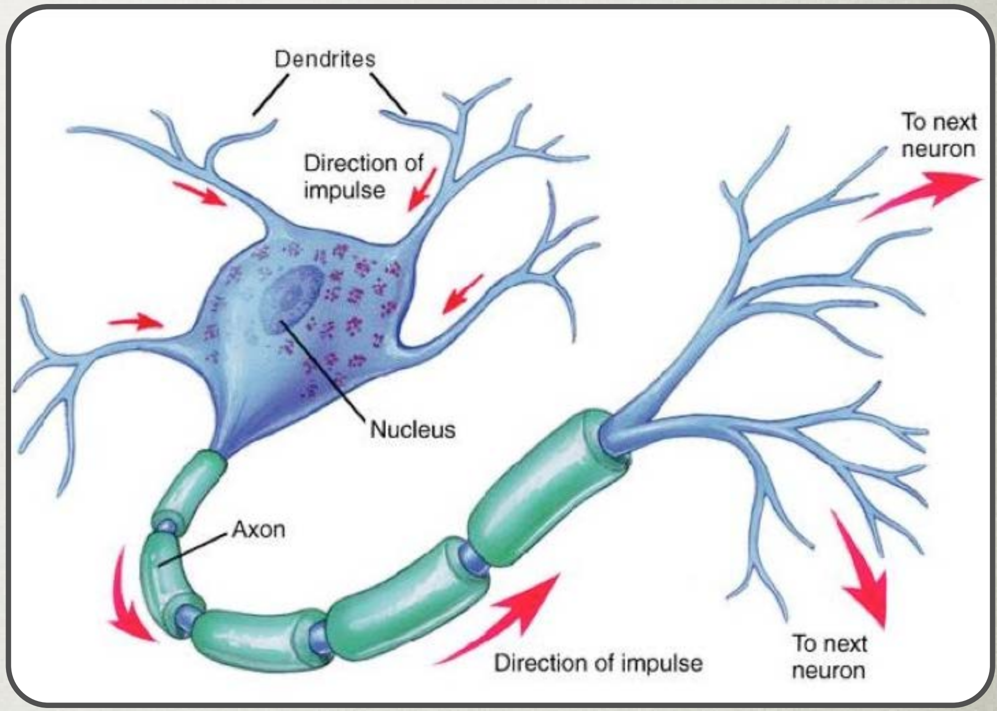
Algorithms in Nature

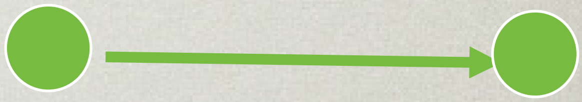
Pruning in neural networks

Neural network development

1. ***Efficient*** signal propagation
   [e.g. information processing & integration]
2. ***Robust*** to noise and failures
   [e.g. cell or synapse failure]
3. ***Cost-aware*** design
   [e.g. energy, metabolic constraints, wiring]



Abstracted to:



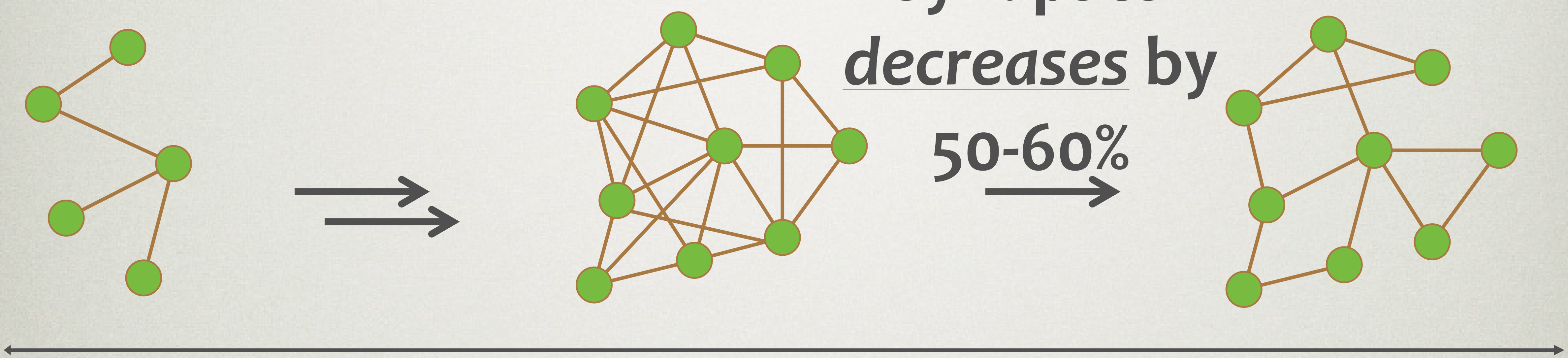
Pre-synaptic neuron; output along axon

Post-synaptic neuron; input via dendrites

[Laughlin & Sejnowski 2003]

Formation of neural networks

Density of synapses *decreases* by 50-60%

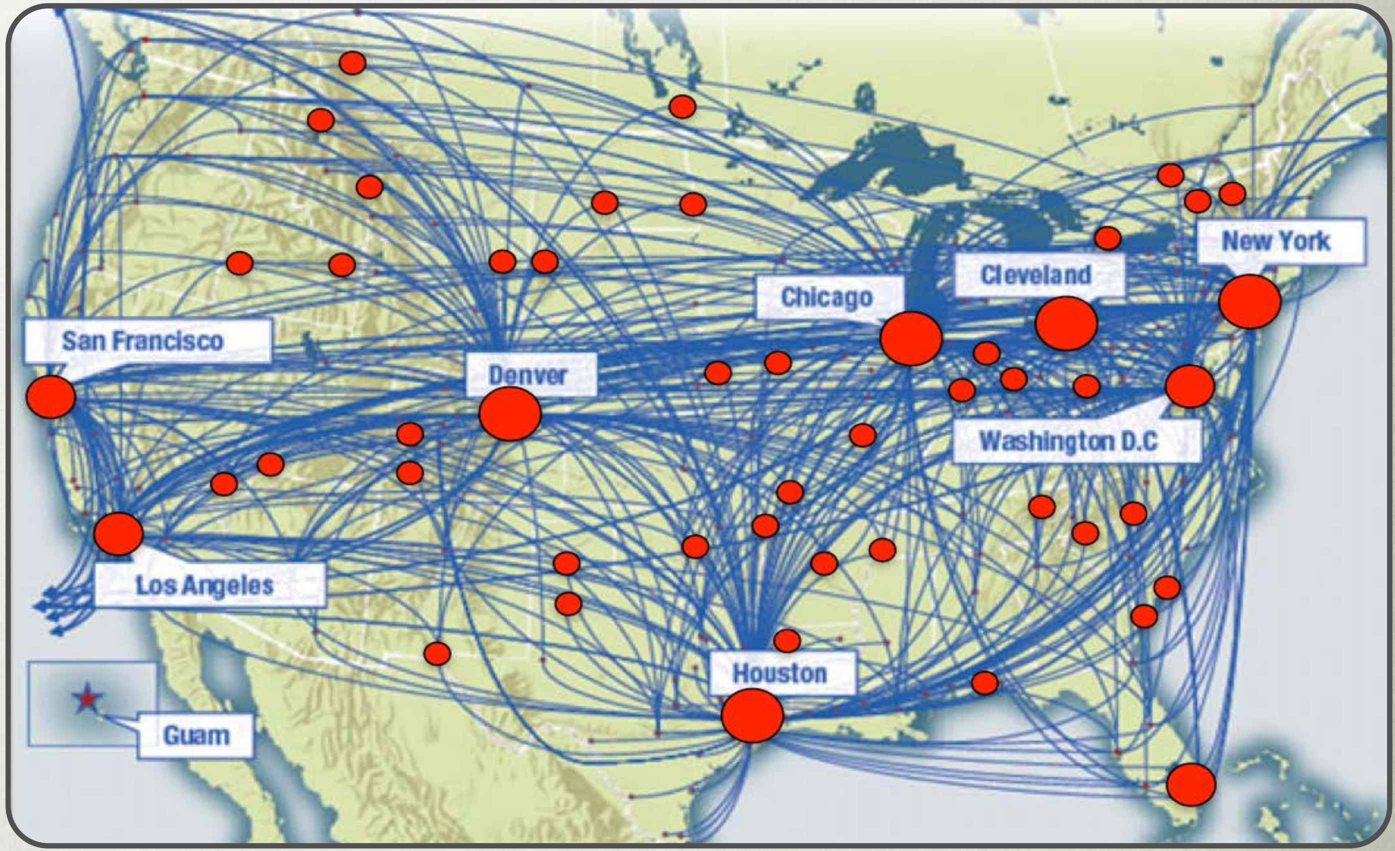


Synaptic pruning occurs in every brain region and organism studied that exhibits learning

Very different from current computational / engineering network design strategies!

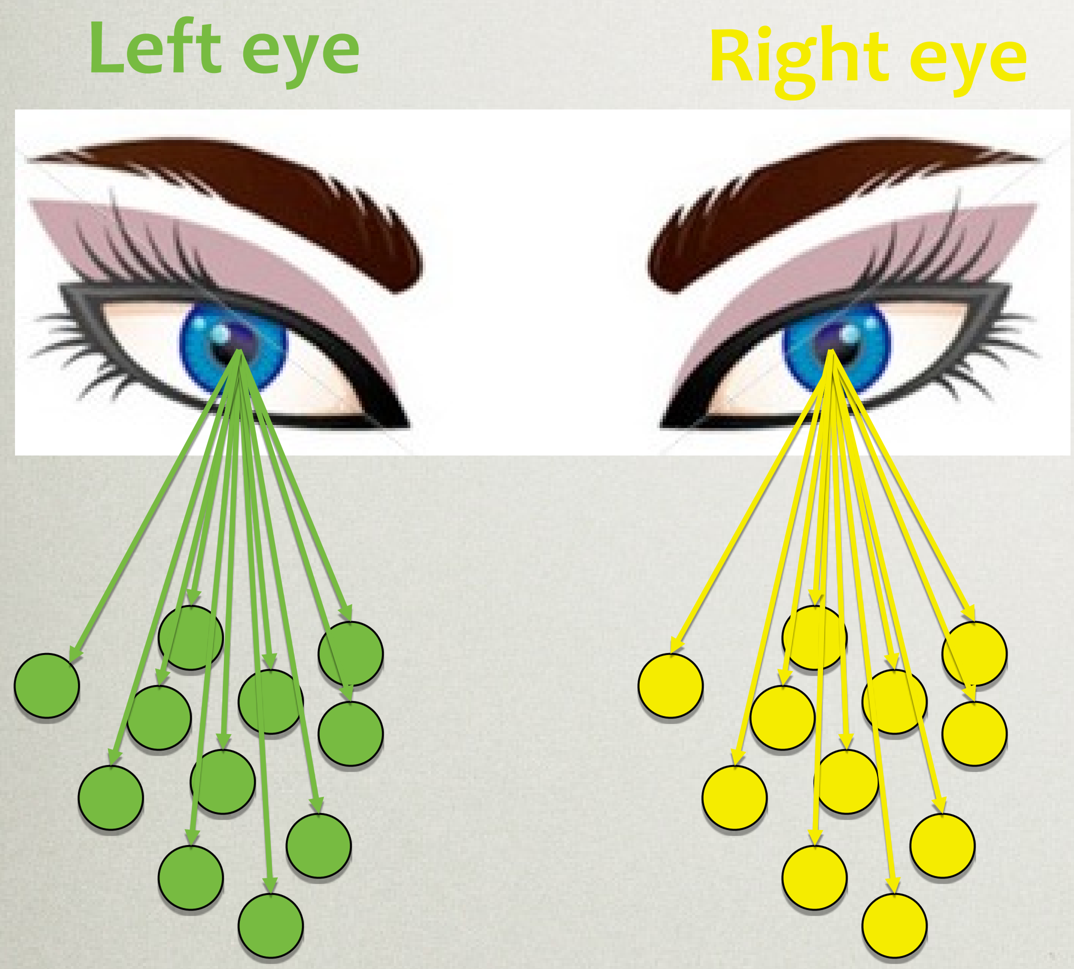
Engineered distributed networks:

- Engineered networks share similar goals: Efficiency, robustness, costs.
- Networks start sparse and can add more connections if needed
- A common starting strategy is based on spanning trees



airline routes, USA

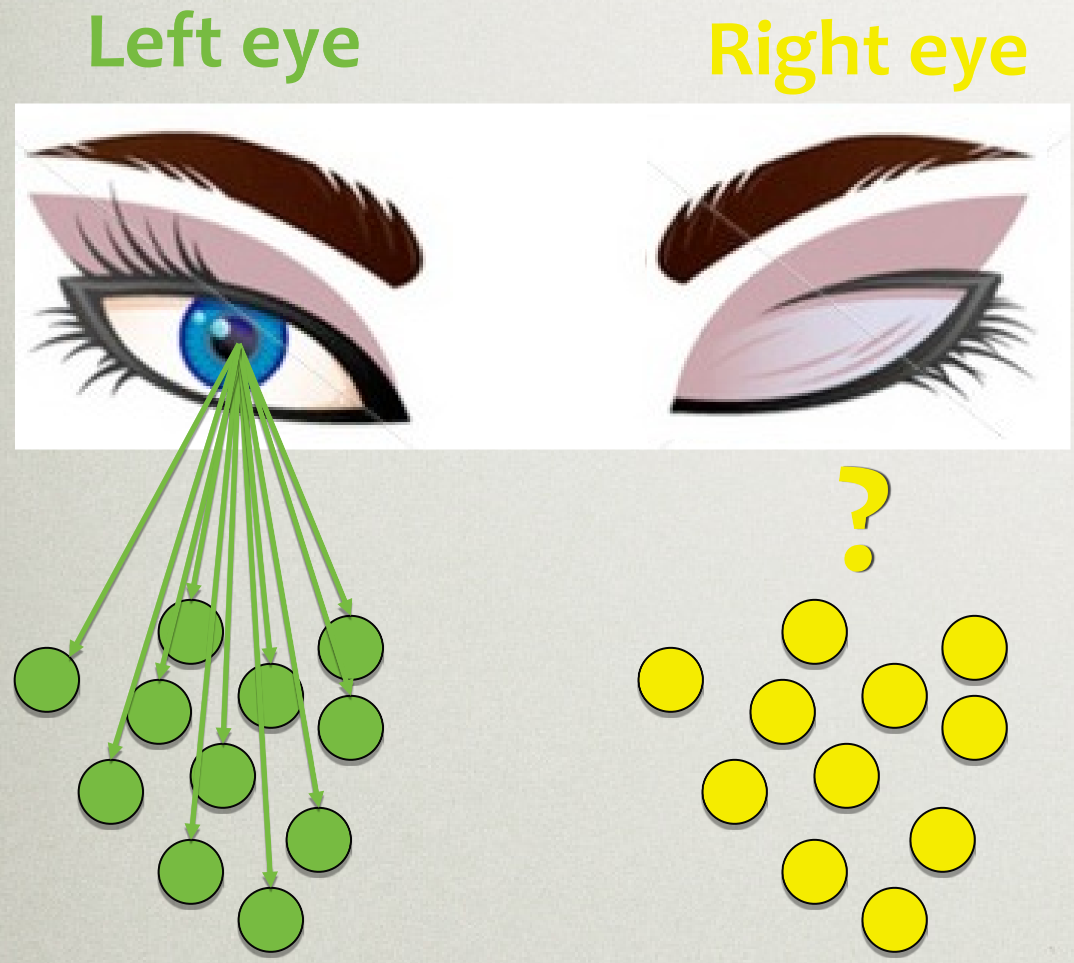
Advantages of pruning



Two sets of neurons that each respond to stimuli from one eye

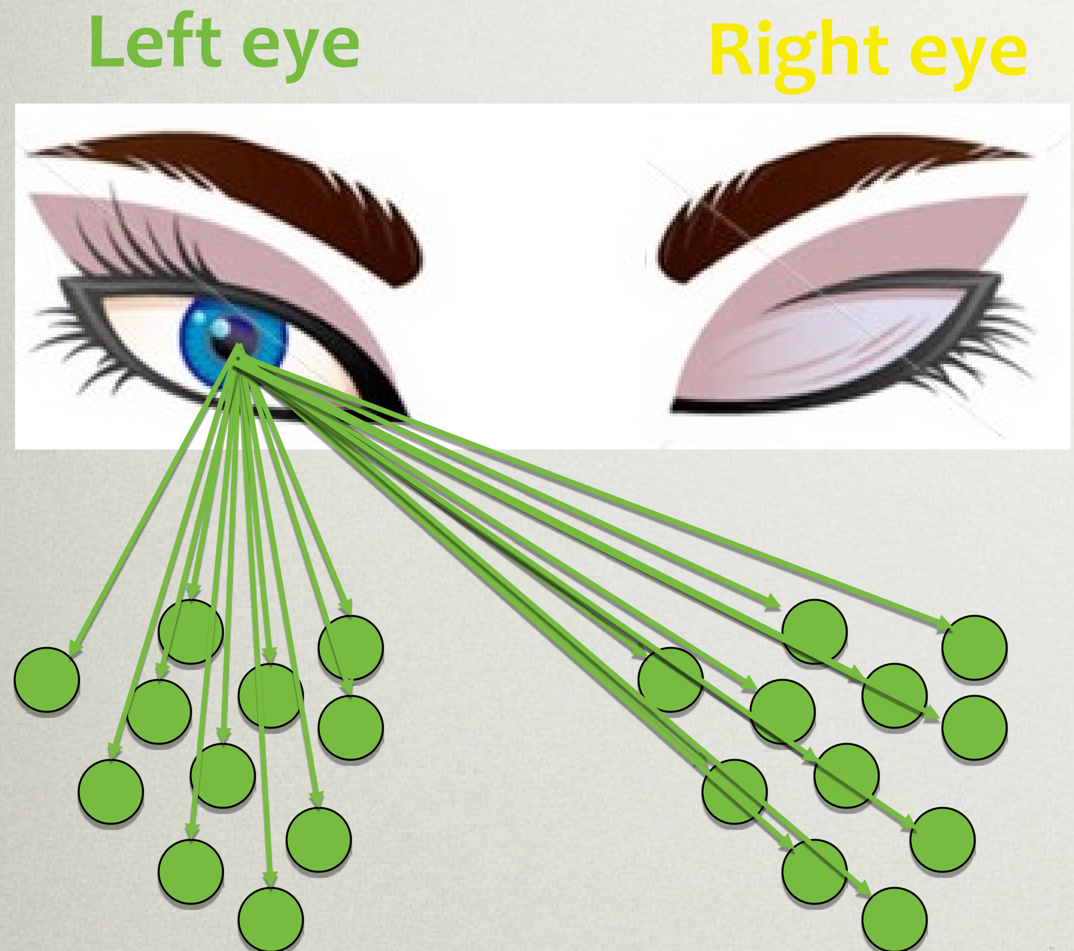
[Hubel & Wiesel,1970s]

Advantages of pruning



What happens to the neurons that now receive no input?

