Image restoration and segmentation by convolutional networks

Sebastian Seung

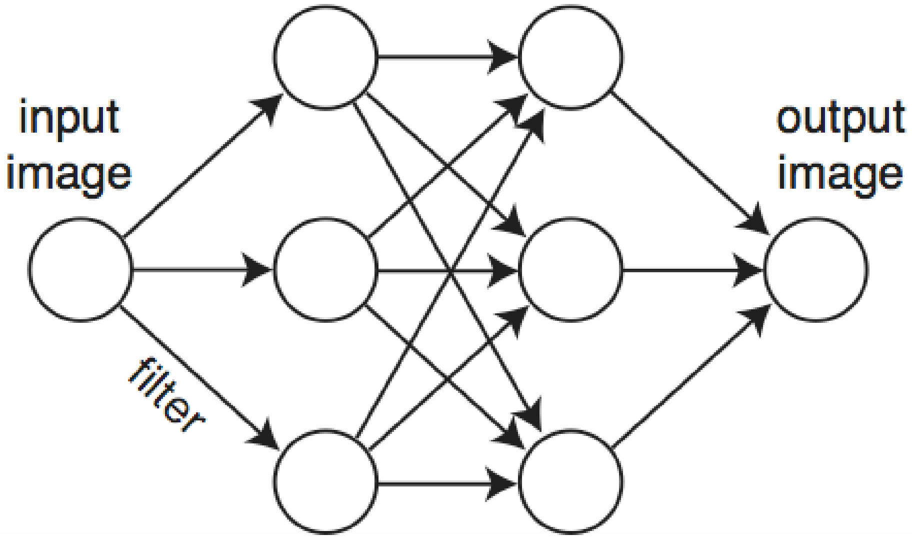
Howard Hughes Medical Institute
and MIT

Outline

- Convolutional networks
- Connectomics
- Binary image restoration
- Markov random fields
- Image segmentation
- Lessons

Convolutional network

- Defined with a directed graph



- node ↔ image, edge ↔ filter

Linear and nonlinear computations

- At edge ab
  - convolution by wab
- At node a
  - addition of results
  - nonlinear activation function

$$I_a = f\left(\sum_b w_{ab} * I_b - \theta_a\right)$$

Relation to neural networks

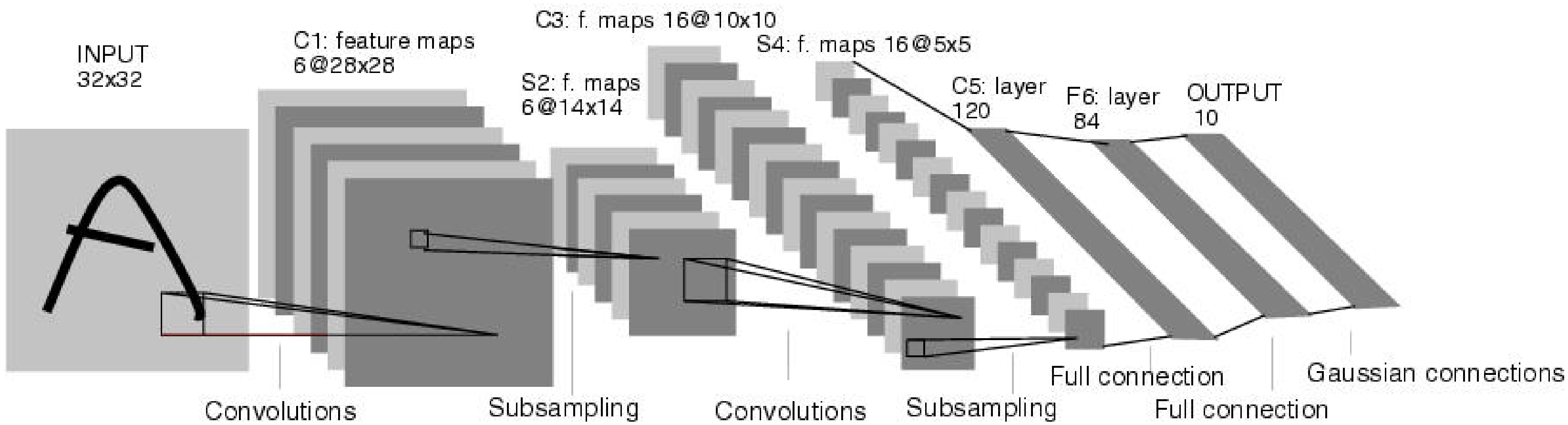
- Can be viewed either as a generalization or as a specialization.
- Gradient learning can be done via backpropagation.

