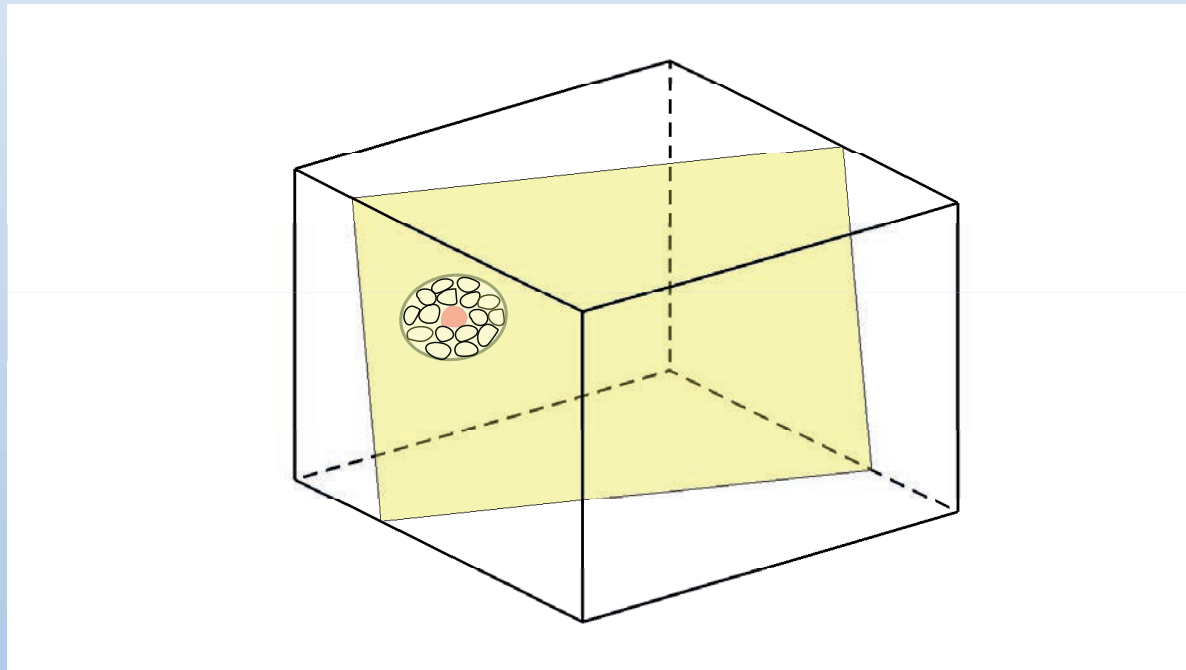
$$p_G^0(y_i = \hat{y}) = \sum_{s_j \in A_k(s_i)} p_T(y_i = \hat{y} | y_j = \hat{y}) p_\theta(y_j = \hat{y} | \mathbf{x}_j)$$

$$p_G^{\tau+1}(y_i = \hat{y}) = \sum_{s_j \in A_k(s_i)} p_T(y_i = \hat{y} | y_j = \hat{y}) p_G^\tau(y_j = \hat{y})$$