Advantages of pruning



Both sets of neurons respond to activity from the *same* eye

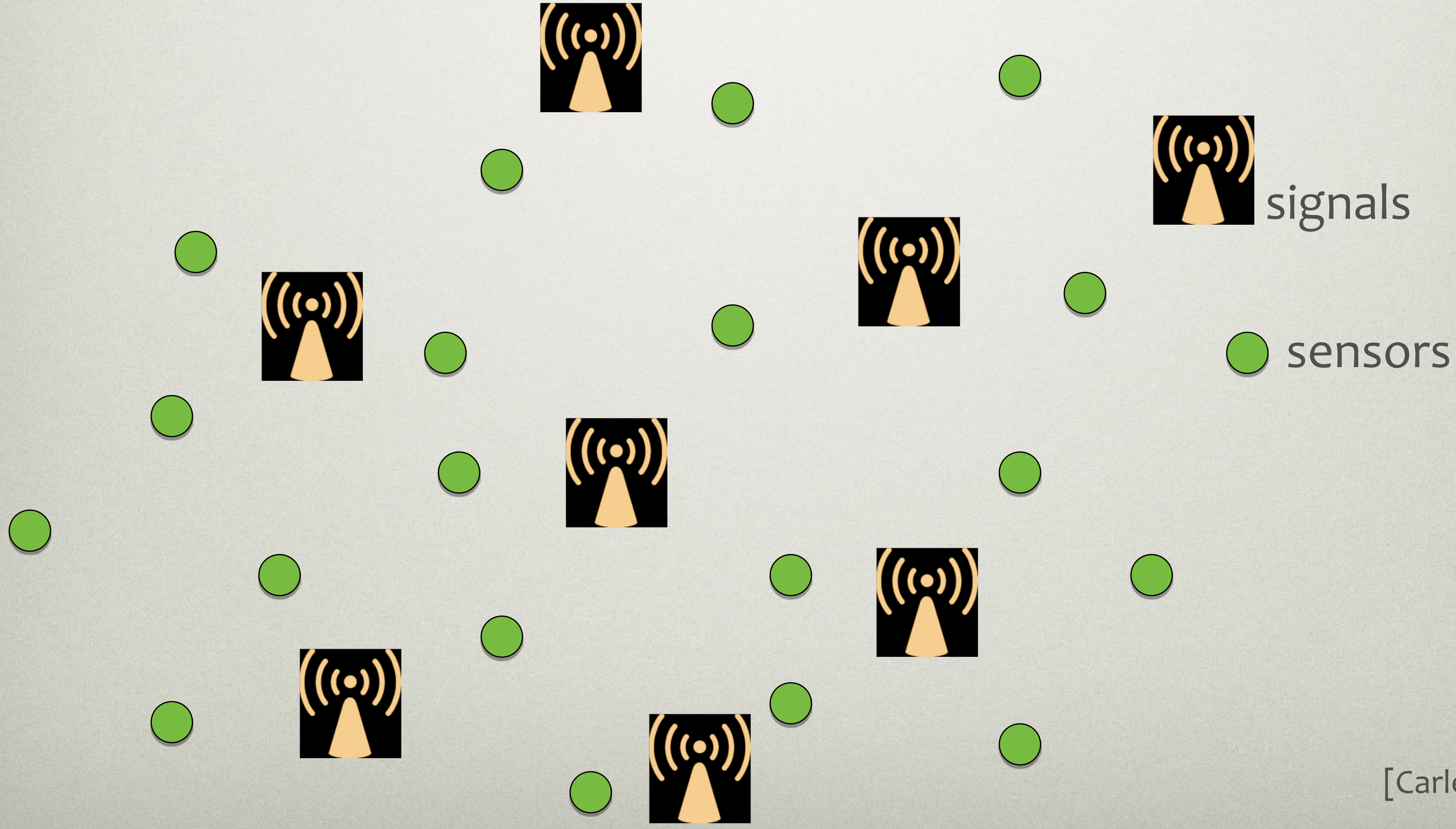
Why does this happen?

* Pool resources to compensate for loss of the right eye

* More *efficient* and *robust* use of neurons and connections

Distributed communication networks

In wireless networks, broadcast ranges are often required to be inferred based on active set of participants

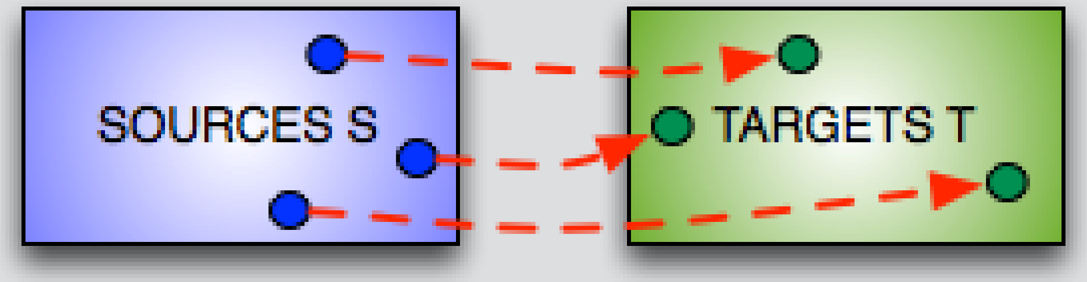


[Carle et al. 2004]

A theoretical model of network design

Input: Given n nodes and source-target pairs $\{(s_i, t_i)\}_{i=1}^{p}$ drawn from some *a priori* unknown distribution $\mathcal{D}$

For example:

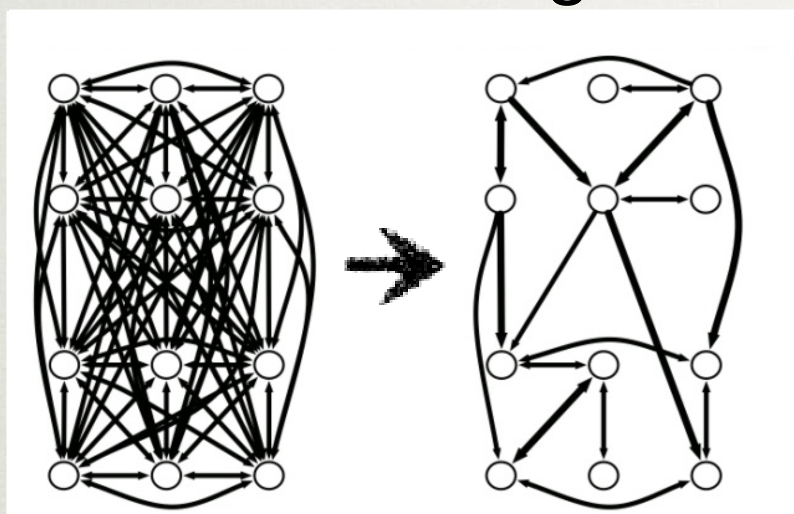


Output: A sparse graph G with $B \ll p$ edges that is "efficient" and "robust" with respect to $\mathcal{D}$