Properties suited for low-level image processing

- Translation invariance
  - inherited from the convolution operation
- Locality
  - filters are typically small

Visual object recognition

- handprinted characters
  - LeCun, Bottou, Bengio, Haffner (1998)
- objects
  - LeCun, Huang, Bottou (2004)

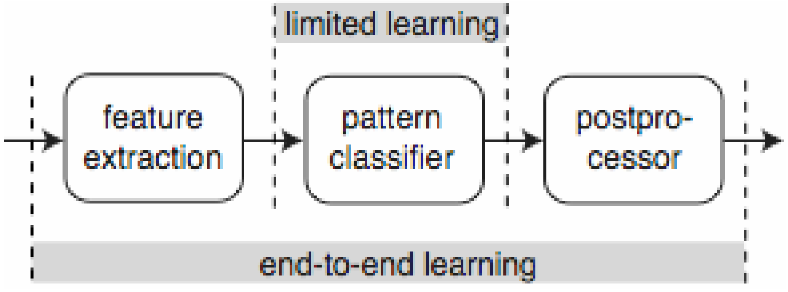


High-level vs. low-level

- High-level vision
  - convolution alternates with subsampling
- Low-level vision
  - no subsampling
  - possibly supersampling

Learning image processing

- Based on hand-designed features
  - Martin, Fowlkes, and Malik (2004)
  - Dollar, Tu, Belongie (2006)
- End-to-end learning



Neural networks for image processing

- reviewed by Egmont-Petersen, de Ridder, and Handels (2002)
- active field in the 80s and 90s
- ignored by the computer vision community
- convolutional structure is novel

Outline

- Convolutional networks
- Connectomics
- Binary image restoration
- Markov random fields
- Image segmentation
- Lessons

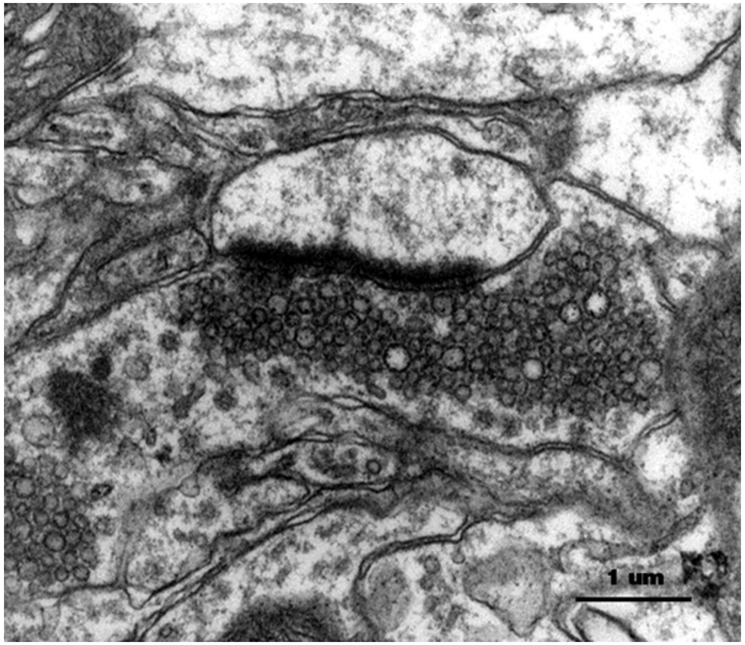
SBF-SEM

QuickTime™ and a
TIFF decompressor
are needed to see this picture.

- Denk & Horstmann, *PLOS Biol.* (2004).
- Briggman & Denk, *Curr. Opin. Neuro.* (2006).

The two problems of connectomics

- Recognize synapses
- Trace neurites back to their sources



Anna Klintsova

What is connectomics?

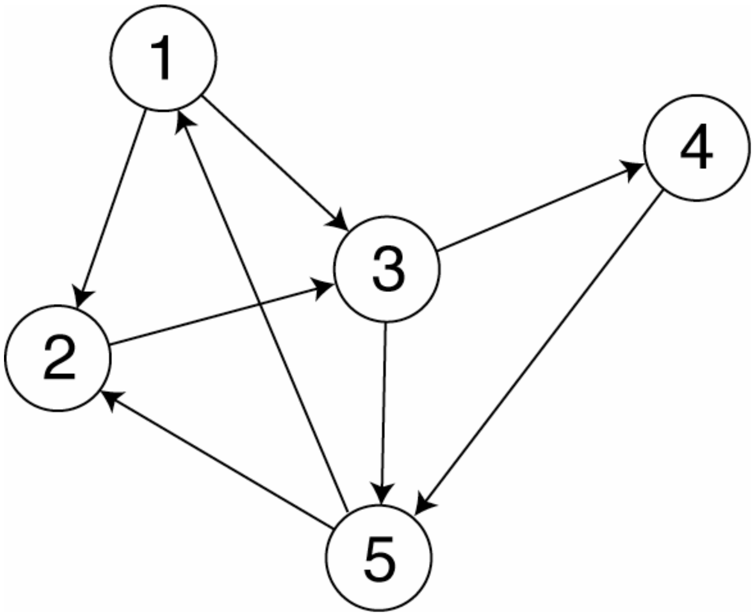
- High-throughput generation of data about neural connectivity
  - data-driven
- Mining of connectivity data to obtain knowledge about the brain
  - hypothesis-driven

Nanoscale imaging and cutting

- Axons and spine necks can be 100 nm in diameter.
- xy resolution: electron microscopy
  - Transmission EM (TEM)
  - Scanning EM (SEM)
- z resolution: cutting

C. elegans connectome

- list of 300 neurons
- 7000 synapses
- 10-20 years to find
- not high-throughput!