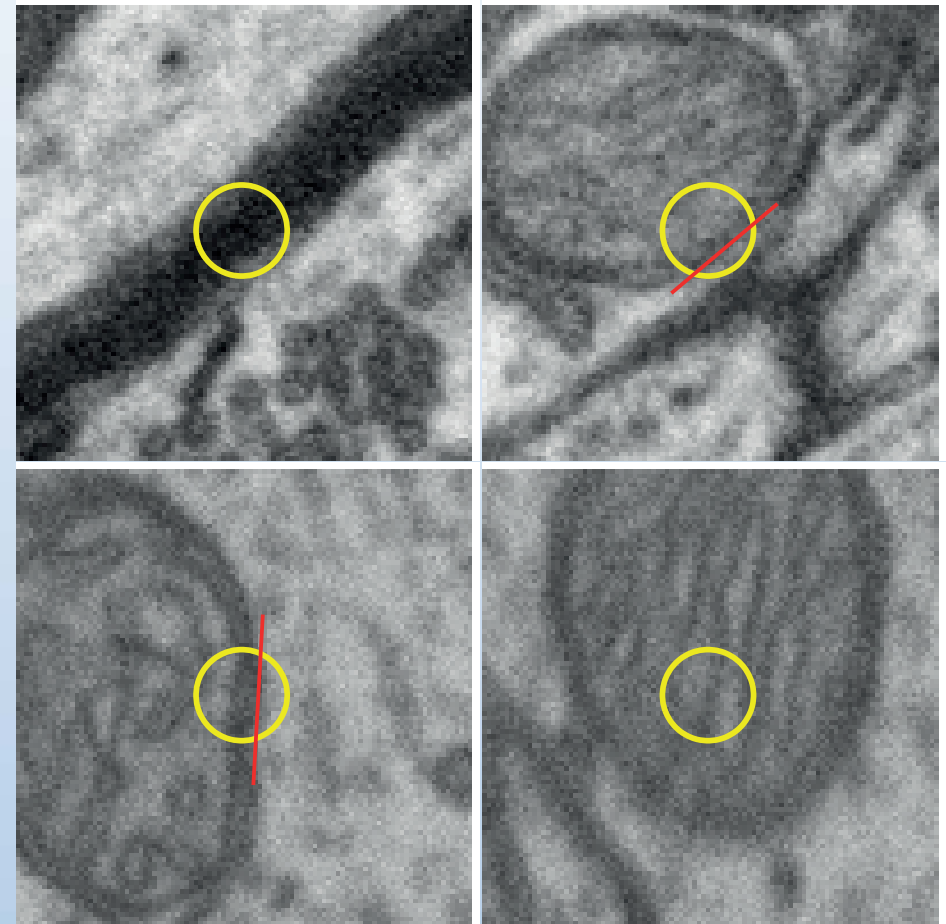
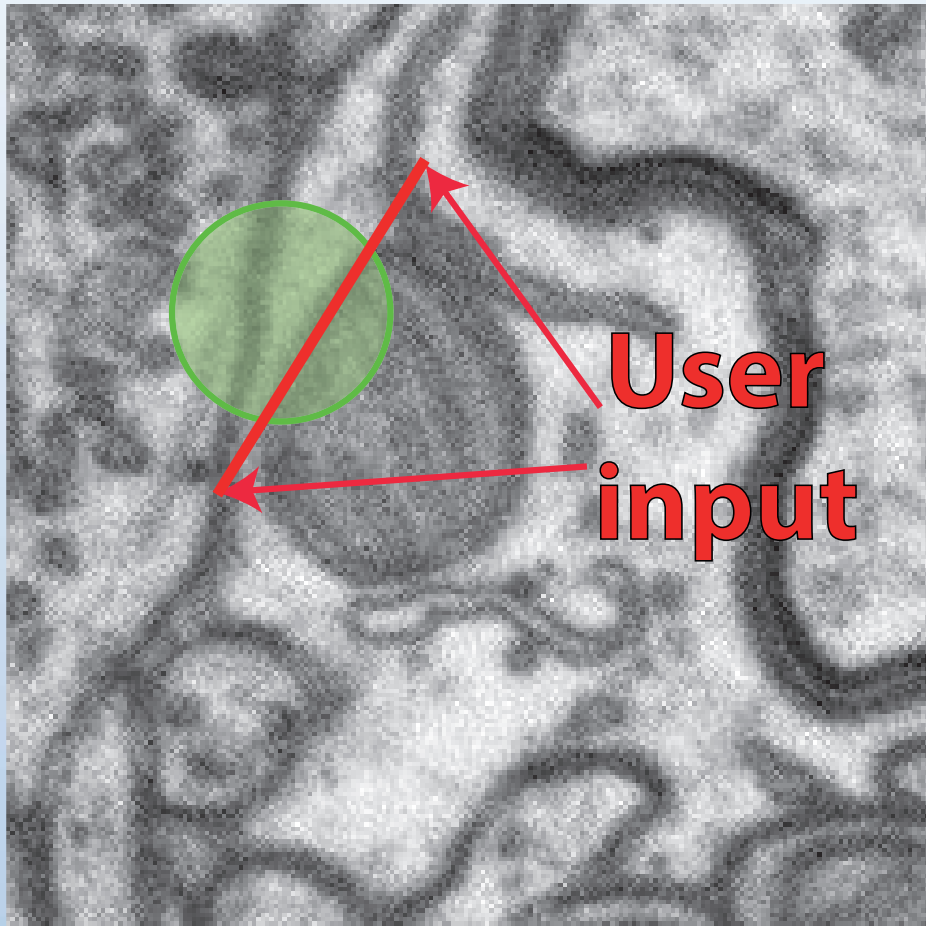
Batch-Mode Active Learning