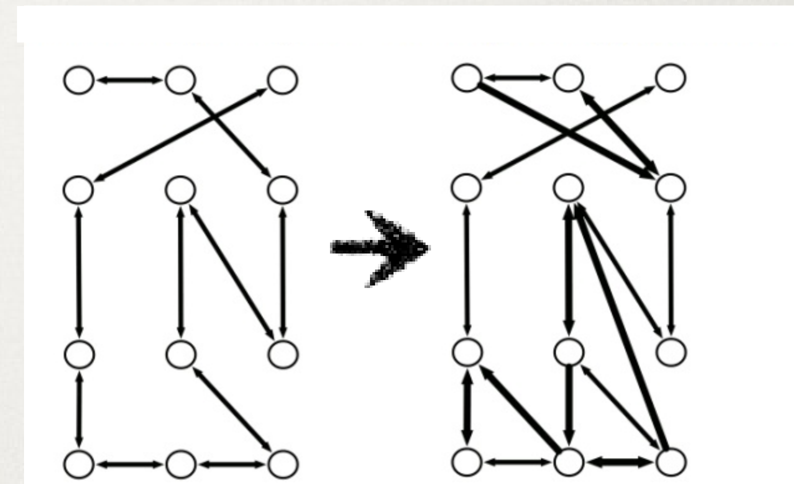
Constraints: Streaming Distributed

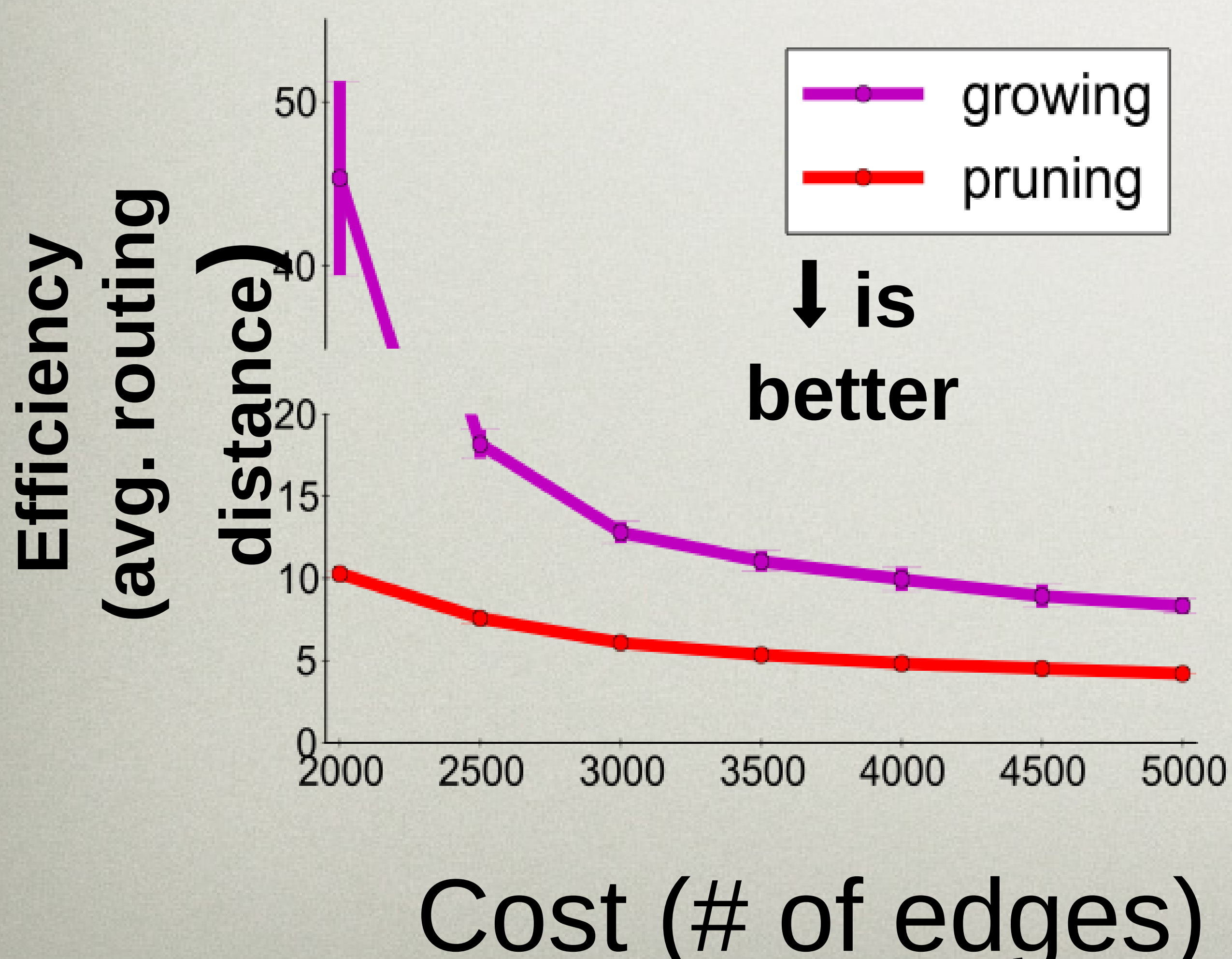
Pruning outperforms Growing

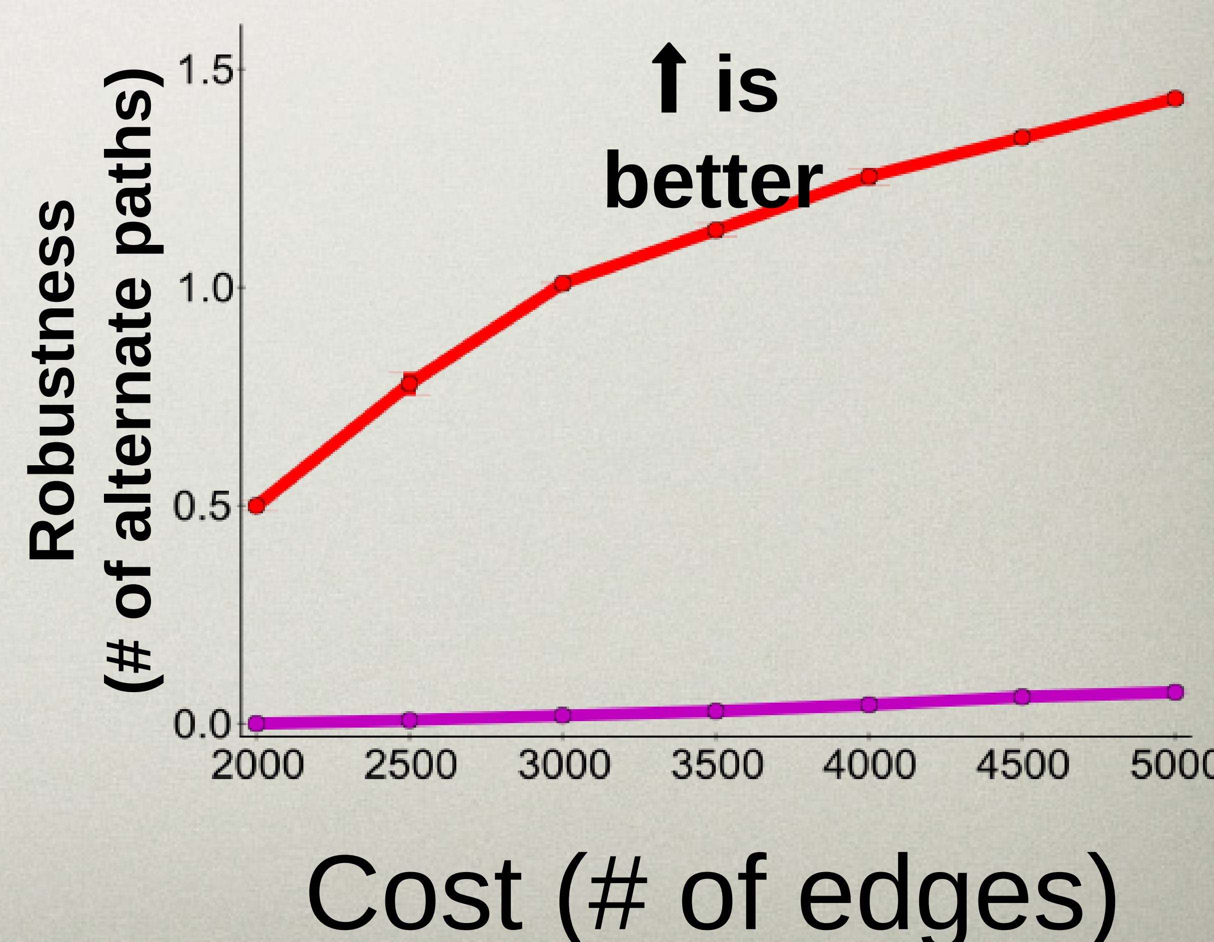
Pruning



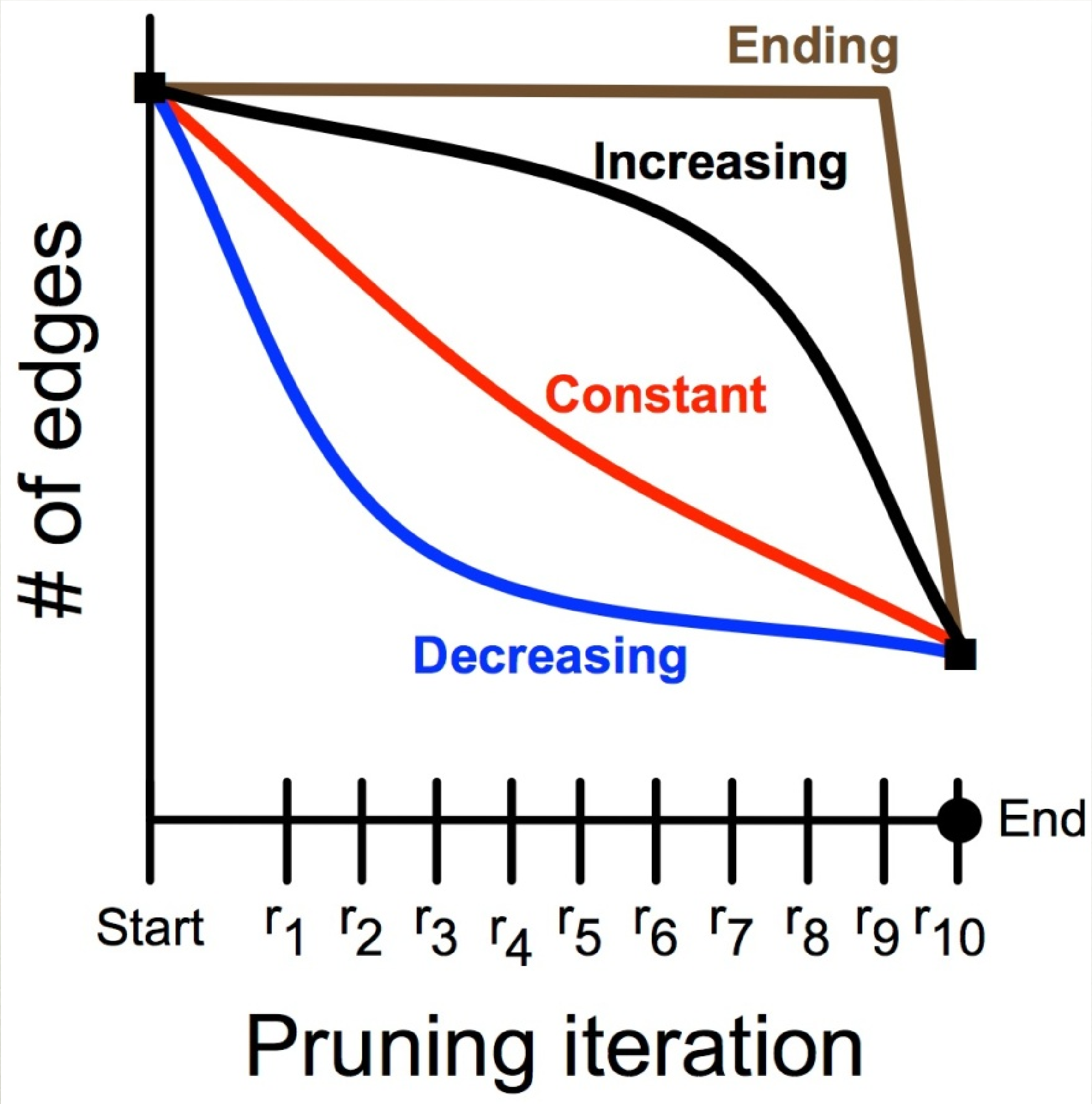
Growing



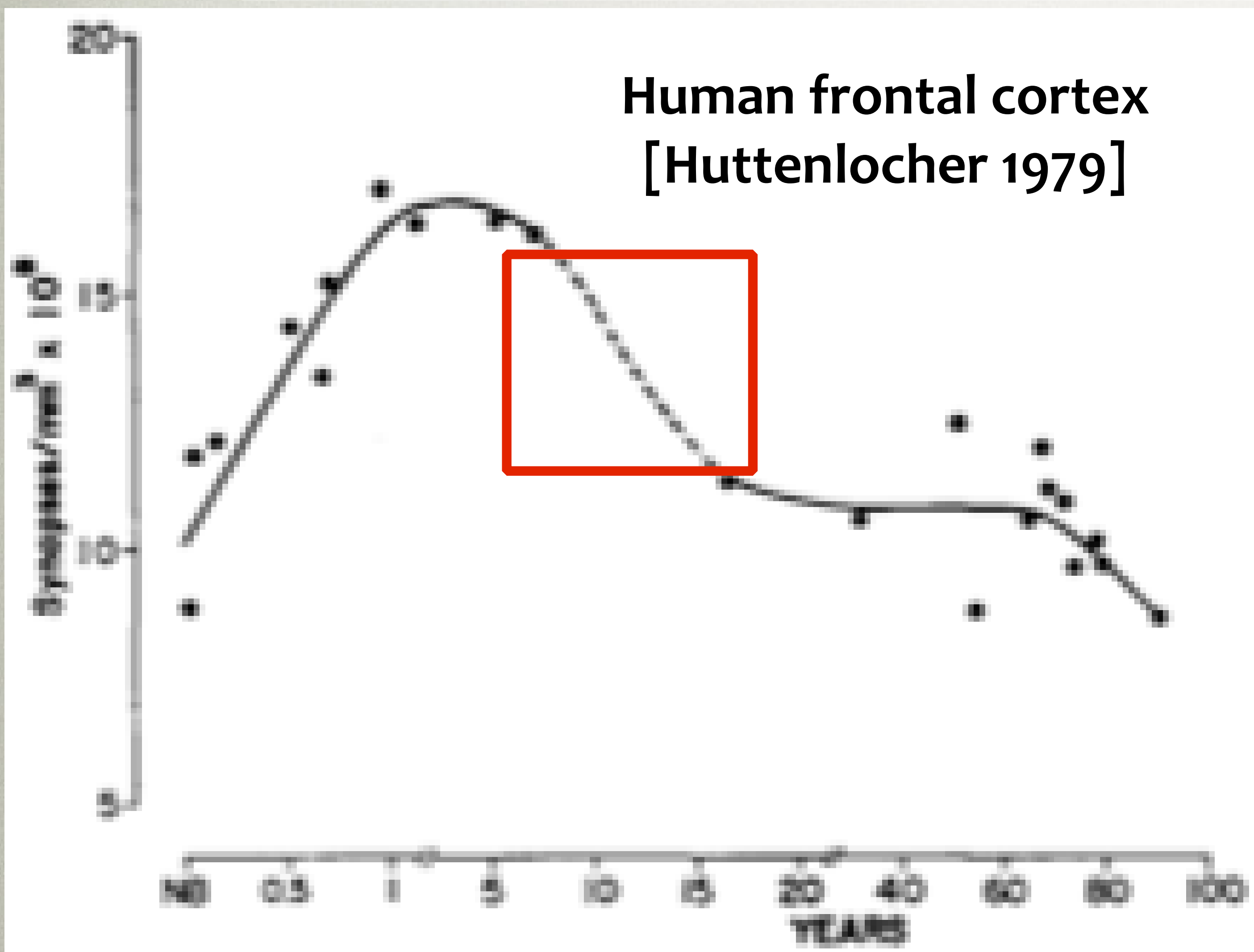


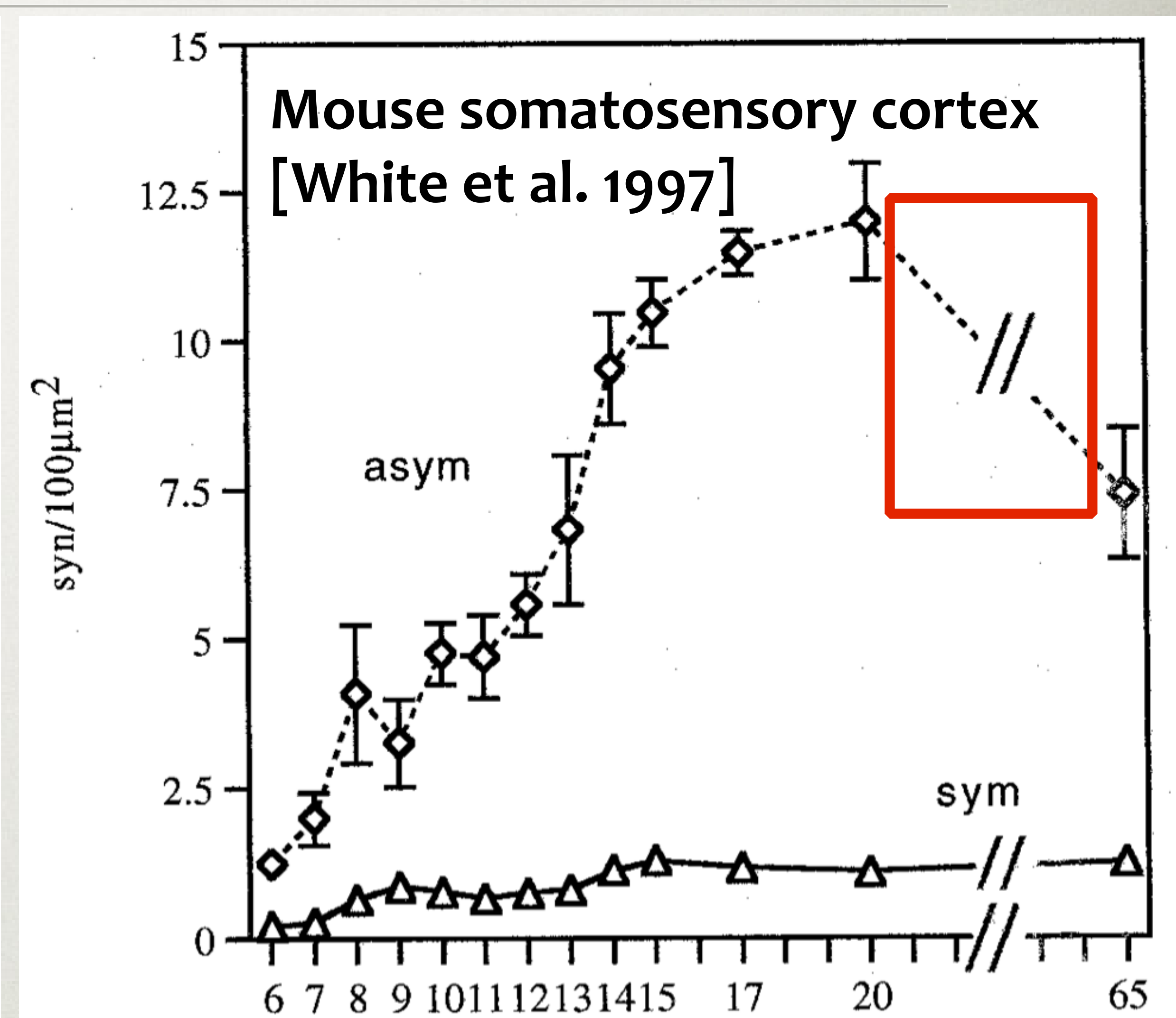


Does the rate of synapse pruning matter?