Near future: teravoxel datsets

- one cubic millimeter
- entire brains of small animals
- small brain areas of large animals
- speed and accuracy are both challenges

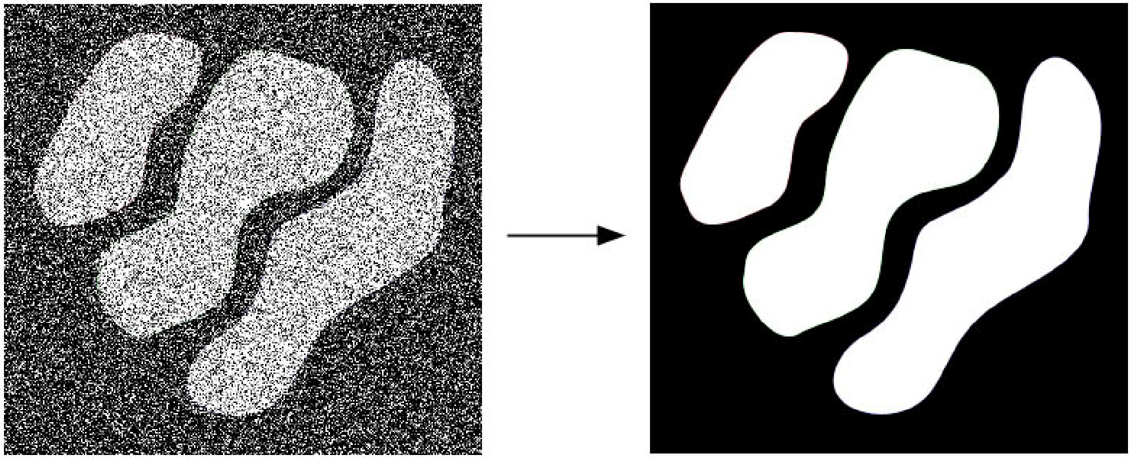
QuickTime™ and a
YUV420 codec decompressor
are needed to see this picture.

Outline

- Convolutional networks
- Connectomics
- Binary image restoration
- Markov random fields
- Image segmentation
- Lessons

Binary image restoration

- Map each voxel to “in” or “out”



Training and test sets

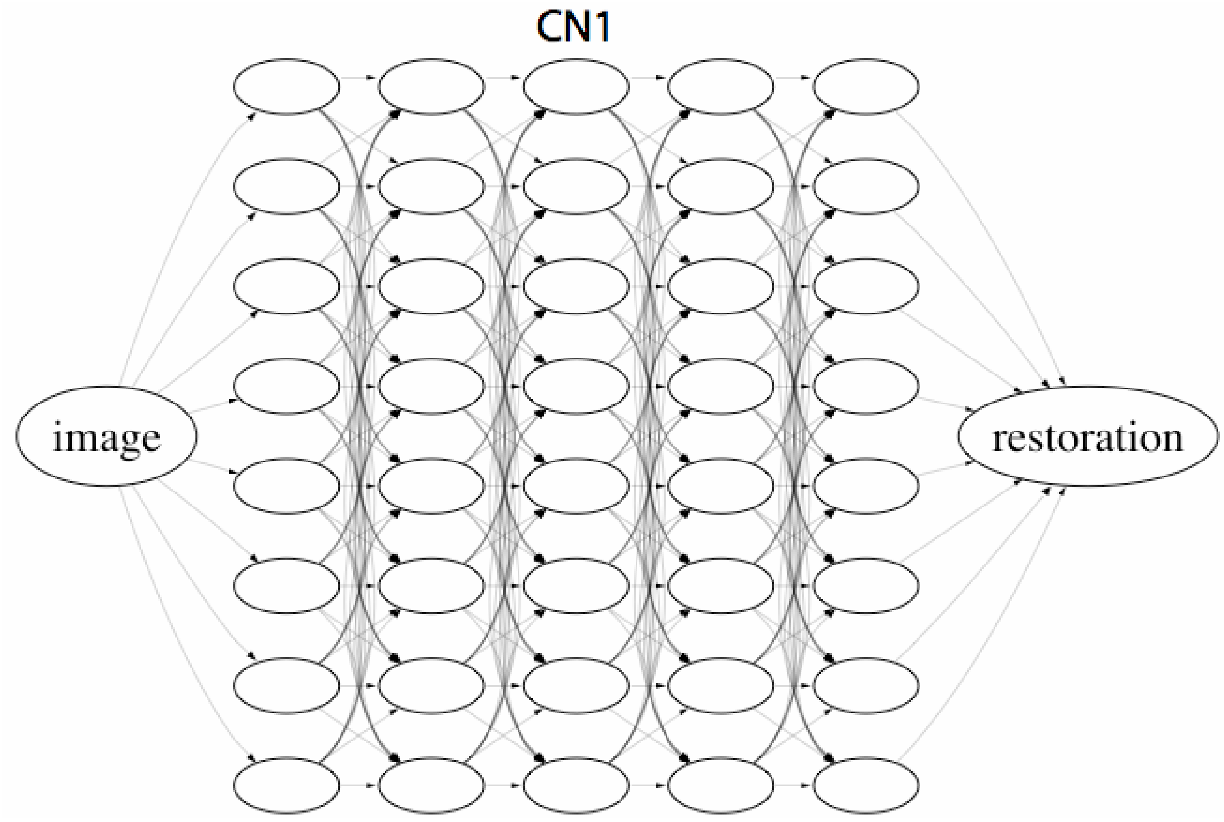
- rabbit retina (outer plexiform layer)
- $800 \times 600 \times 100$ image at $26 \times 26 \times 50$ nm
- boundaries traced by two humans
  - disagreement on 9% of voxels
  - mostly subtle variations in boundary placement
- 0.5/1.3 megavoxel training/test split