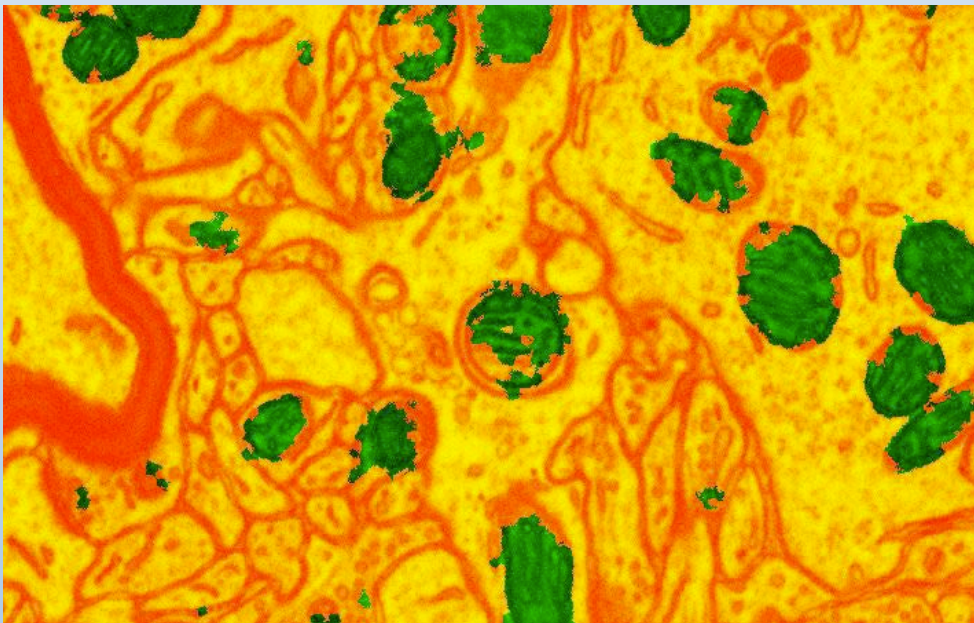
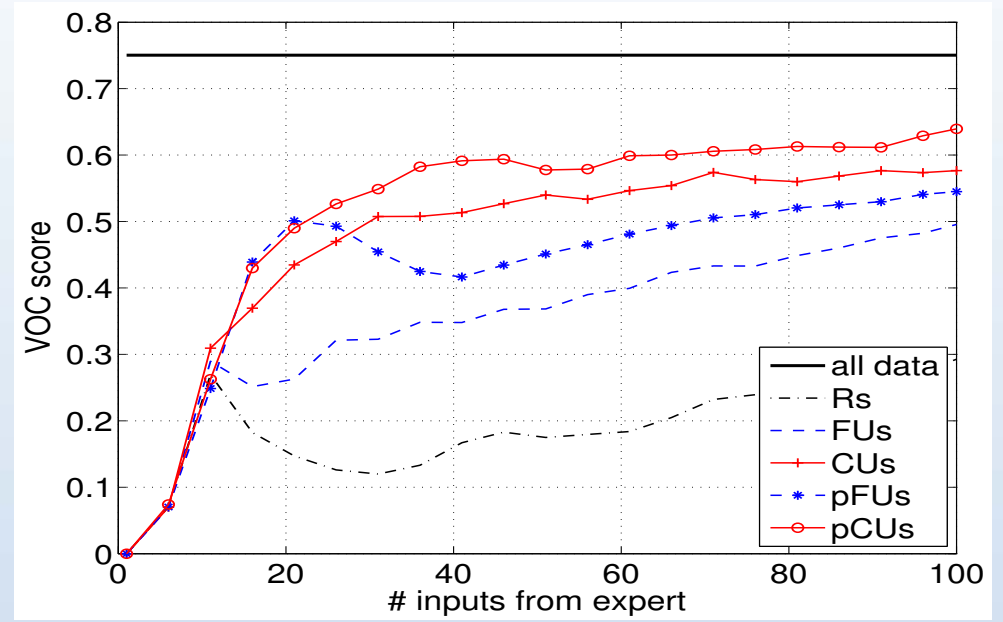
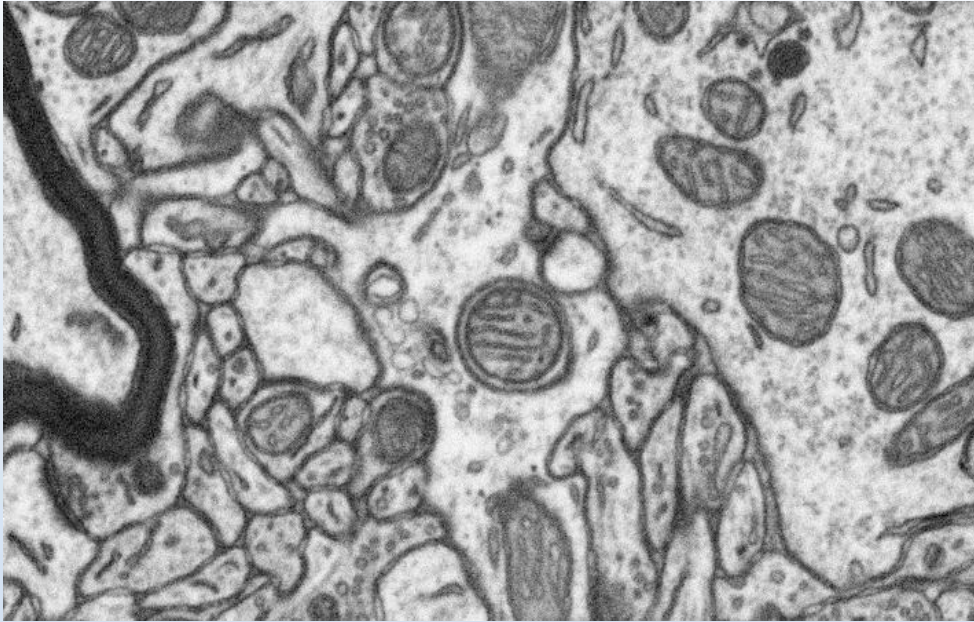
- Branch and Bound Search for an Optimal Slice.
- Much easier task for the expert annotator.



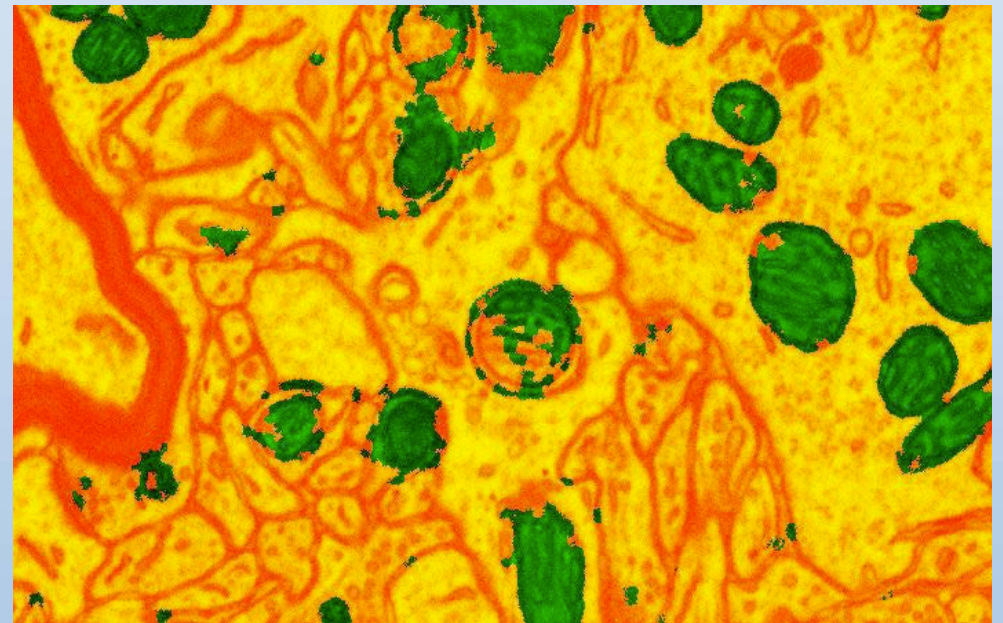
Batch-Mode Geometry-Based Active Learning