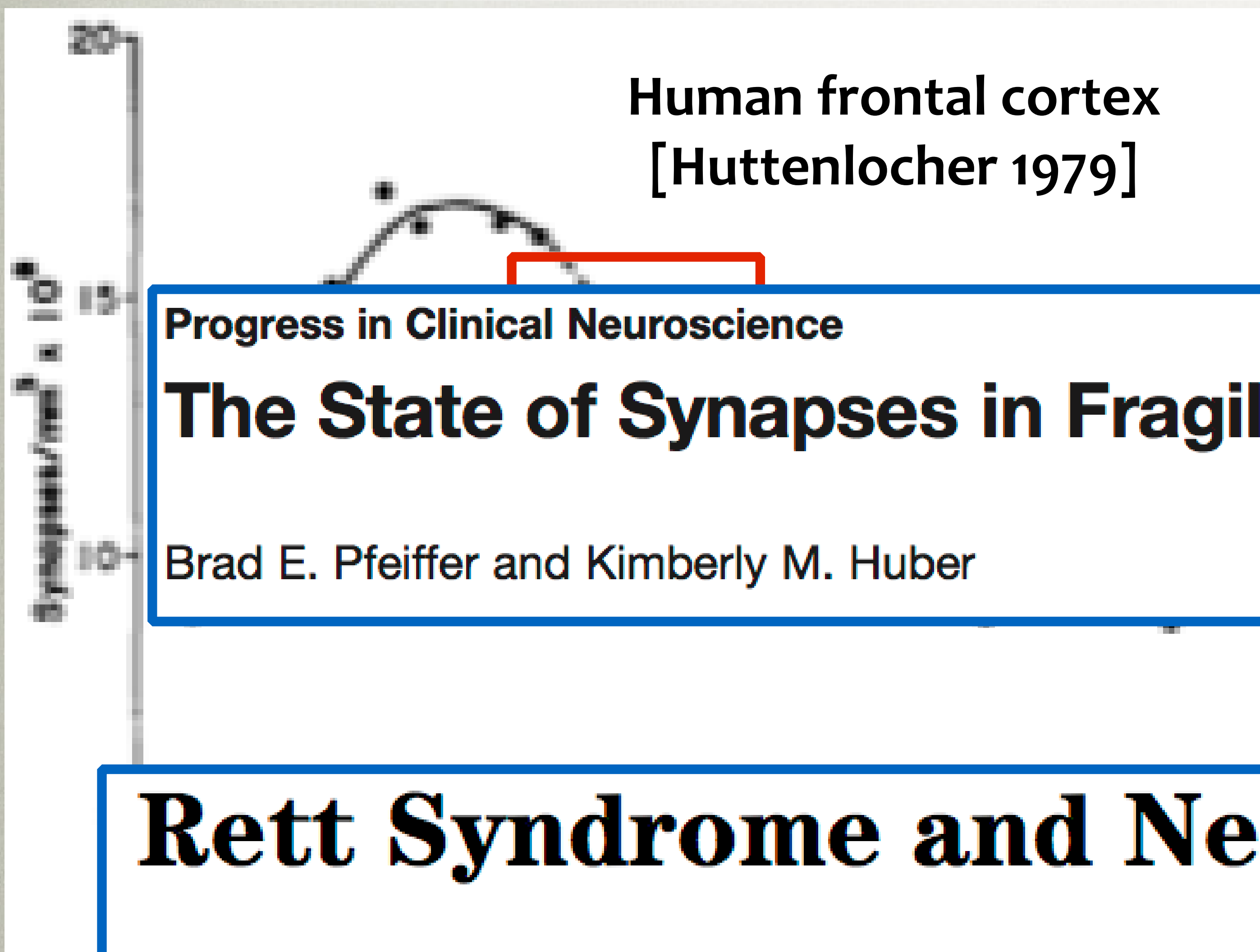


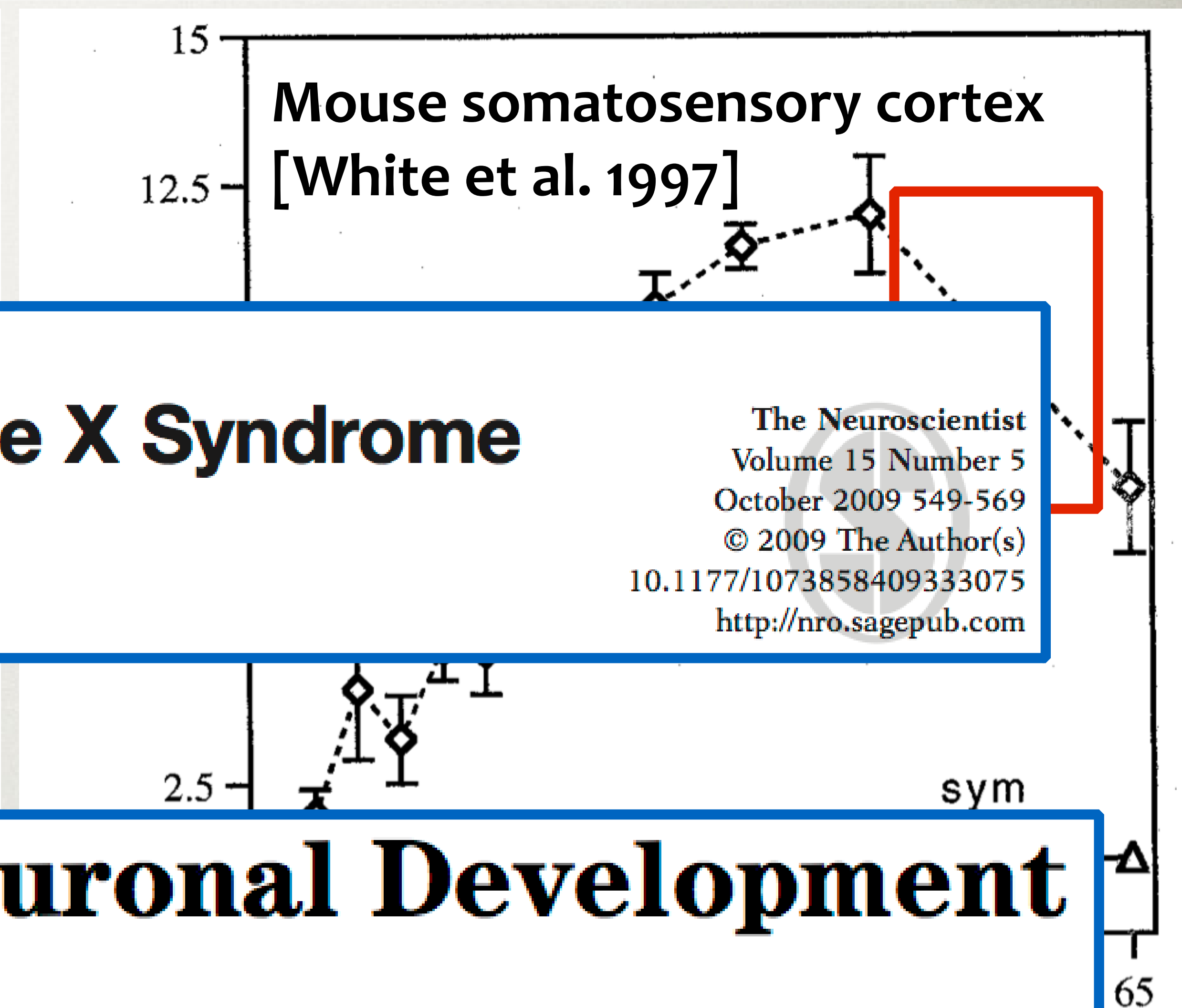
Pruning rates have been ignored in the literature





Pruning rates have been ignored in the literature





Progress in Clinical Neuroscience

The State of Synapses in Fragile X Syndrome

Brad E. Pfeiffer and Kimberly M. Huber

The Neuroscientist
Volume 15 Number 5
October 2009 549-569

10.1177/1073858409333075
http://nro.sagepub.com

Rett Syndrome and Neuronal Development

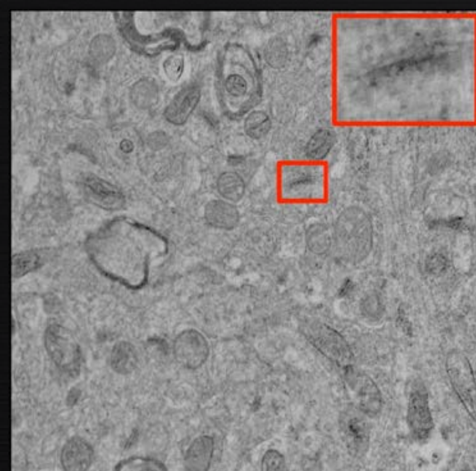
Michael V. Johnston, MD; Mary E. Blue, MD; Sakkubai Naidu, MD

Experimental techniques to detect synapses

Fast data collection

Slow data collection

Conventional EM



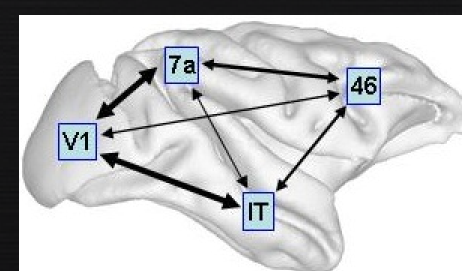
✓ Detect synapses, ultrastructure, pre- and post-synaptic neurons, etc

✗ Low-throughput analysis

Array Tomography [Micheva+Smith, 2007]

✗ Low-throughput analysis, cumbersome experimental technique

MRI [Honey et al. 2007]



✗ Detect synapses

?

✓ Detect synapses and measure synapse strength

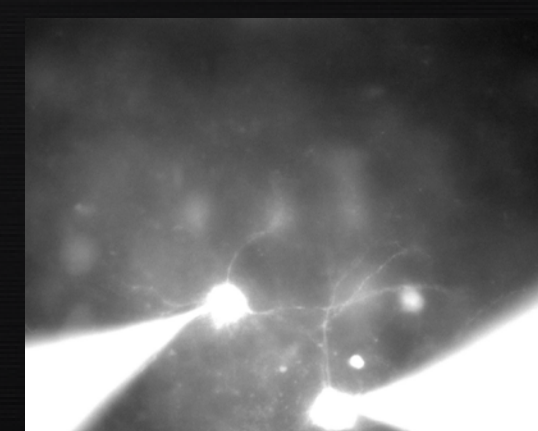
✓ High-throughput data analysis and collection

✗ Limited synapse types, failure rates, etc

mGRASP [Kim et al. 2012]

✗ Requires transgenic mouse

Electrophysiology



✓ Detect synapses, failure rates, neuron properties, etc

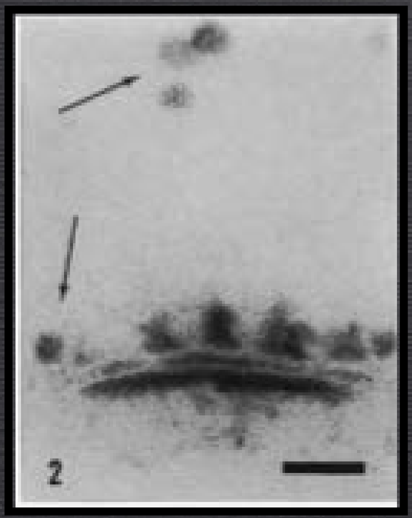
✗ Low-throughput collection

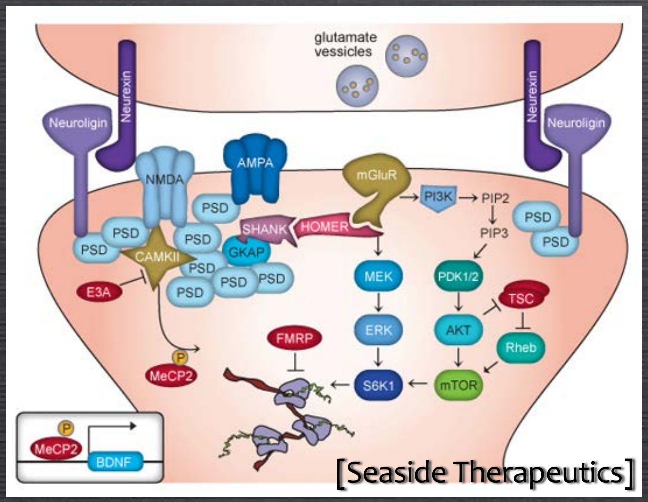
Slow data analysis

Fast data analysis

EPTA-staining

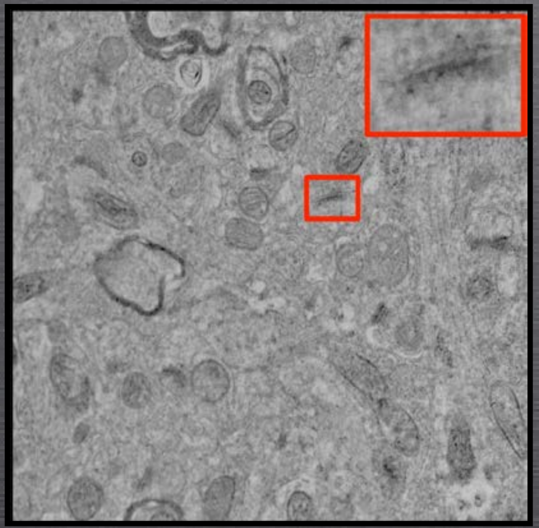
[Bloom and Aghajanian, *Science* 1966]





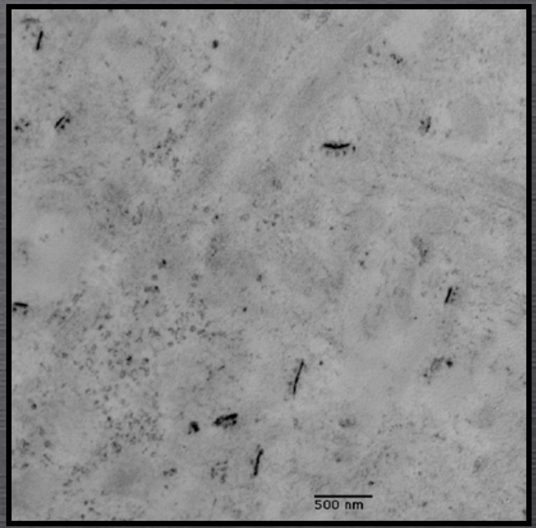
Ethanolic phosphotungstic acid (EPTA) targets proteins most prominently in the pre- and post-synaptic densities

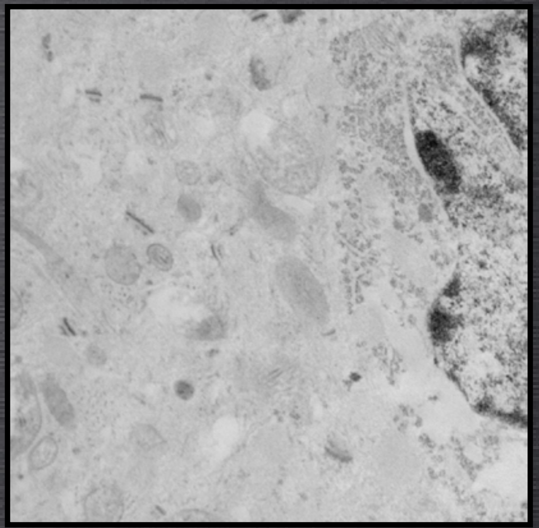
Conventional EM

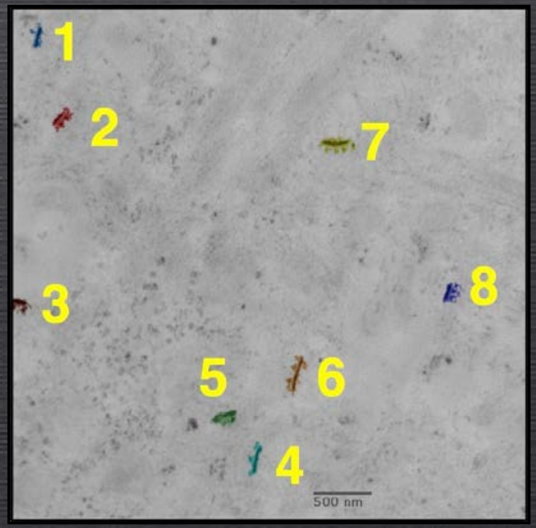


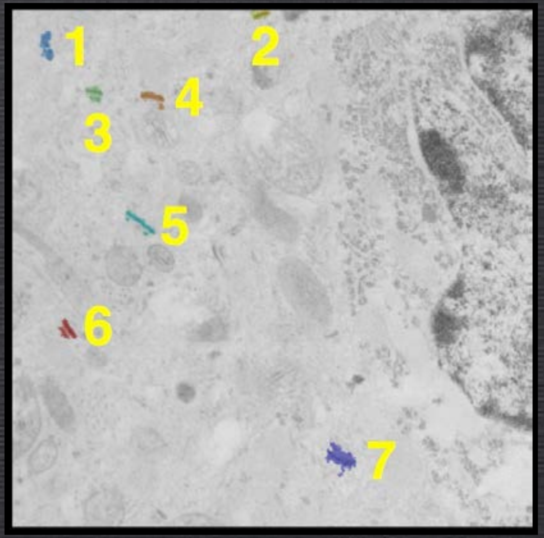
Hard to discern synapses

EPTA-based EM





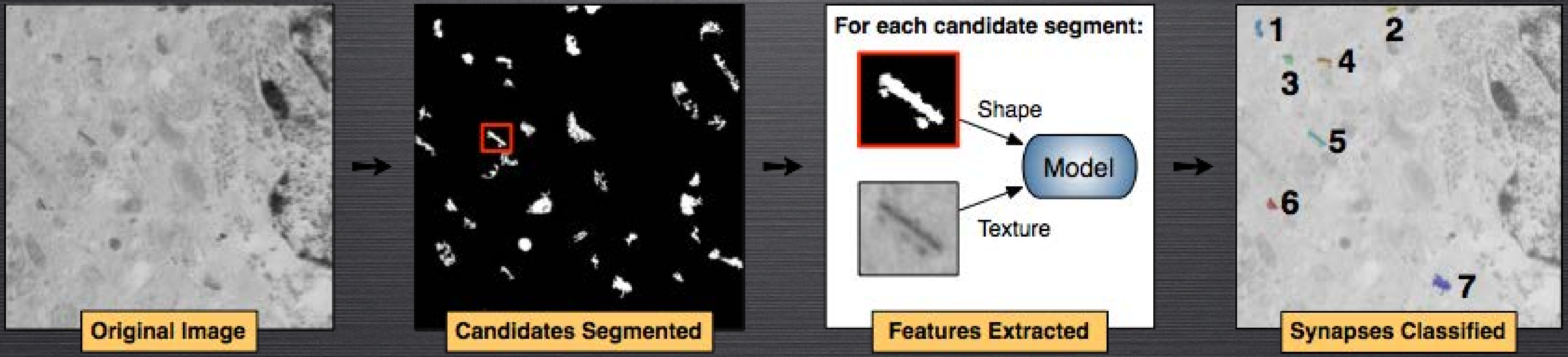




Pipeline for detecting synapses

EM images are inherently noisy due to variations in the:

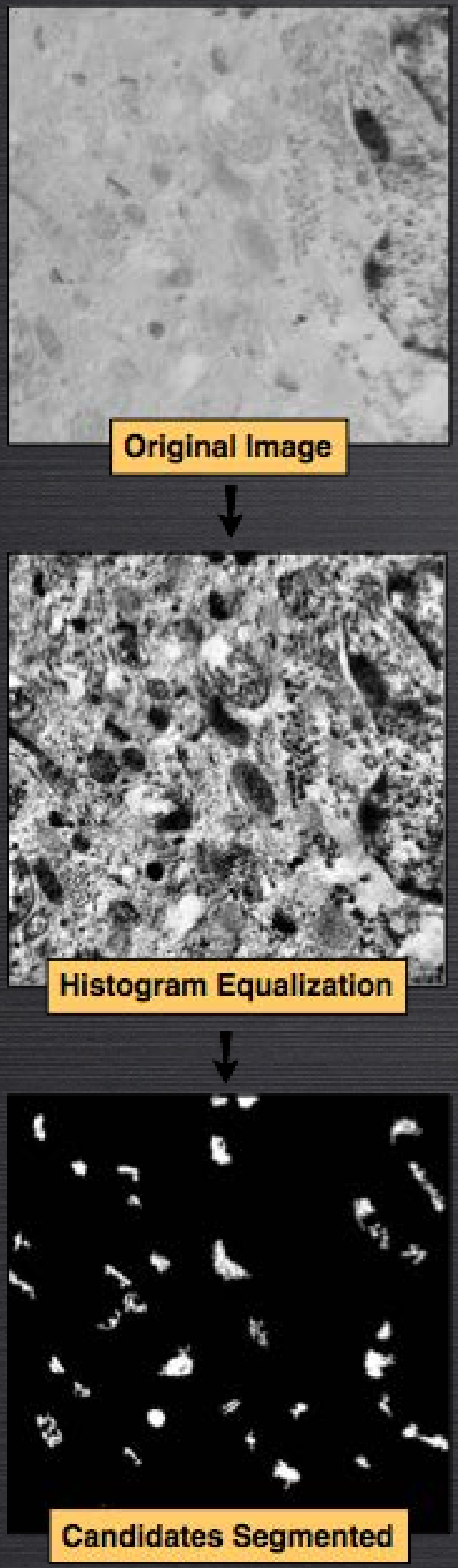
1. Tissue sample (e.g. age, brain region)
2. EPTA chemical reactions
3. Image acquisition process (e.g. microscope, illumination, focus)