Baseline performance

- Guessing “in” all the time: 25% error
- Simple thresholding
  - training error 14%
  - test error 19%
- Thresholding after smoothing by anisotropic diffusion
  - not significantly better

CN1: a complex network

- 5 hidden layers, each containing 8 images



Gradient learning

- each edge: $5 \times 5 \times 5$ filters
- each node: bias
- 35,041 adjustable parameters
- cross-entropy loss function
- gradient calculation by backpropagation

QuickTime™ and a
YUV420 codec decompressor
are needed to see this picture.

CN1 halves the error rate of simple thresholding

- The test error is about the same as the disagreement between two humans.
- The training error is less.

Outline

- Convolutional networks
- Connectomics
- Binary image restoration
- Markov random fields
- Image segmentation
- Lessons

Model of image generation

- Clean image x is drawn at random
  - Image prior p(x)
- and corrupted to yield noisy image y
  - Noise model p(y|x)
- restoration by MAP inference

$$\arg\max_{x} p(x|y)$$

What image prior?

- Intuition
  - Geman and Geman (1984)
- Unsupervised learning
  - Examples of noisy images only
  - Roth and Black (2005)
- Supervised learning
  - Examples of noisy and clean images

Markov random field

- Prior for binary images

$$p(x) \propto \exp\left(\frac{1}{2} \sum_i x_i (w * x)_i + \sum_i b x_i \right)$$

- Translation-invariant interactions
  – filter w
  – external field b

MRF learning

- maximum likelihood
  - Boltzmann machine
  - MCMC sampling
- maximum pseudolikelihood $p(x_i \mid x_{-i})$
  - Besag (1977)

MRF inference

- maximize the posterior

$$p(x \mid y) \propto \exp\left(\frac{1}{2} \sum_i x_i (w * x)_i + \sum_i b_i x_i \right)$$

- simulated annealing
- min-cut algorithms
  - polynomial time for nonnegative w
  - Greig, Porteous, and Seheult (1989)
  - Boykov and Kolmogorov (2004)

MRF performance is similar to thresholding

- Pseudolikelihood might be a bad approximation to maximum likelihood
- Min-cut inference might not perform MAP, if the weights are of mixed sign.
- Maximizing p(x,y) might be misguided

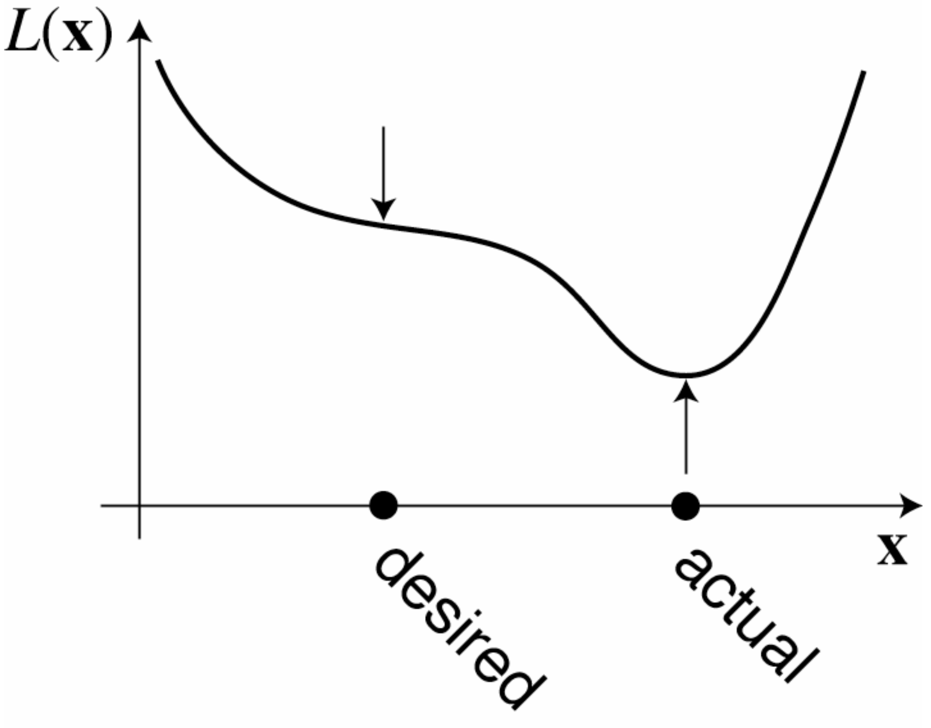
Conditional random field

- Learn by maximizing the posterior
- Pseudolikelihood was really bad
- Zero temperature Boltzmann learning
  - min-cut for inference
  - contrastive update

$$\Delta w_j \propto \left\langle \sum_i x_{i+j} x_i \right\rangle_0 - \left\langle \sum_i x_{i+j} x_i \right\rangle_\infty$$