EM Segmentation Results

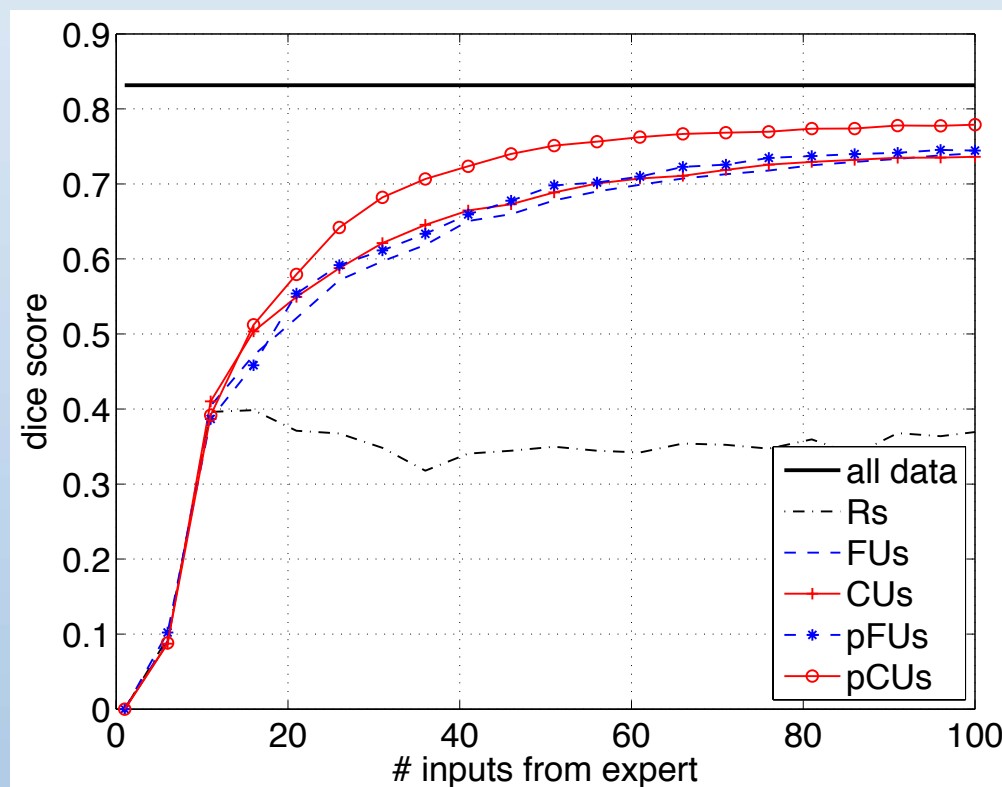
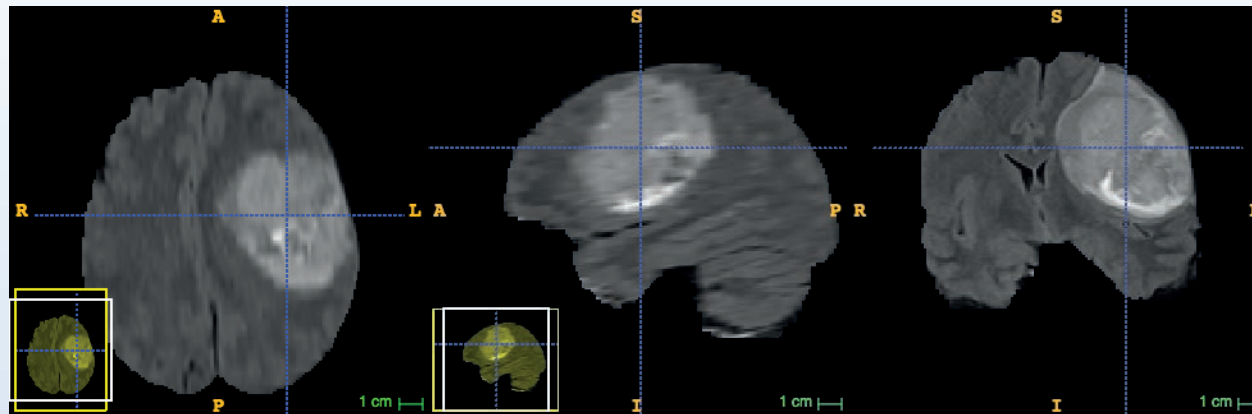