Step 1. Unsupervised segmentation

Step 2. Extract window and normalize

Steps 3+4. Extract features and build classifier



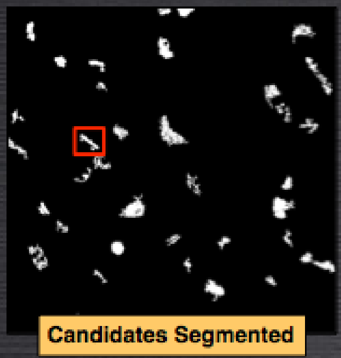
Step 1. Image segmentation

Adaptive histogram equalization [Zuiderveld, 1994]:

* Enhances contrast in each local window to match a flattened histogram; windows combined using bilinear interpolation to smoothen boundaries

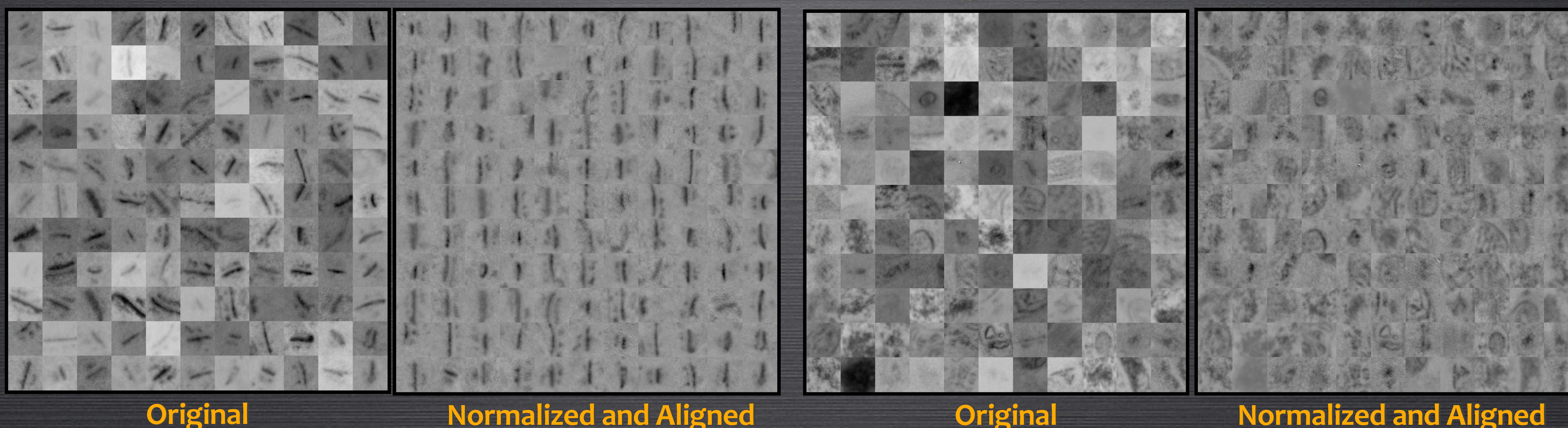
Unsupervised segmentation:

* Binarize using a single *sample-independent* threshold (10%)
* Lose only 1% of synapses in this step (two adjacent synapses get merged)



Step 2. Reduce heterogeneity

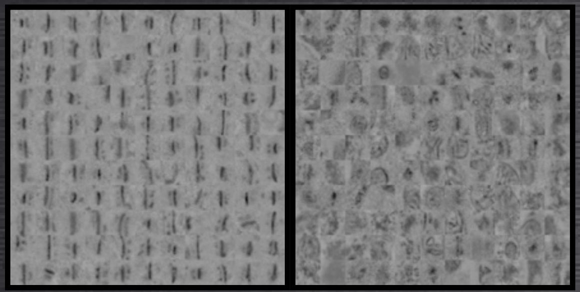
Positive windows (synapses) — Negative windows (non-synapses)



* Extract surrounding window: 75x75-pixel window *W* (~325nm^2) around segment centroid.

* Normalize window: $$W = \frac{W - \mu(W)}{\sigma(W)}$$

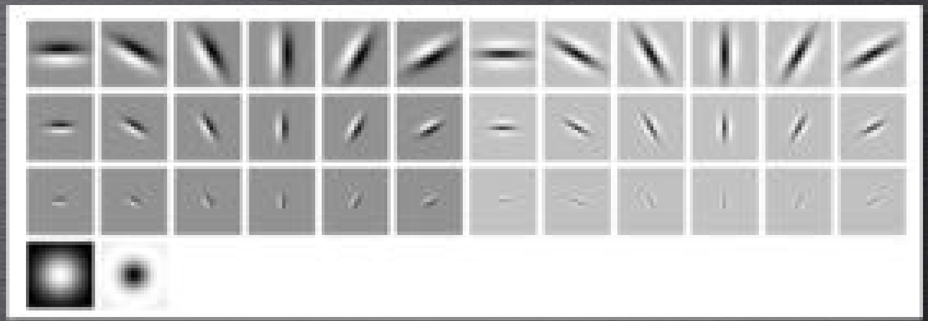
* Align vertically: Hough transform



Step 3. Extract features

Texture: a common cue used by humans when manually segmenting EM images [Arbelaez et al. 2011]

MR8 filter bank: 38 filters (max of 6 orientations at 3 scales for 2 oriented filters, + 2 isotropic) = 8-dim filter response vector at each pixel

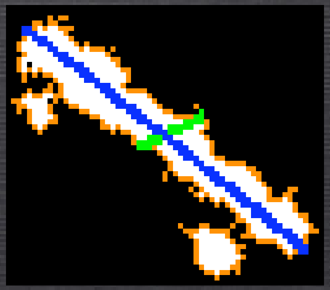


[Varma and Zisserman, 2004]

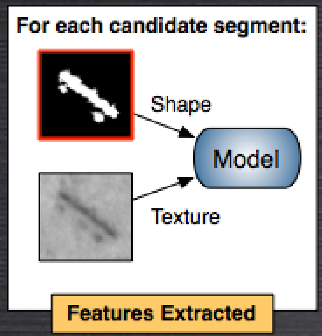
HoG: histogram of oriented gradients [Dalal+Triggs, 2005]

Shape: synapses are typically long and elongated

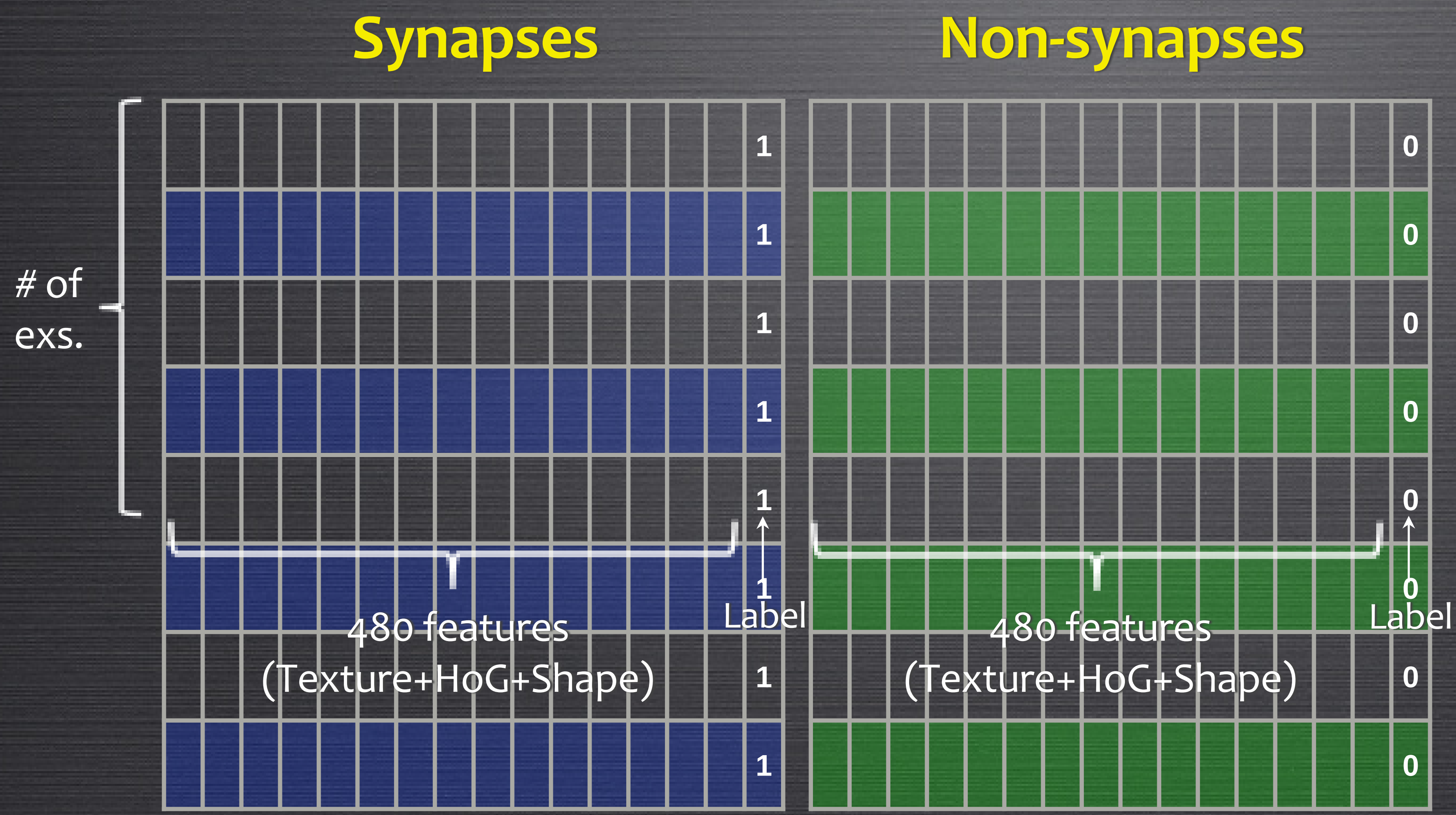
10 features for each segment: Length, Width, Perimeter, Area, etc.



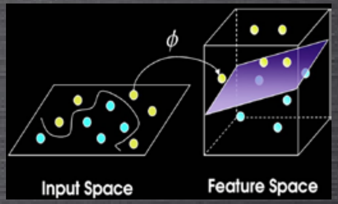
⇒ **Overall**: each window represented by a $\mathrm{480\text{-}dim\ vector} \in \mathbb{R}^n$ scaled to $[0, 1]$



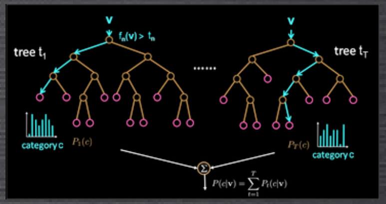
Step 4. Build classifier



SVM [Chang+Lin, 2011]

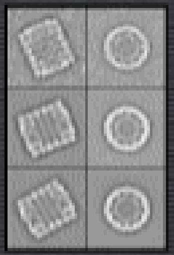


Random Forest [Breiman, 2001]

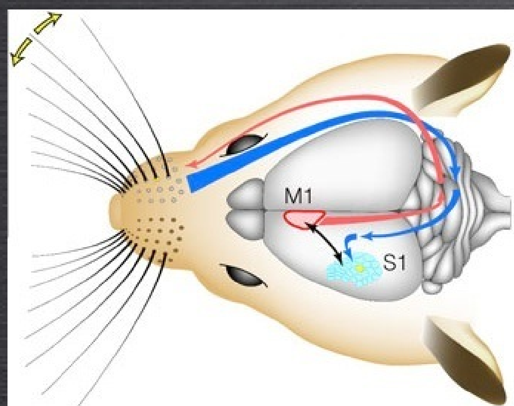


AdaBoost [Freund+Schapire, 1995]

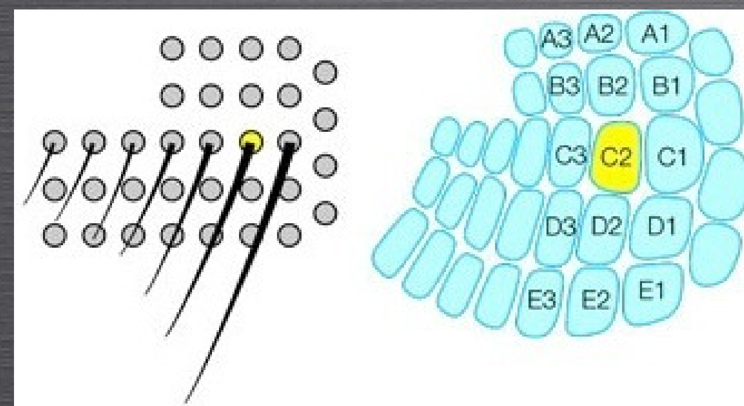
Template Matching [Roseman, 2004]



Experiments performed and data collected



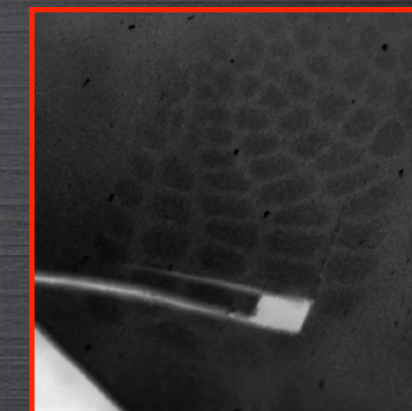
Somatosensory (whisker) cortex in the mouse



1-1 somatotopic mapping from whiskers to columns

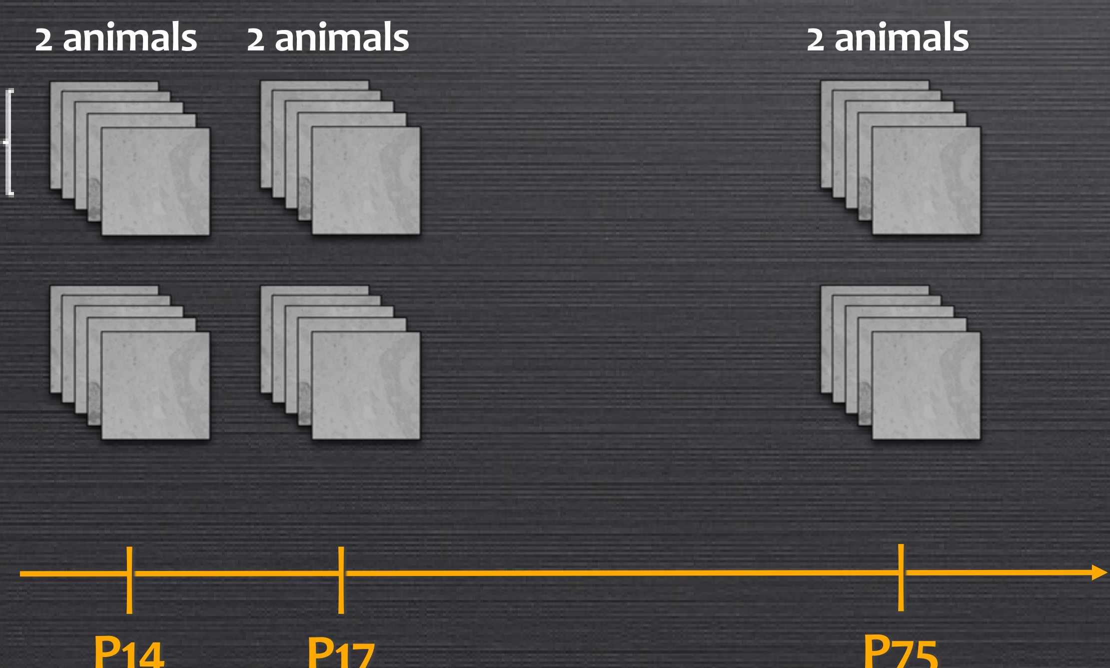
[Aronoff+Petersen, 2008]

Staining barrels with cytochrome oxidase



Dissecting D1 barrel

130 images per animal covering 3,000 um²



Post-natal age (day) of mouse

Accurately detecting synapses in EPTA images

Training data: for P14 and P17, we manually labeled 11% of the 520 EPTA images (counting 230 synapses and 2062 non-synapses)

10-fold cross-validation

SVM outperformed all other methods:

AUC ROC = 96.4%

AUC PR = 73.8%

At default classifier threshold (0.5):