  - constraint w to be nonnegative

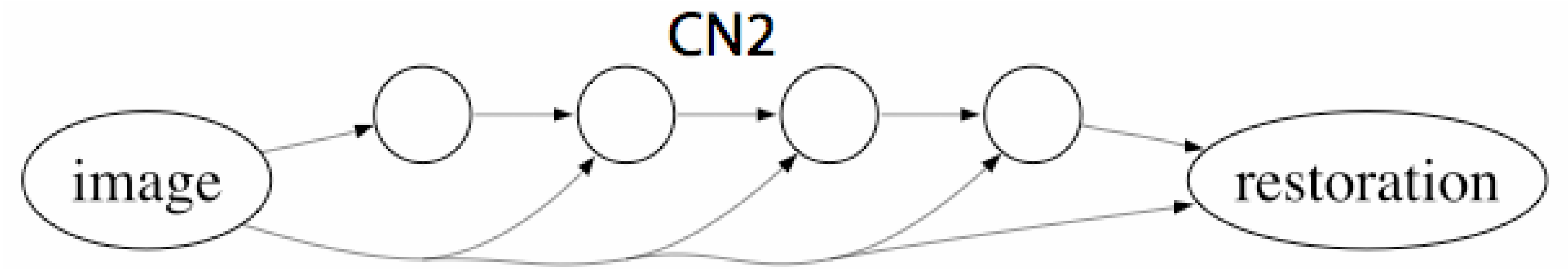
Contrastive Hebbian learning



CRF performance is similar to thresholding

- Perhaps the CRF cannot represent a powerful enough computation.
- To test this hypothesis, try a convolutional network with a simple architecture.