10000 datapoints

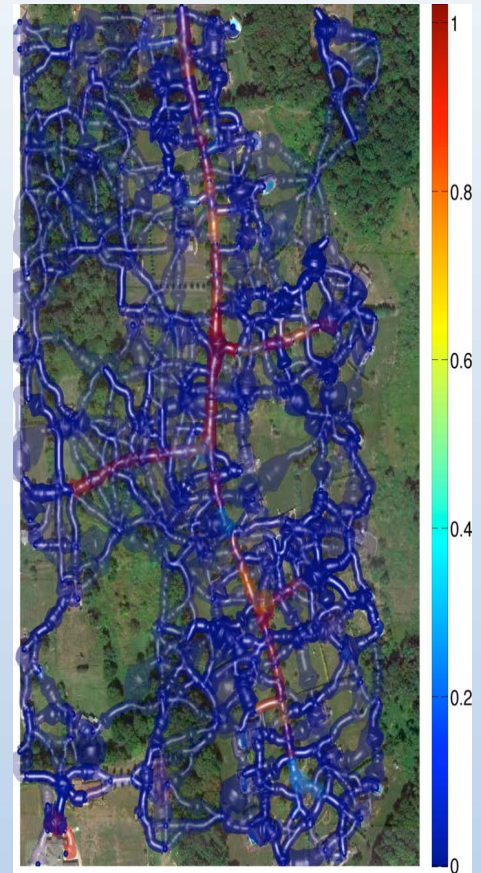
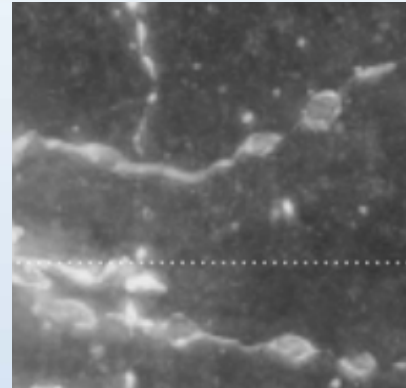
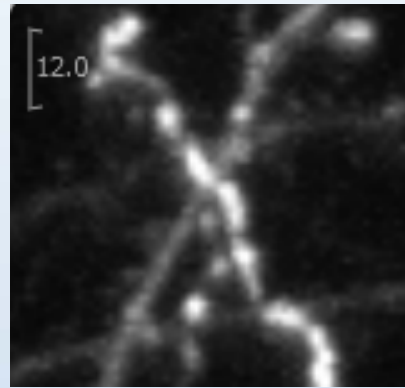


100 datapoints

MRI Segmentation Results



Delineation



Ambiguous cases:

- Branch crossovers
- Occluded paths
- Linear structures that are not part of the ground truth

—> Annotator should focus on paths that violate the smoothness of probabilities along the branches.

Probability Propagation

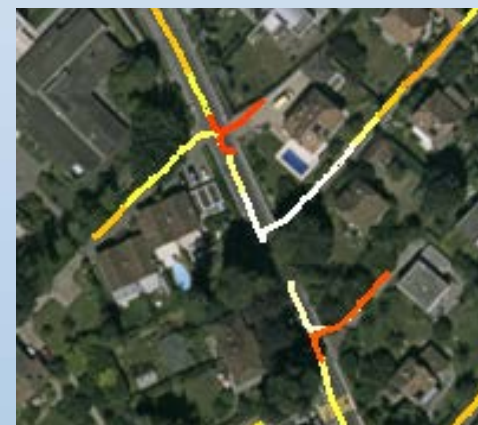
- Propagate the probabilities P_0 assigned by the base classifier around the overcomplete graph to obtain:

$$P^* = (I - \alpha S)^{-1} P_0 \quad \text{Zhou et al., NIPS 04}$$

- Calculate entropy *after* the propagation
- Identify ambiguous image regions



Uncertainty
before propagation



Uncertainty
after propagation



Density Query

◦ A selected batch E^* should be:

- *Informative* –entropy
- *Representative* –global similarity
- *Diverse* –labelled and intra-batch similarity

Introduce a *density score*:

- σ_G – global similarity
- σ_L – labelled similarity
- σ_I – intra-batch similarity

$$\sigma_G(E) = \sum_{i \in E} \sum_{1 \leq j \leq N} w_{ij}$$

$$\sigma_L(E) = \sum_{i \in E} \sum_{l \in L} w_{il}$$

$$\sigma_I(E) = \sum_{i \in E} \sum_{j \in E} w_{ij}$$

$$\mu(E) = \frac{\sigma_G(E) - \sigma_L(E) - \sigma_I(E)}{\sigma_G(E)},$$

Combine density and uncertainty components:

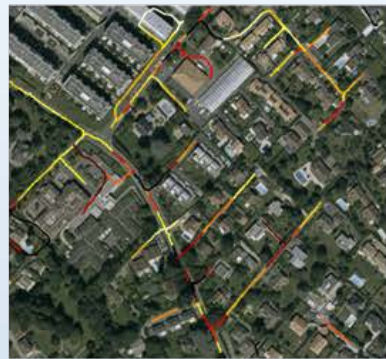
$$E^* = \arg \max_{E \in E_k} \mu(E) \left(\sum_{i \in E} H(\mathbf{x}_i) \right)$$