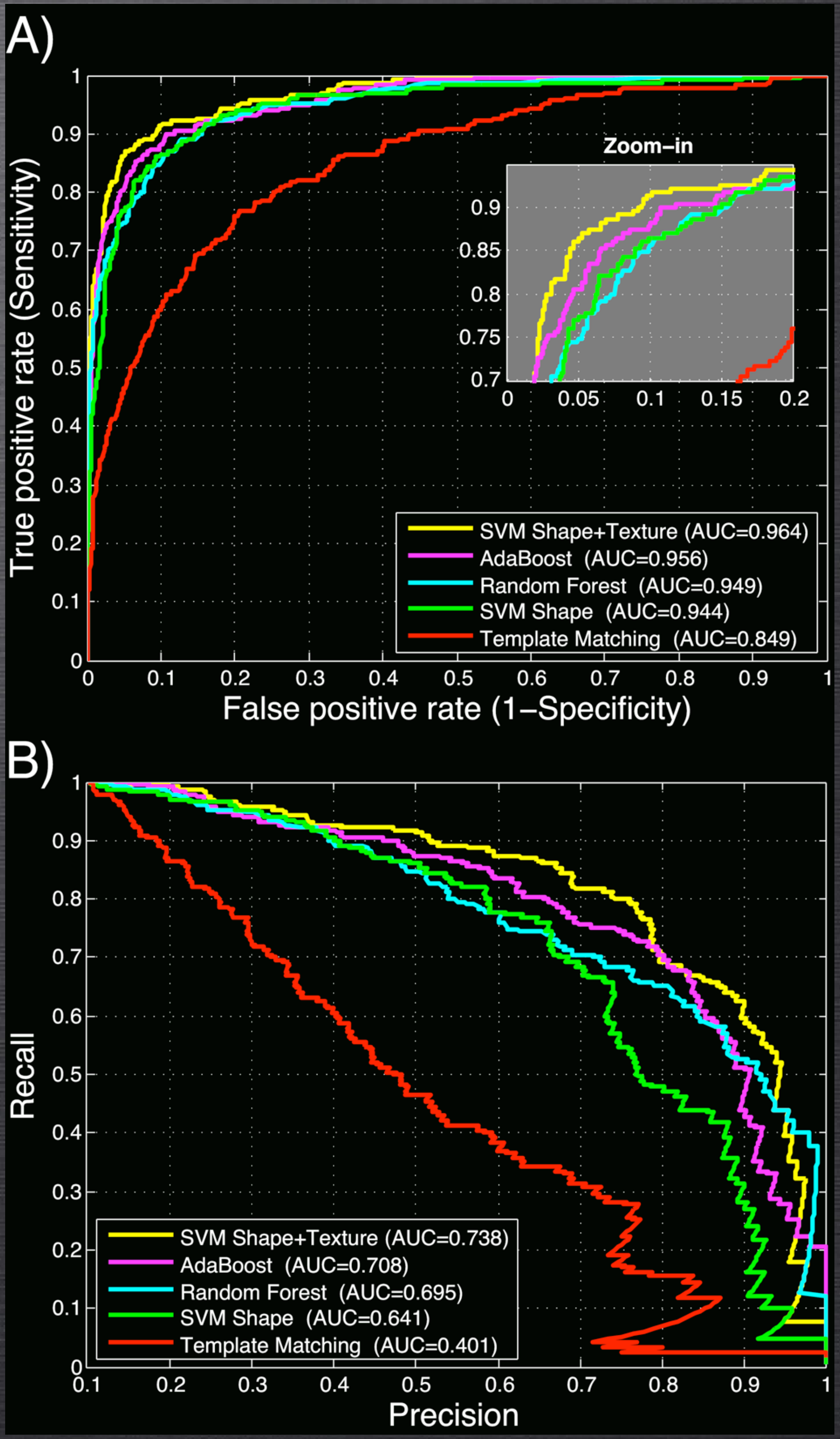
Precision = 83.3%

Recall = 67.8%

Validation against independent human annotation of 30 EPTA images:

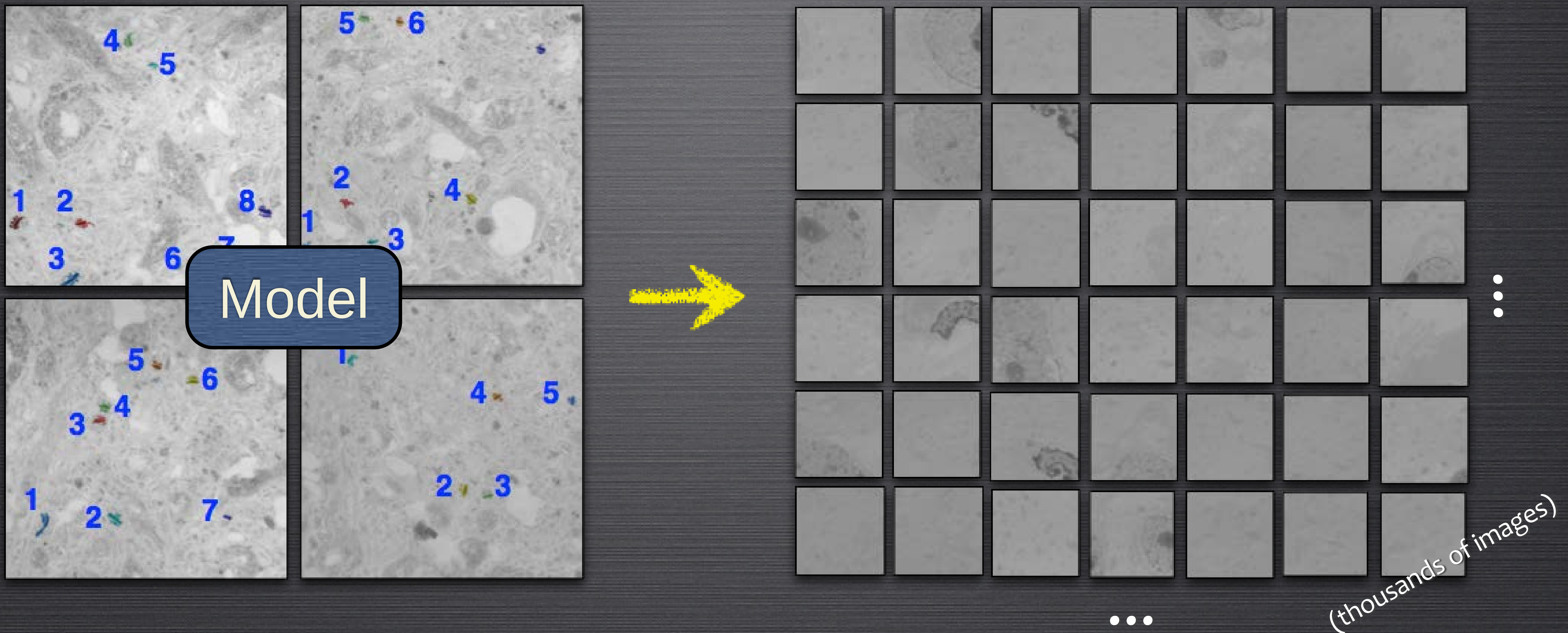
Precision = 87.3%

Recall = 66.6%



Labeled images from Sample A used to build classifier

Unlabeled images from Sample B to analyze; variable staining and noise vs. A



It would be laborious to build a new classifier for every new sample...

Can we improve the model by leveraging the enormous number of unlabeled images available?

Co-training algorithm

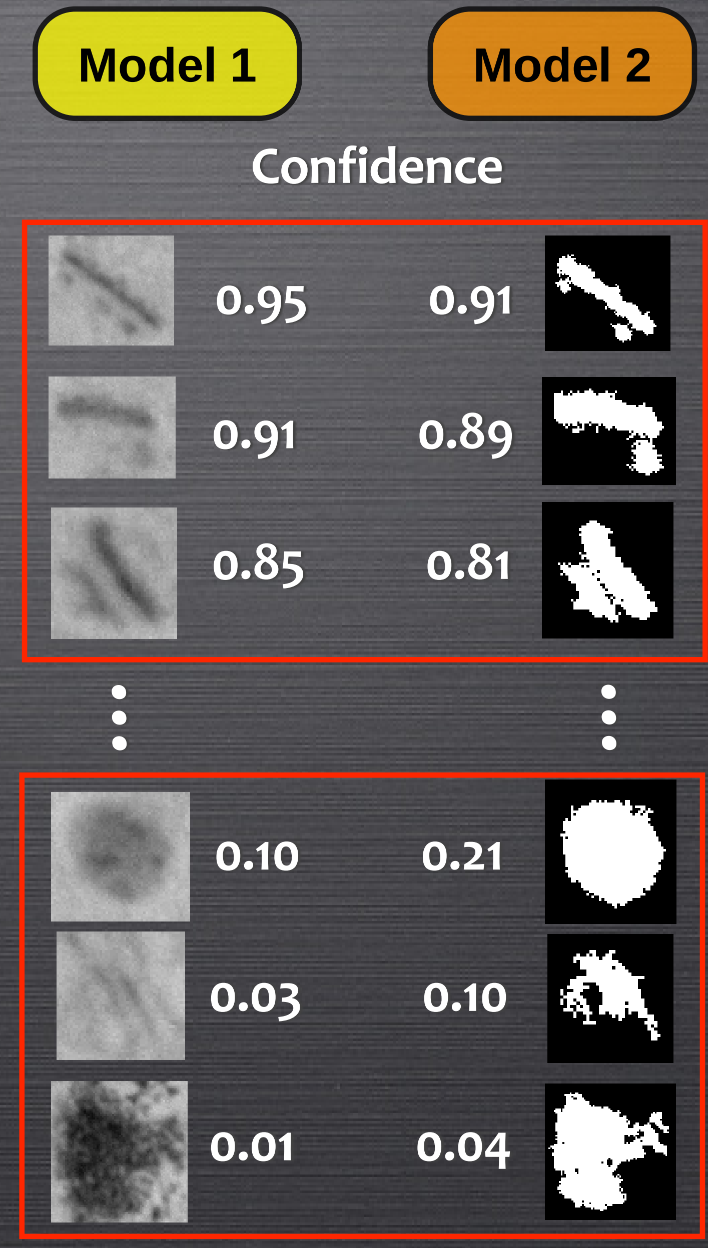
[Blum and Mitchell, *COLT* 1998]

Labeled images from Sample A

Model 1
Texture+HoG

Model 2
Shape

→ Apply each model to *Unlabeled* images from Sample B



Co-trained B+
Keep top k%

} Discard

Co-trained B–
Keep same pos:neg ratio

Blum and Mitchell (1998) proved that under some conditions, the target concept can be learned (PAC model) using few labeled and many unlabeled examples using such a co-training algorithm.

⇒ Retrain single model on examples from: Labeled A, Co-trained B+ and B–

Semi-supervised learning improves classification accuracy

Labeled P75, Unlabeled P14

		—Accuracy—		—AUC—	
Train/Test	Co-training	Positive	Negative	Prec-Recall	ROC
P75 / P14	No	66.36%	98.20%	73.65%	96.91%
P75 / P14	Yes (0.5%)	72.90%	**98.60%**	**75.75%**	**97.14%**
P75 / P14	Yes (1.5%)	**74.77%**	96.91%	73.06%	96.65%

⇒ Baseline

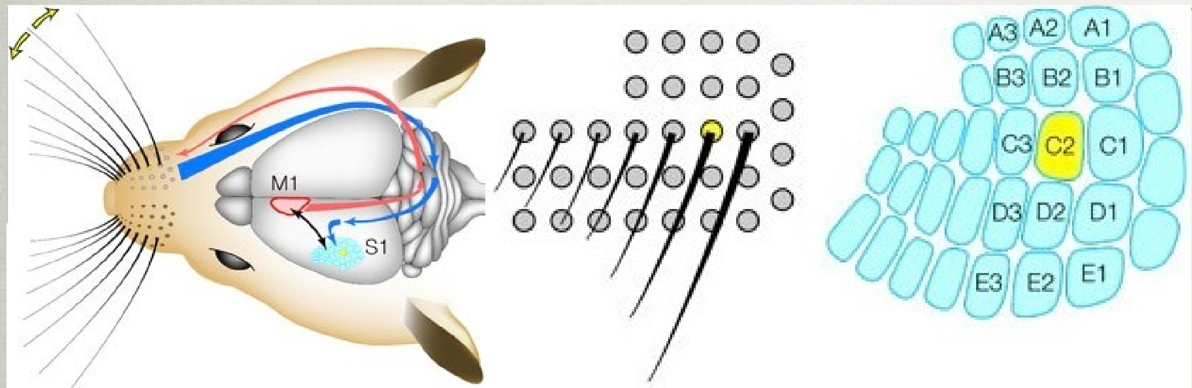
Percentage of unlabeled examples to include in co-trained classifier

[redacted]

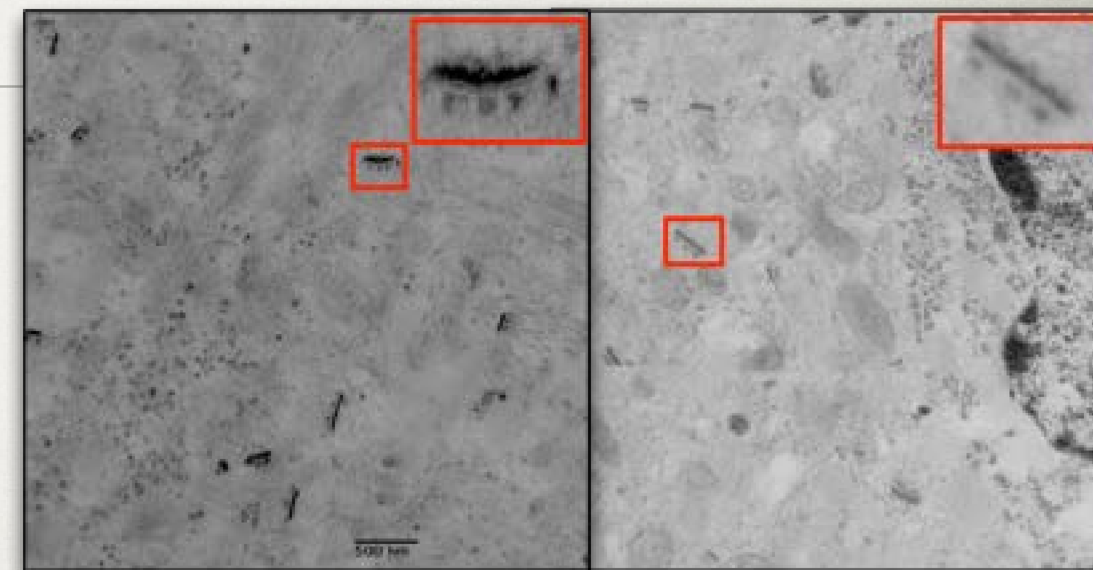
Co-training increases accuracy of positive examples by 8-12% and AUC by 1-4%

... but including too many unlabeled examples (1.5%) can decrease performance

Experimentally quantifying pruning rates



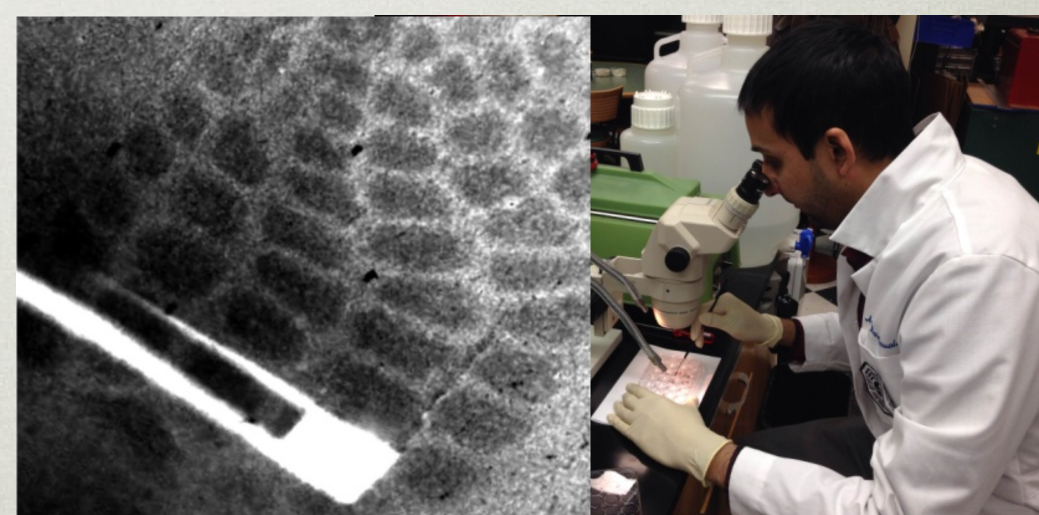
Mouse somatosensory cortex:
whiskers ⇒ columns