CN2: simple network



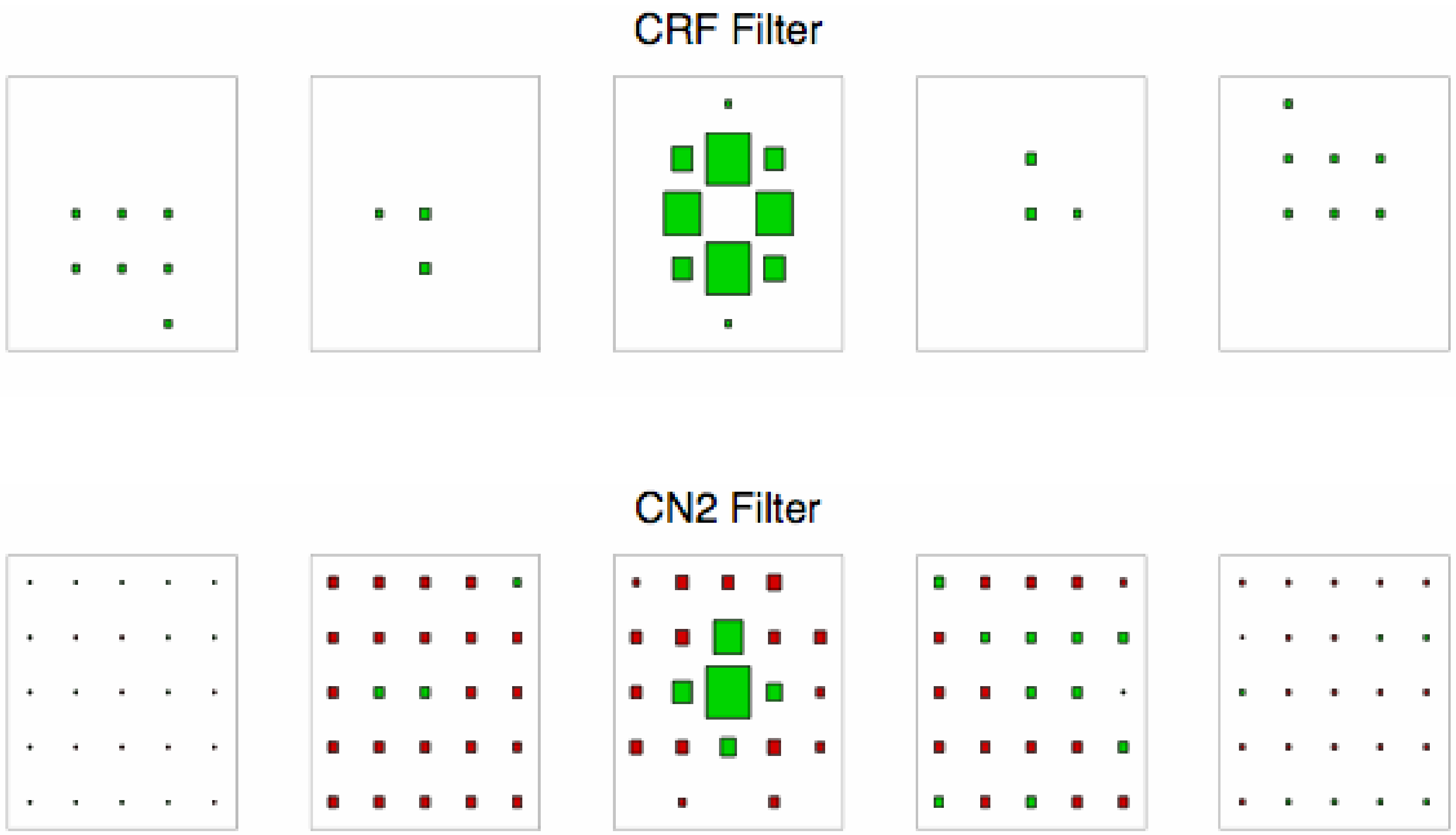
- Mean field inference for the CRF

$$\mu_i = \tanh\left((w * \mu)_i + y_i + b\right)$$

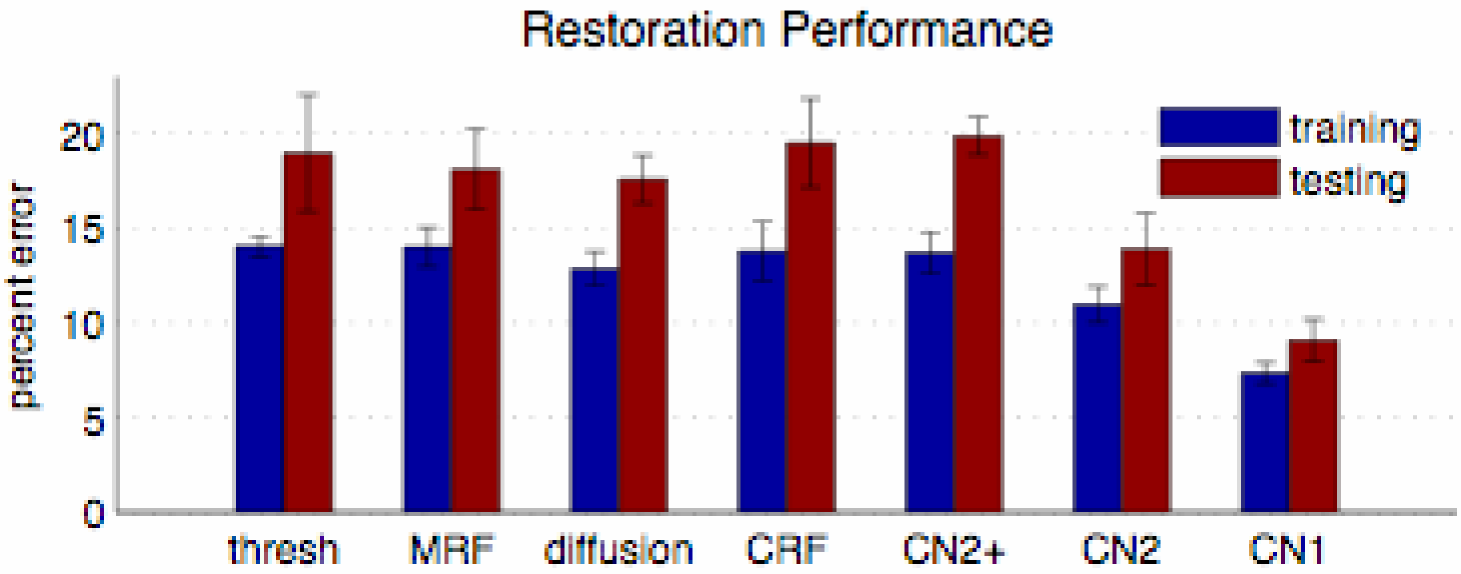
Nonnegativity constraints hurt performance

- CN2+ performed the same as the CRF and thresholding.
- CN2 performed better than thresholding, but not as well as CN1