Algorithm's behavior

Decreasing ambiguity



5 samples



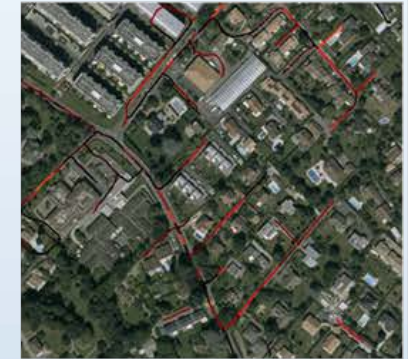
15 samples



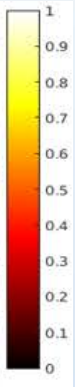
25 samples



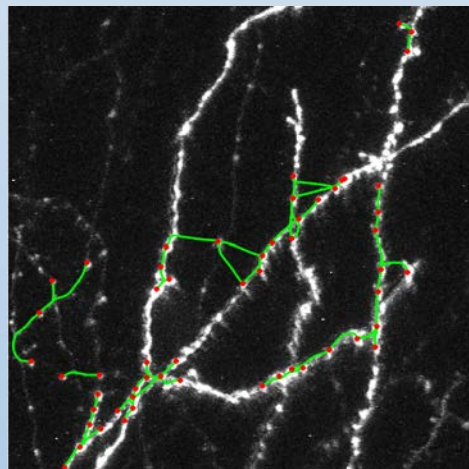
35 samples



45 samples



Frequently queried pairs



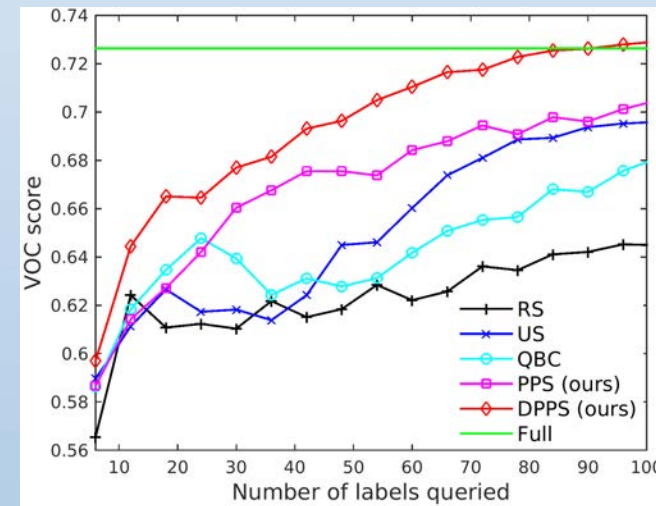
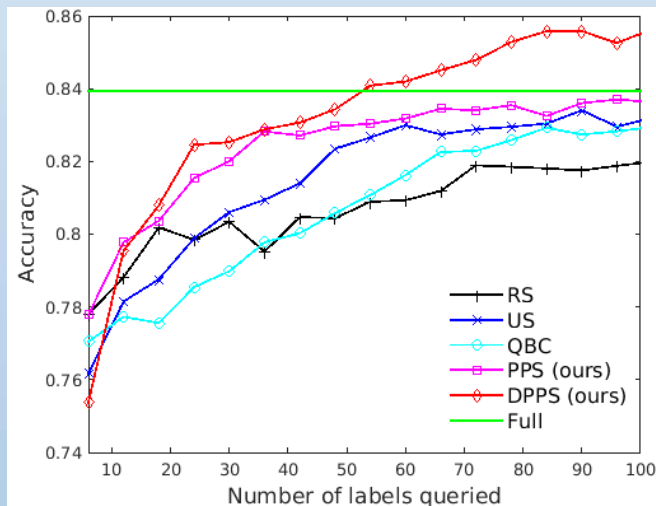
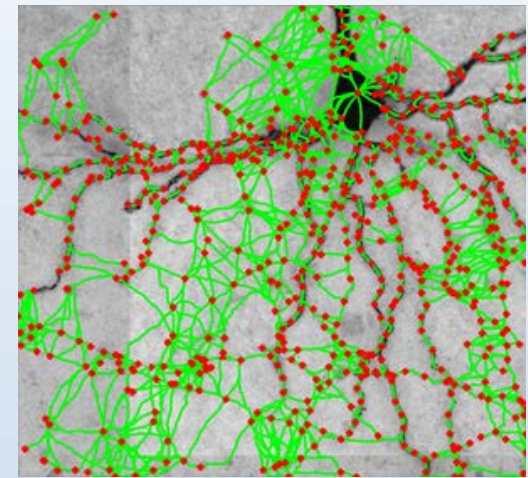
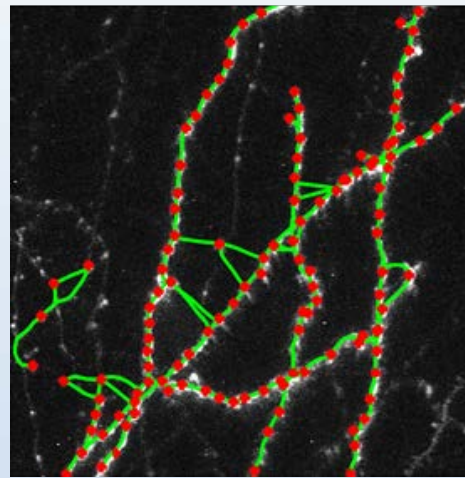
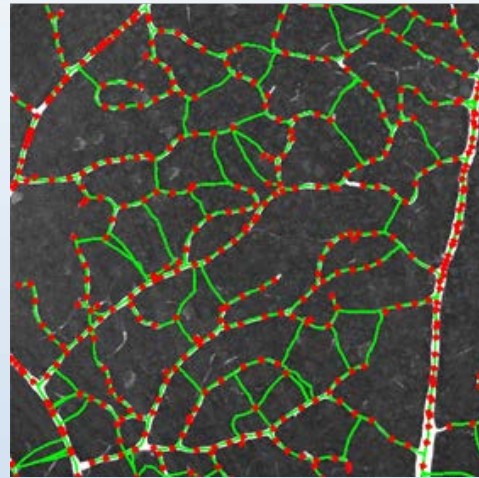
Path Classification Results