Electron microscopy images

imaging

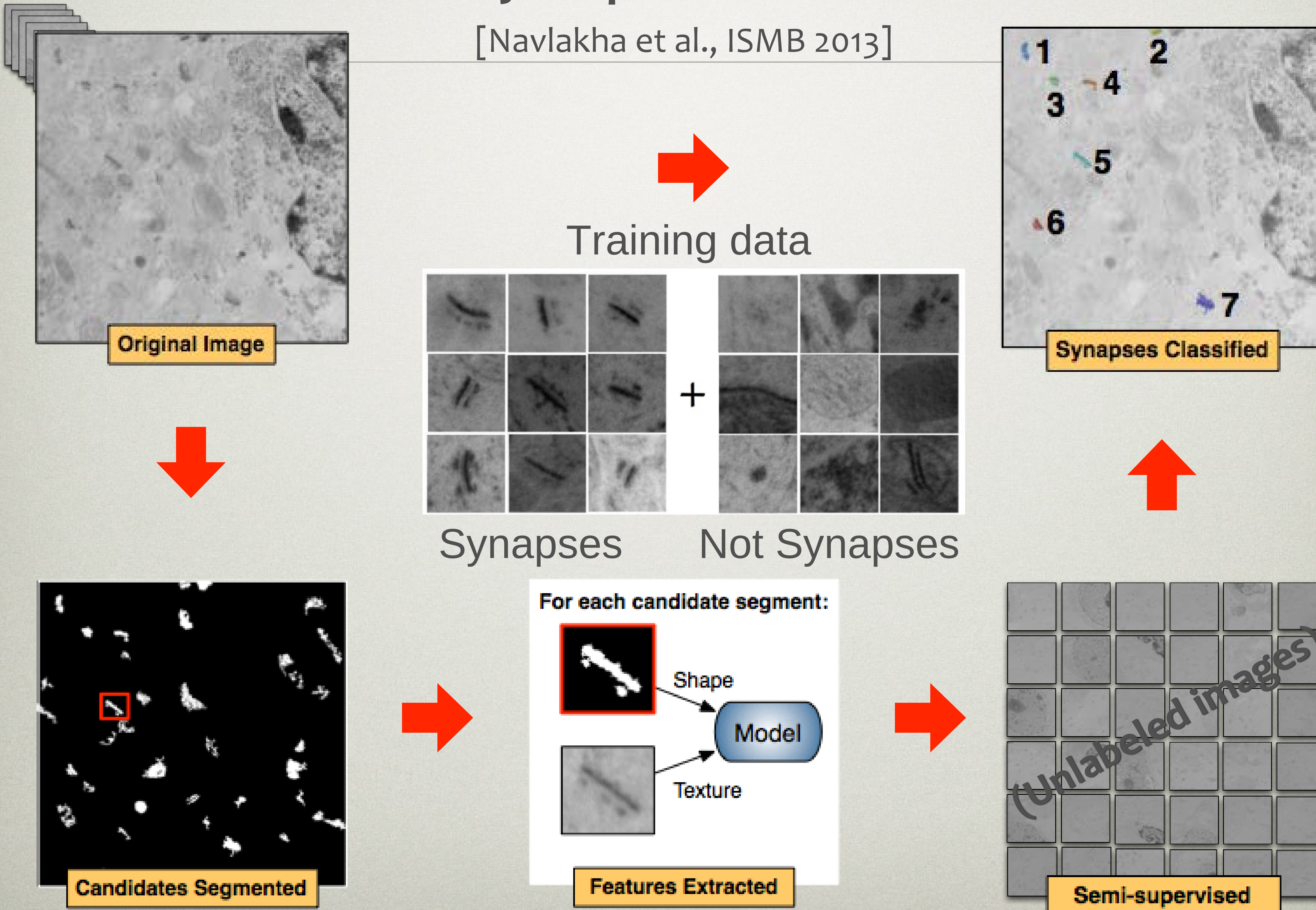
slice brain



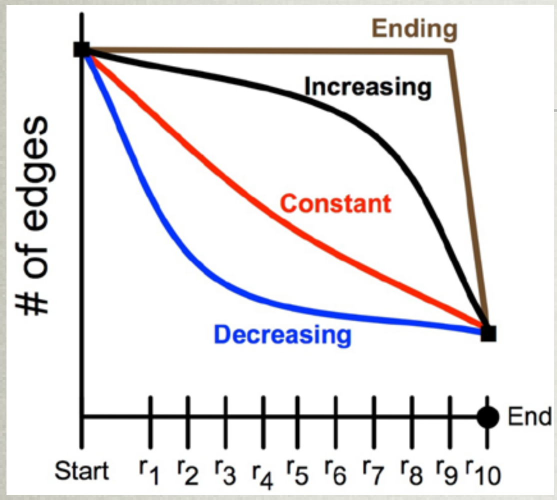
stain & extract D1 column

Machine learning algorithms to count synapses

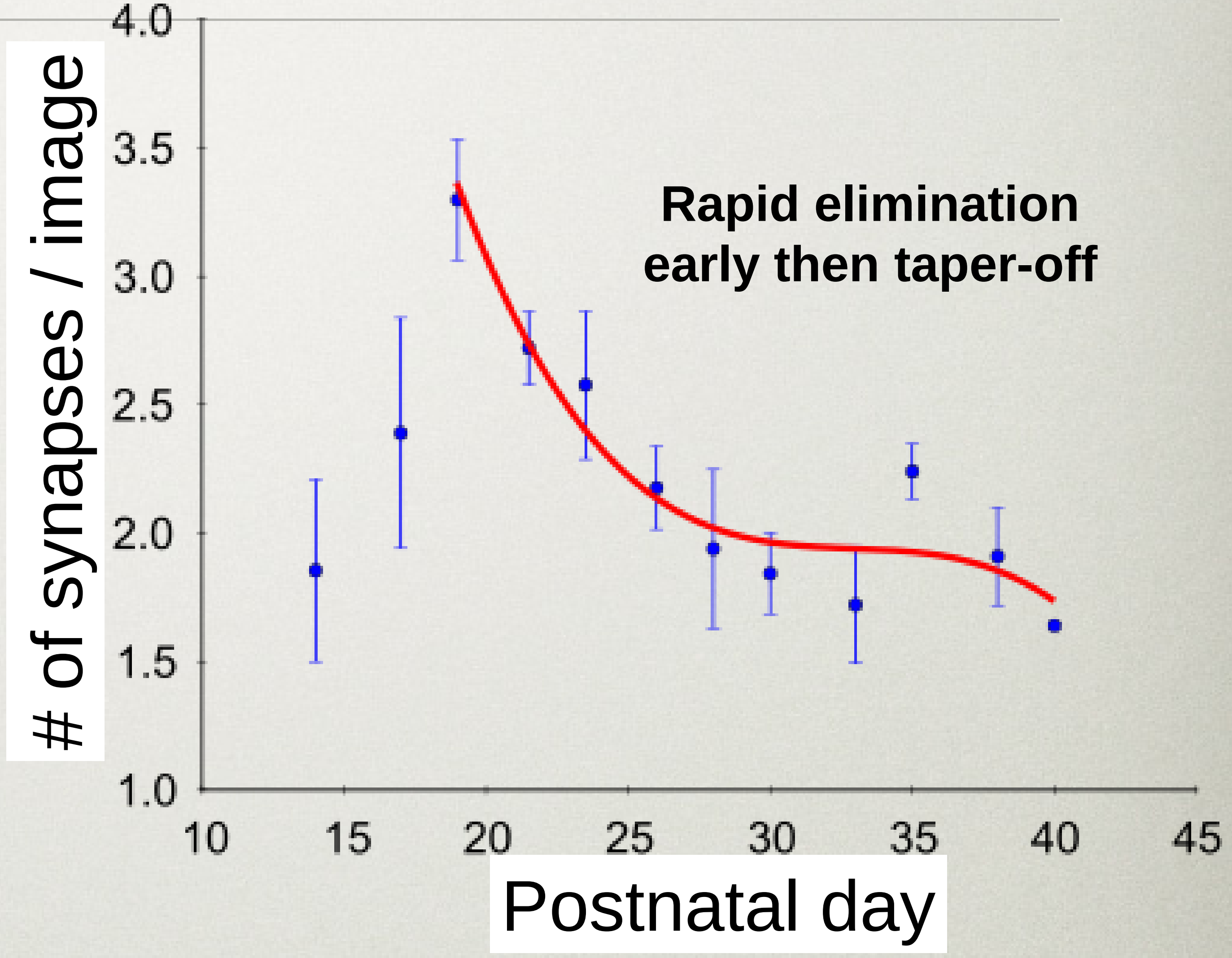
[Navlakha et al., ISMB 2013]