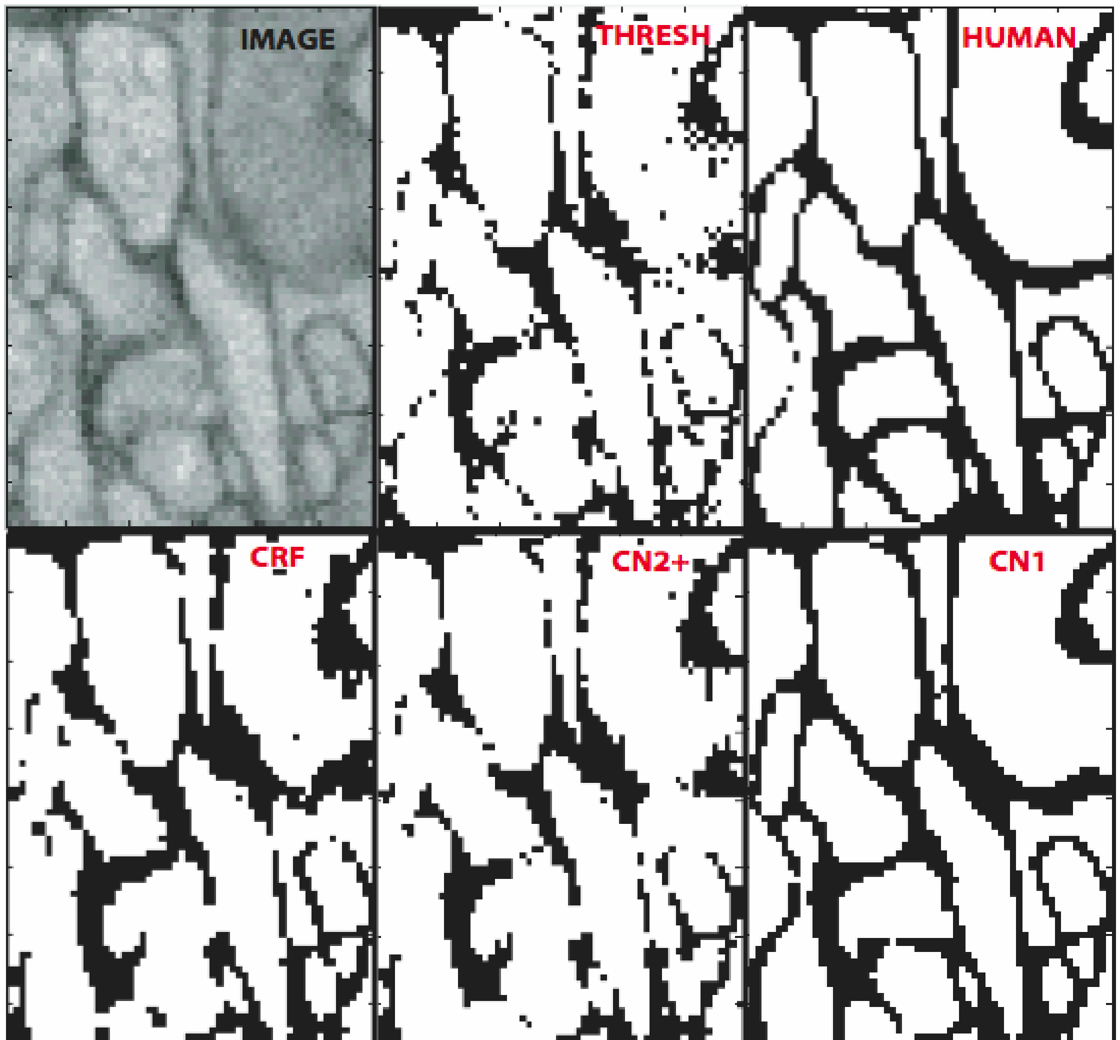
Filter comparison



Comparison of restoration performance



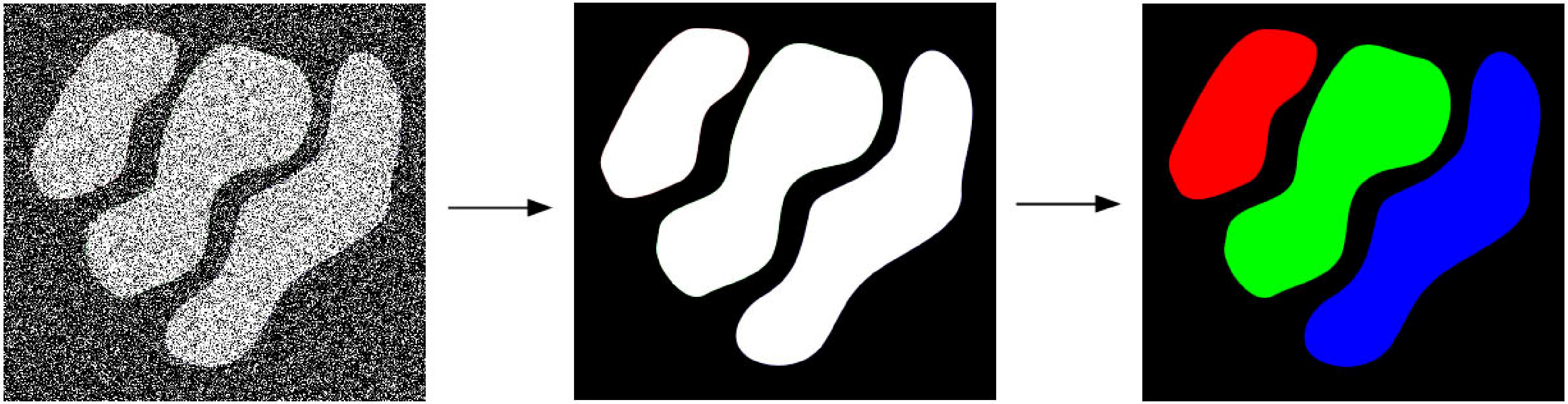
Restored images



Outline

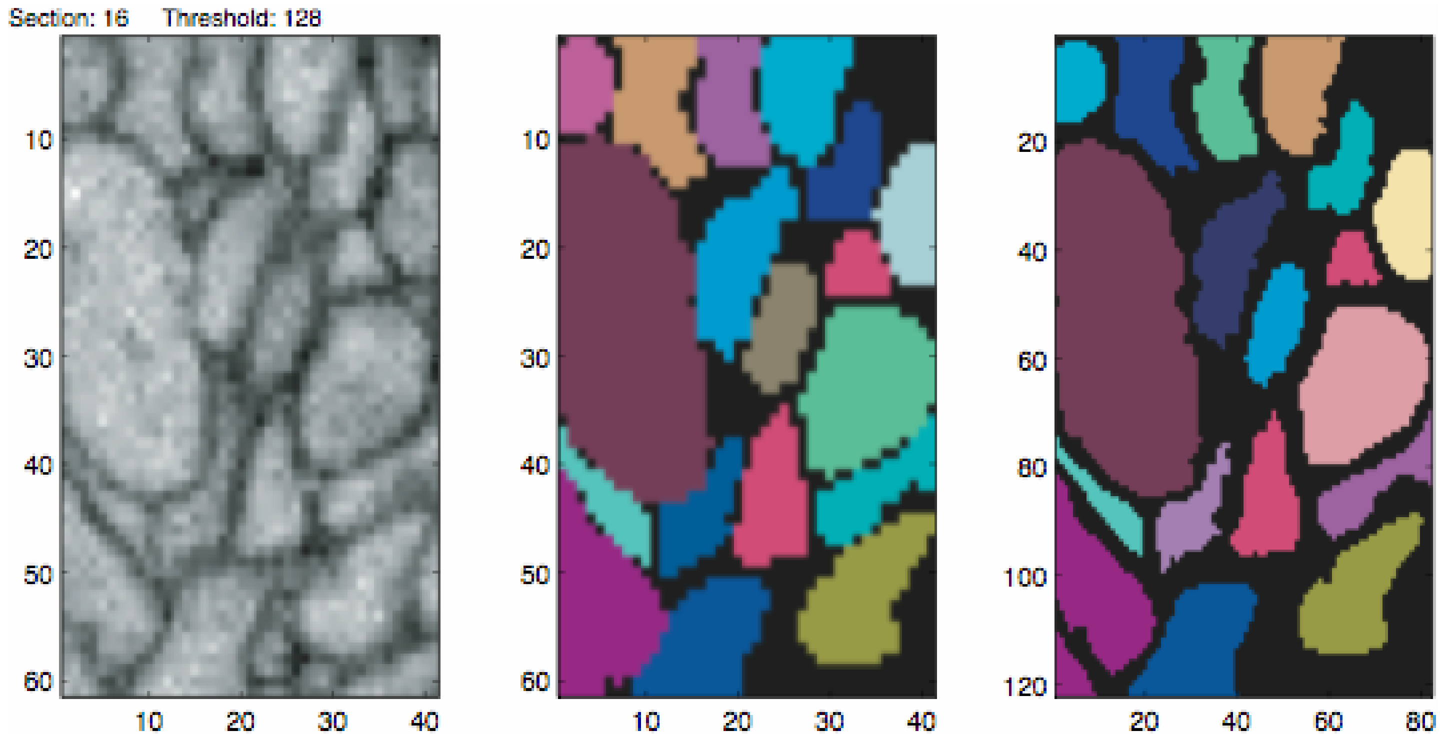
- Convolutional networks