Roads

Blood vessels

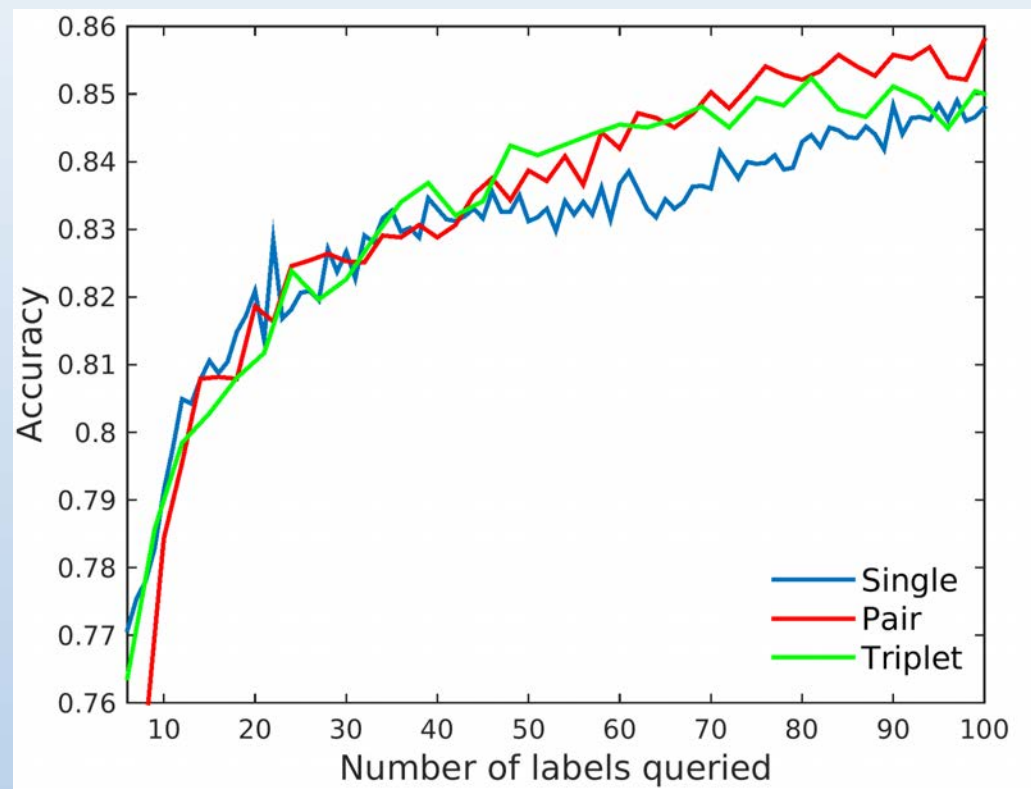
Axons

Neurons

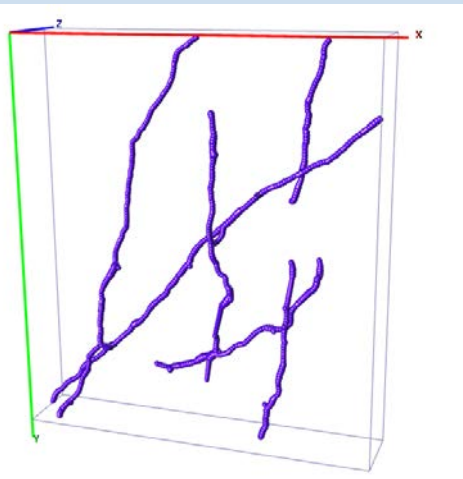
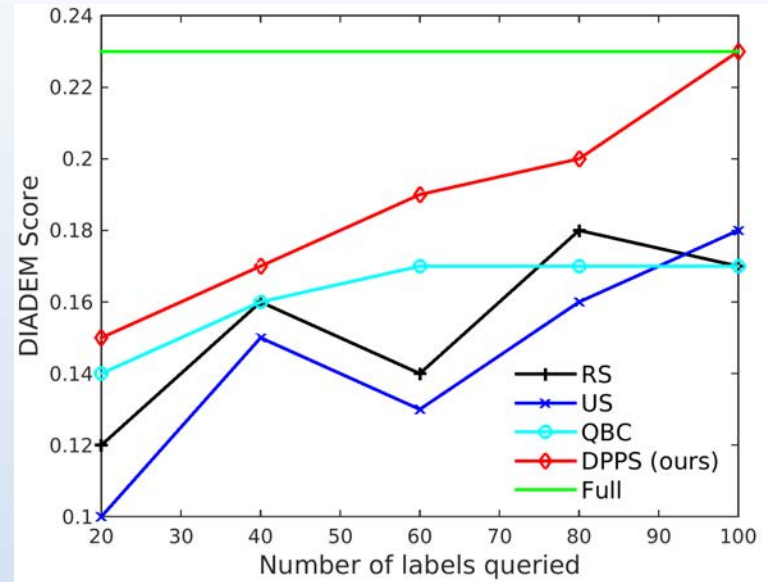


Batch size

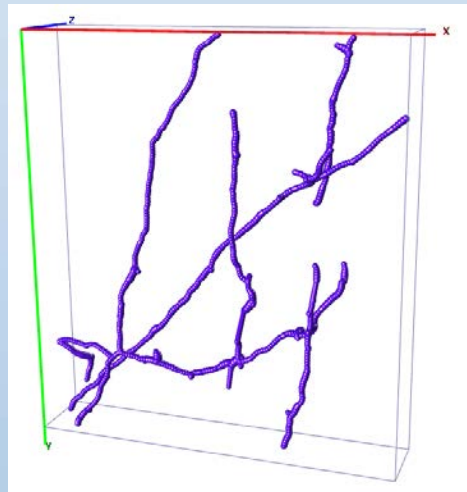
- *Singletons* – do not capture geometry
- *Pairs* – capture geometry and enable frequent classifier update
- *Triples* – classifier not updated frequently enough