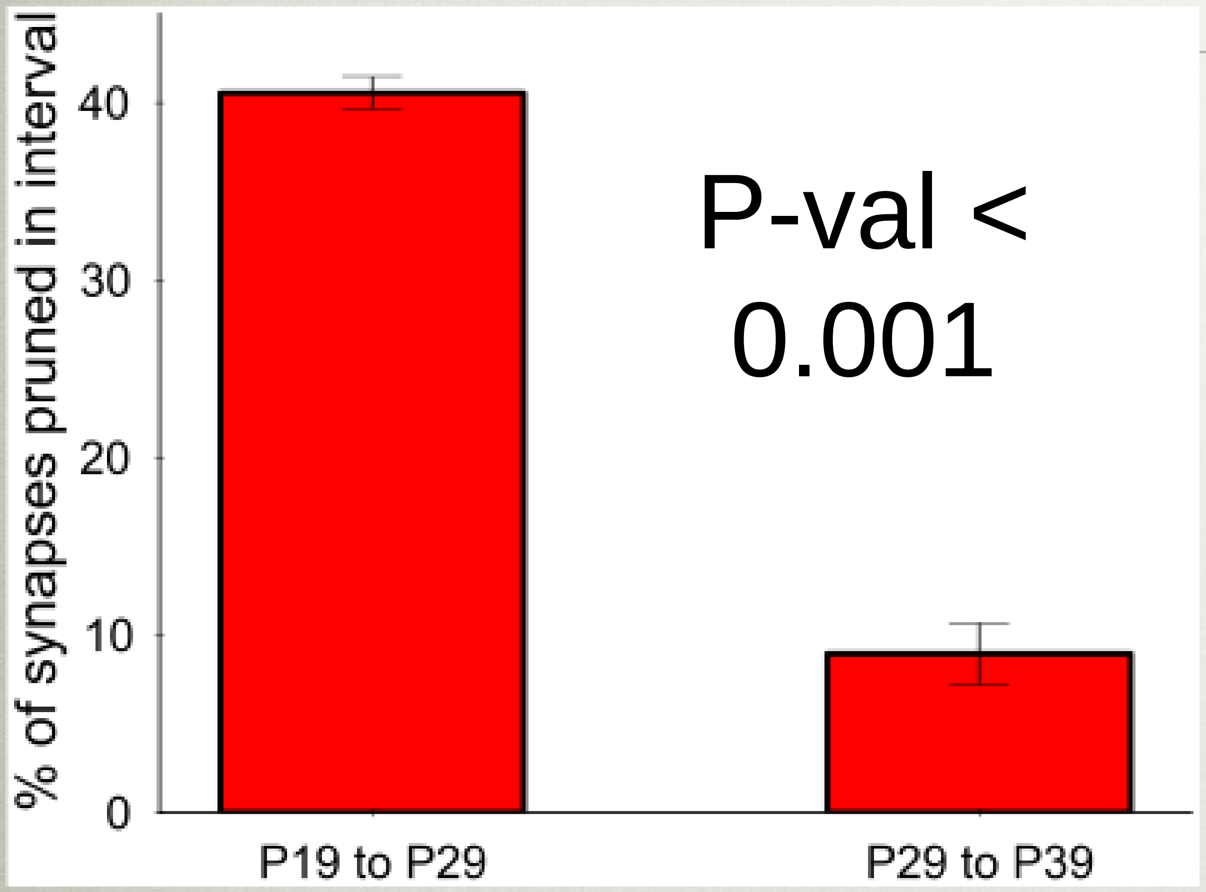
Pruning rates in the cortex

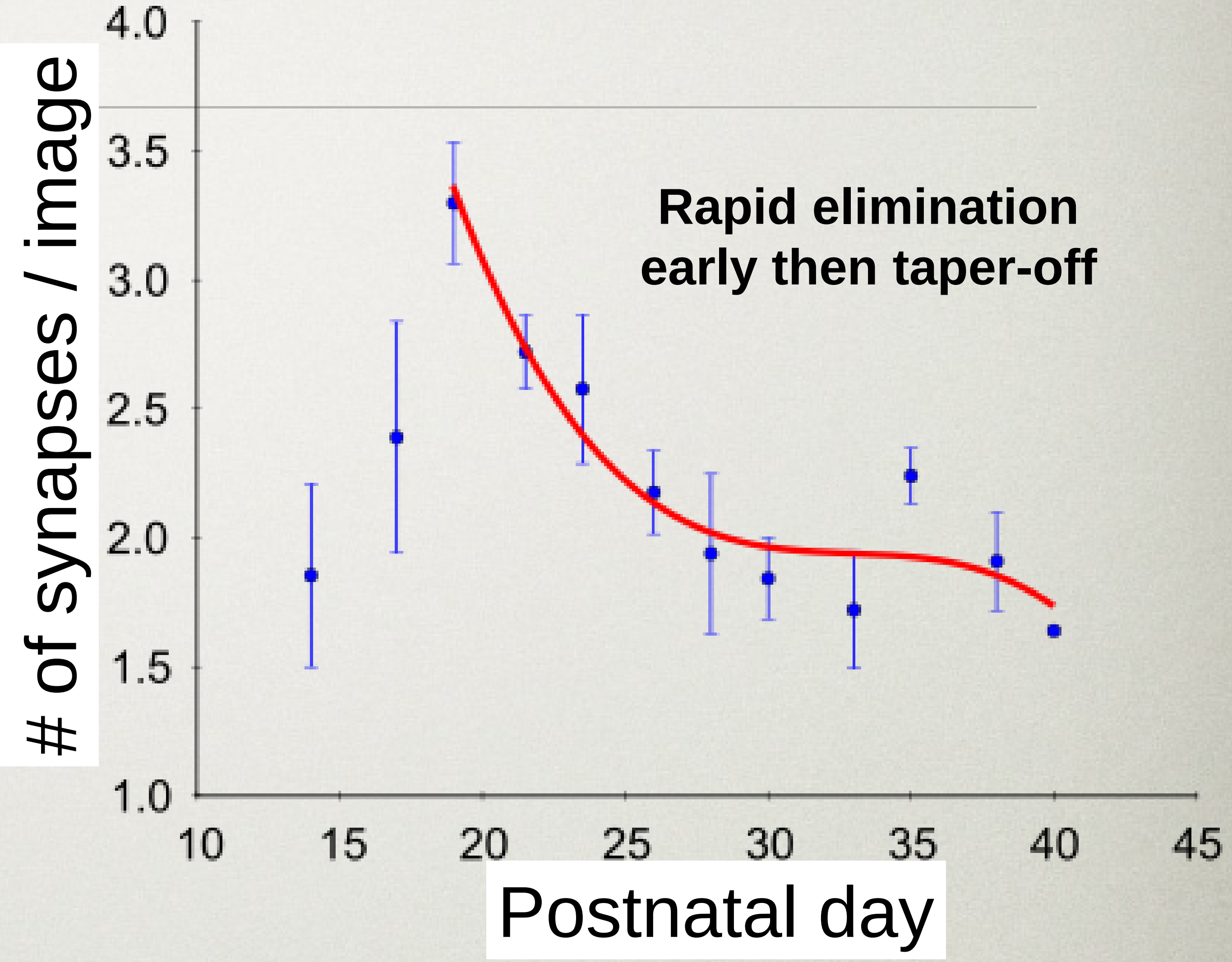


16 time-points
41 animals
9754 images
42709 synapses



Pruning rates are decreasing



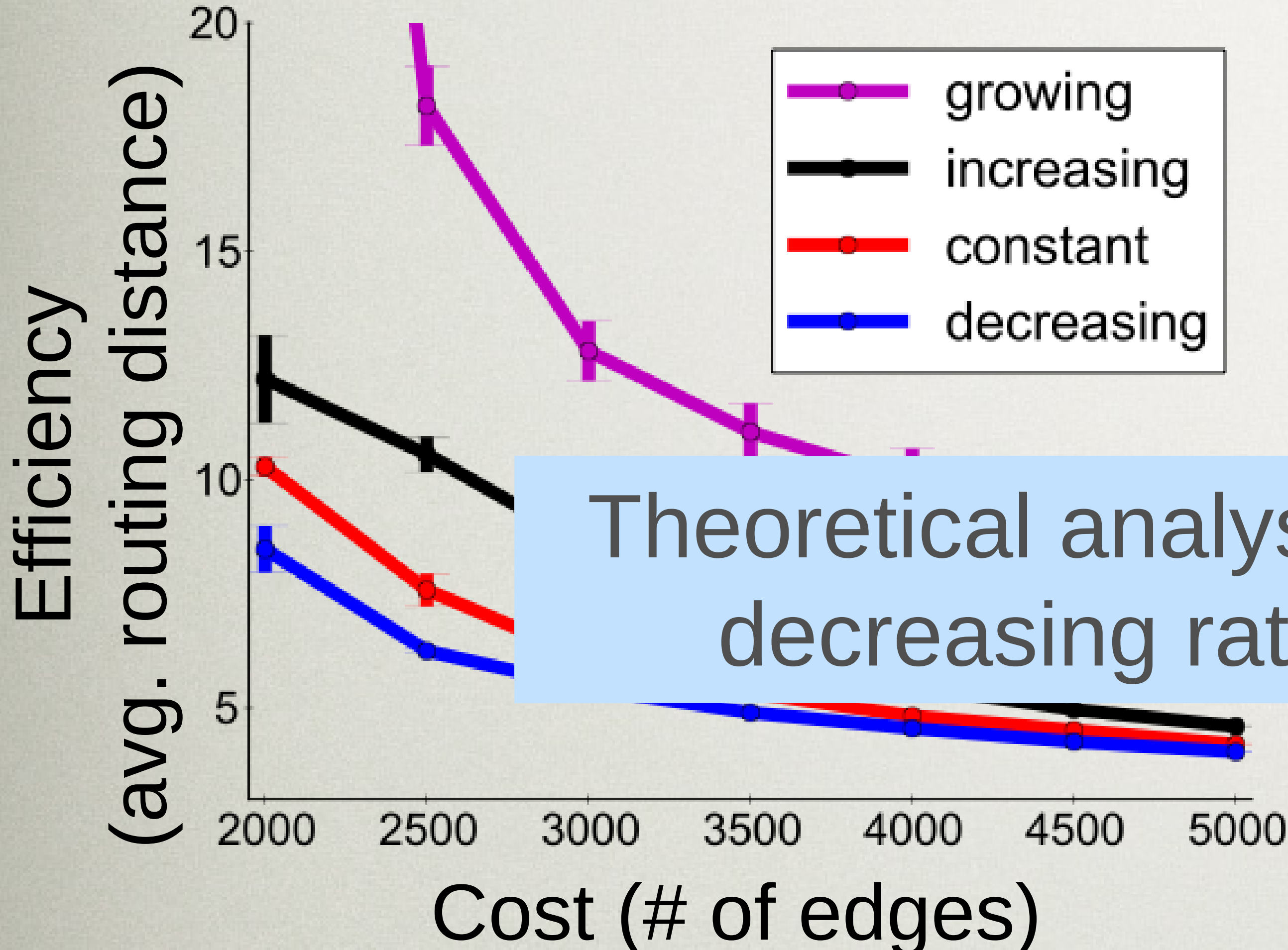


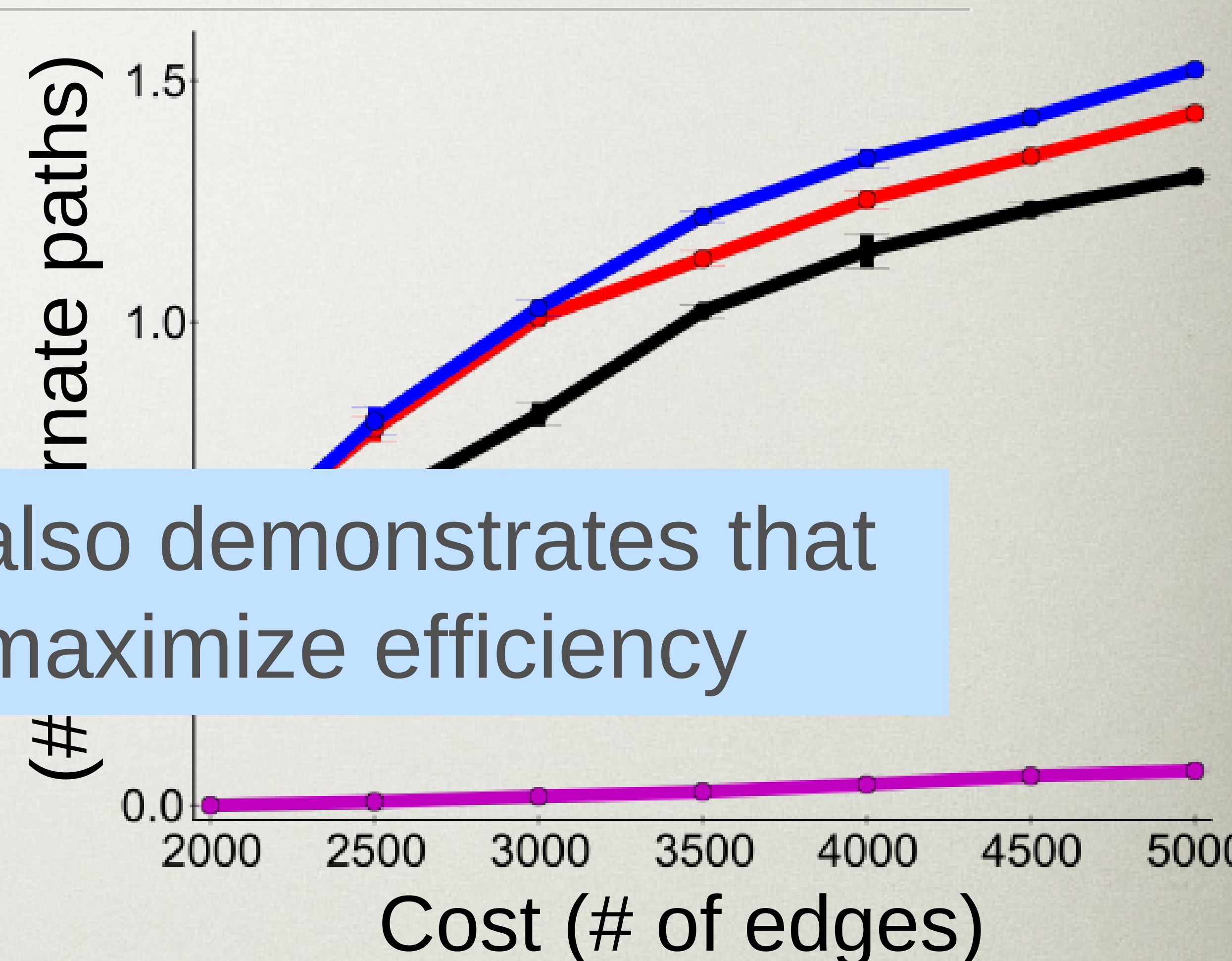
- Decreasing rate remove aggressively at the beginning

But ….

- The process is distributed
- Provides more time for the network to stabilize
- More cost effective

Decreasing rates further optimize network function





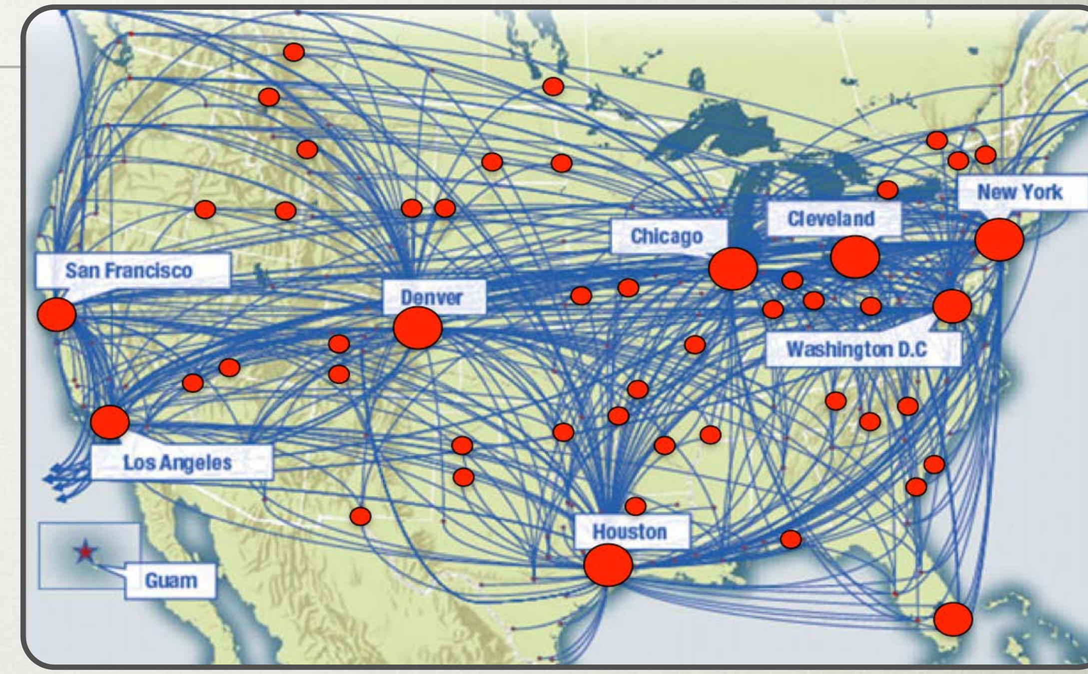
Theoretical analysis also demonstrates that decreasing rates maximize efficiency

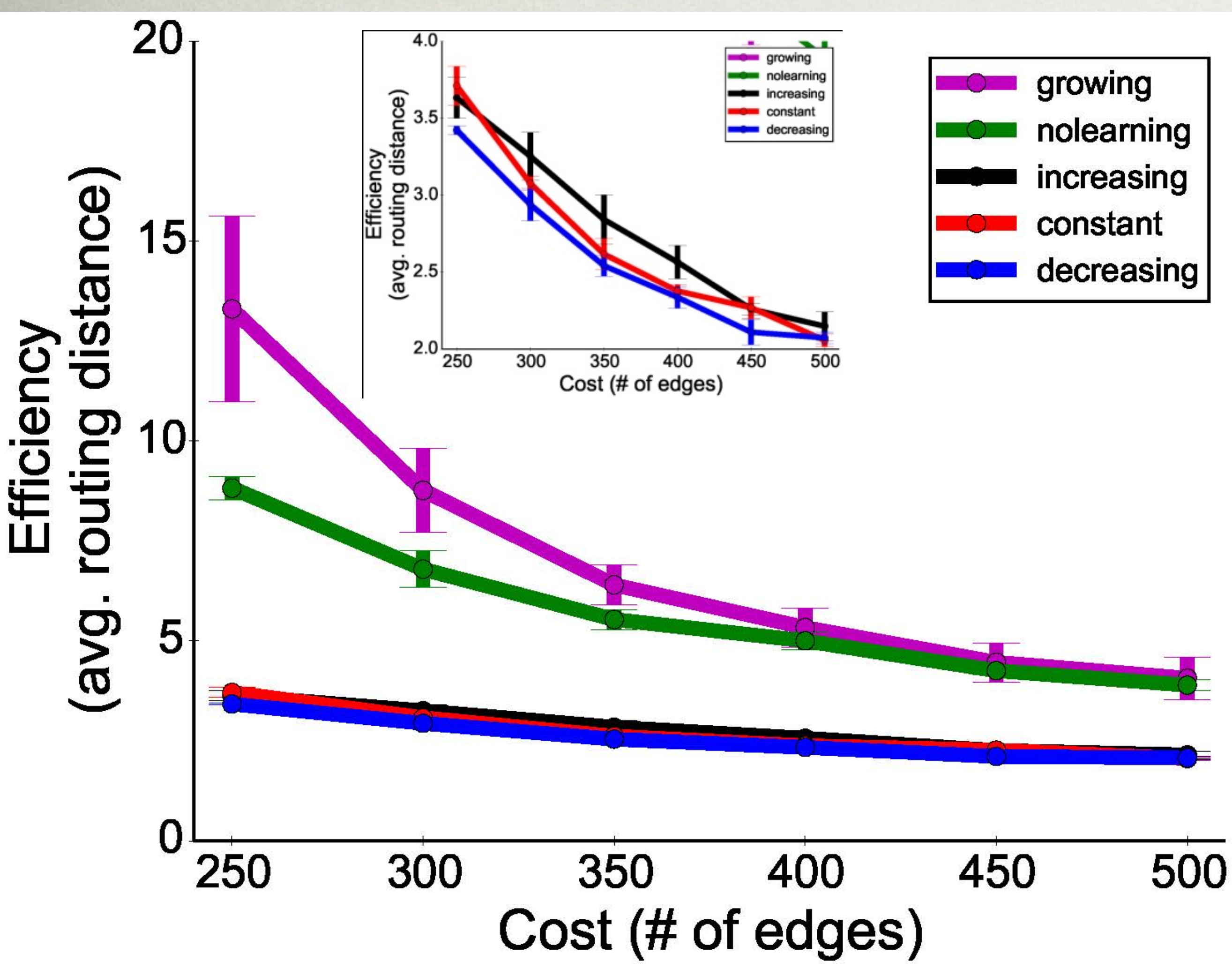
Decreasing rates 30% more efficient than increasing (20% > constant)

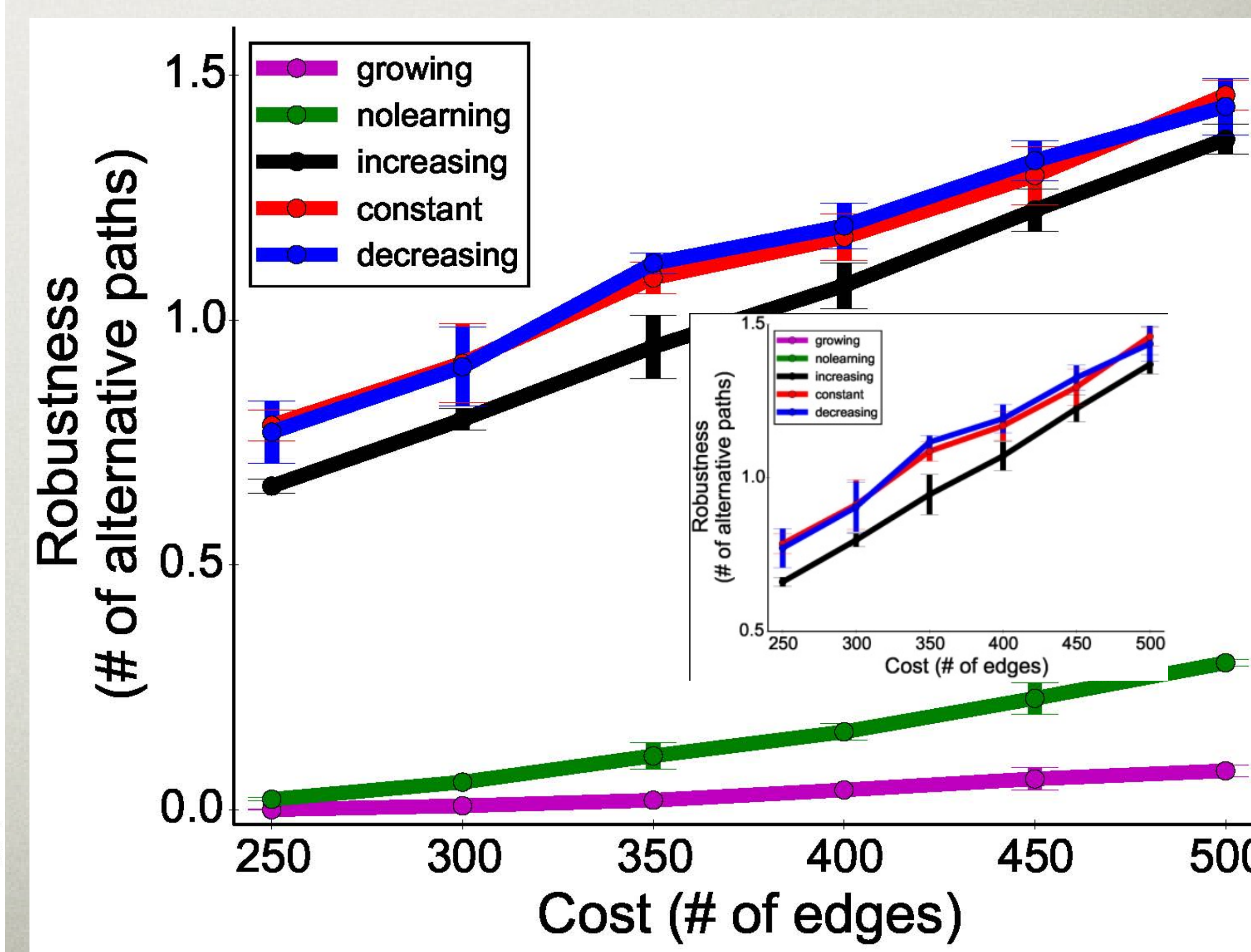
Slightly better fault tolerance

Application to routing airline passengers

- Use start / end city as source / target
  - > 800,000 trips between 122 cities covering 3 months of domestic US travel.
  - Assuming equal cost for each segment.







Conclusions

Reproduced a 60-year-old EM technique to selectively stain synapses coupled with *high-throughput and fully automated* analysis

* Feasible for large or small labs; no specialized transgenics required

Studied changes in synapse density + strength in the developing cortex

* May enable screening of pharmacologically-induced or plasticity-related changes in synapse density and morphology in the brain

Semi-supervised learning can be used to build robust classifiers using unlabeled data, which is often plentiful in bioimaging problems.