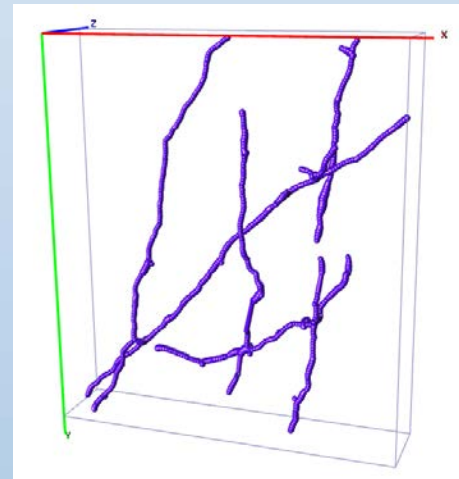
Delineation Results



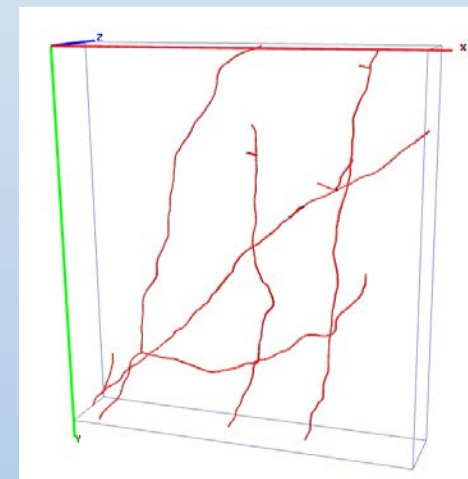
Random



Uncertainty

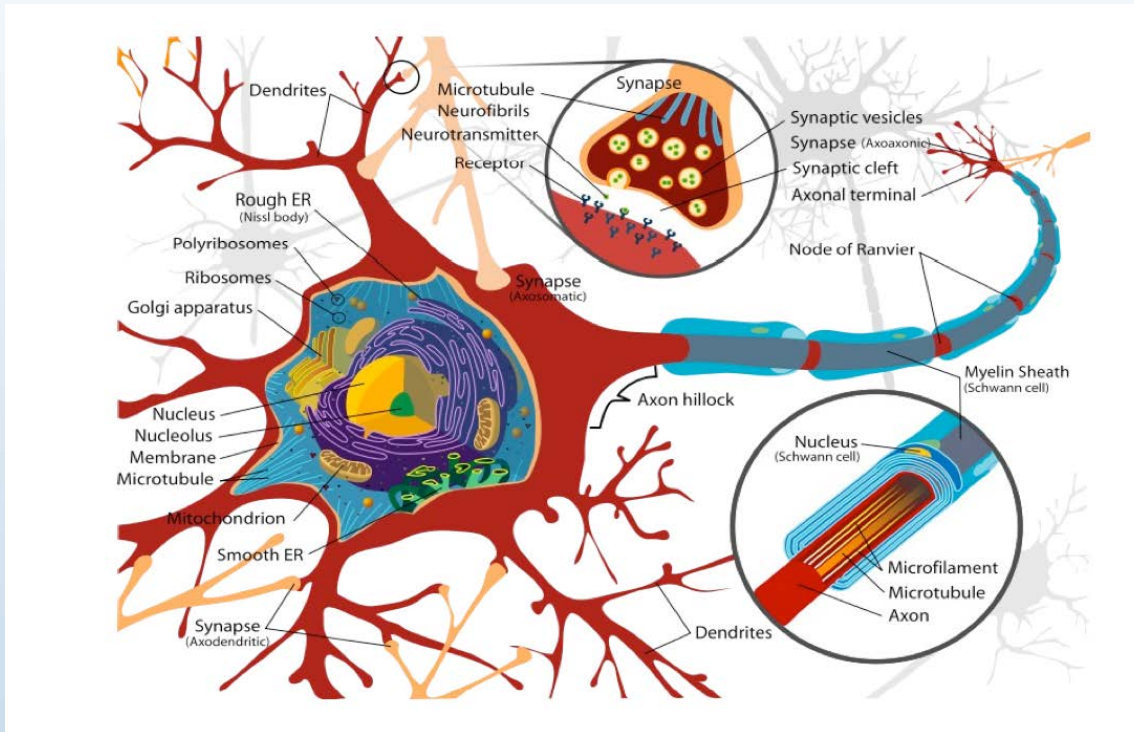


Ours



Ground truth

Conclusion



- A human brain contains approximately 100 billion neurons and 100 trillion synapses.
- It would take 1000 Exabytes to store an uncompressed digitization at 5nm resolution.

- Techniques that rely on Machine Learning are required but the training data is hard to obtain.
- Both Domain Adaptation and Active Learning can help but the interfaces must be carefully designed.
- We are going to have to learn how to learn.

Thanks To

- Radhakrishna Achanta
- Karim Ali
- Carlos Becker
- Roger Bermudez
- Mario Christoudias
- François Fleuret
- German Gonzales
- Anne Jorstad
- Graham Knott
- Ksenia Konyushkova
- Vincent Lepetit
- Yunpeng Li
- Aurelien Lucchi
- Pablo Marquez Neila
- Agata Mosinska
- Felix Schuermann
- Amos Sironi
- Raphael Sznitman
- Kevin Smith
- Engin Turetken



References

Segmentation in EM images

C. J. Becker, K. Ali, G. Knott and P. Fua. Learning Context Cues for Synapse Segmentation, IEEE Transactions on Medical Imaging, 2014.

A. Lucchi, C. J. Becker, P. Marquez Neila and P. Fua. Exploiting Enclosing Membranes and Contextual Cues for Mitochondria Segmentation. In Conference on Medical Image Computing and Computer Assisted Intervention, 2014.

Domain Adaptation

C. J. Becker, C. M. Christoudias and P. Fua. Domain Adaptation for Microscopy Imaging, IEEE Transactions on Medical Imaging, 2015.

R. Bermúdez-Chacón, C. Becker, M. Salzmann and P. Fua. Scalable Unsupervised Domain Adaptation for Electron Microscopy. In Conference on Medical Image Computing and Computer Assisted Intervention 2016.

Active Learning

K. Konyushkova, R. Sznitman and P. Fua. Introducing Geometry in Active Learning for Image Segmentation. in International conference in Computer Vision, Santiago, 2015.

A. Mosinska-Domanska, R. Sznitman, P. Glowacki and P. Fua. Active Learning for Delineation of Curvilinear Structures. in Conference on Computer Vision and Pattern Recognition, 2016.