- Connectomics
- Binary image restoration
- Markov random fields
- Image segmentation
- Lessons

Image restoration and segmentation

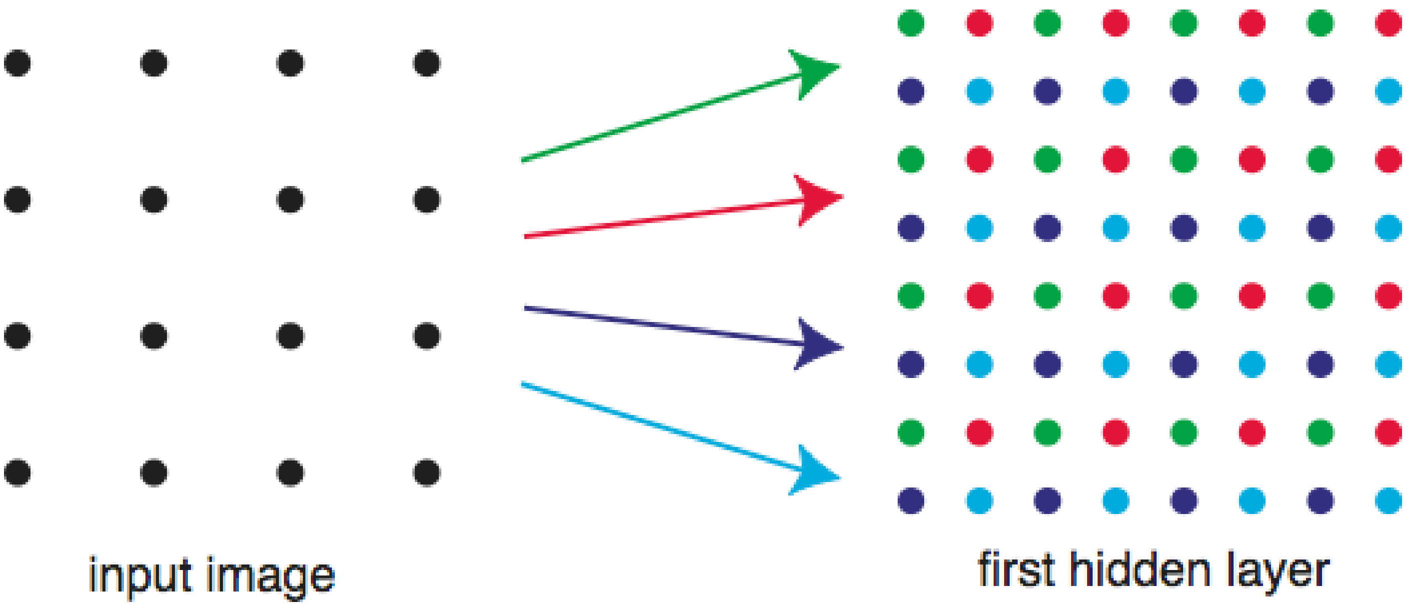


A problem due to inadequate image resolution

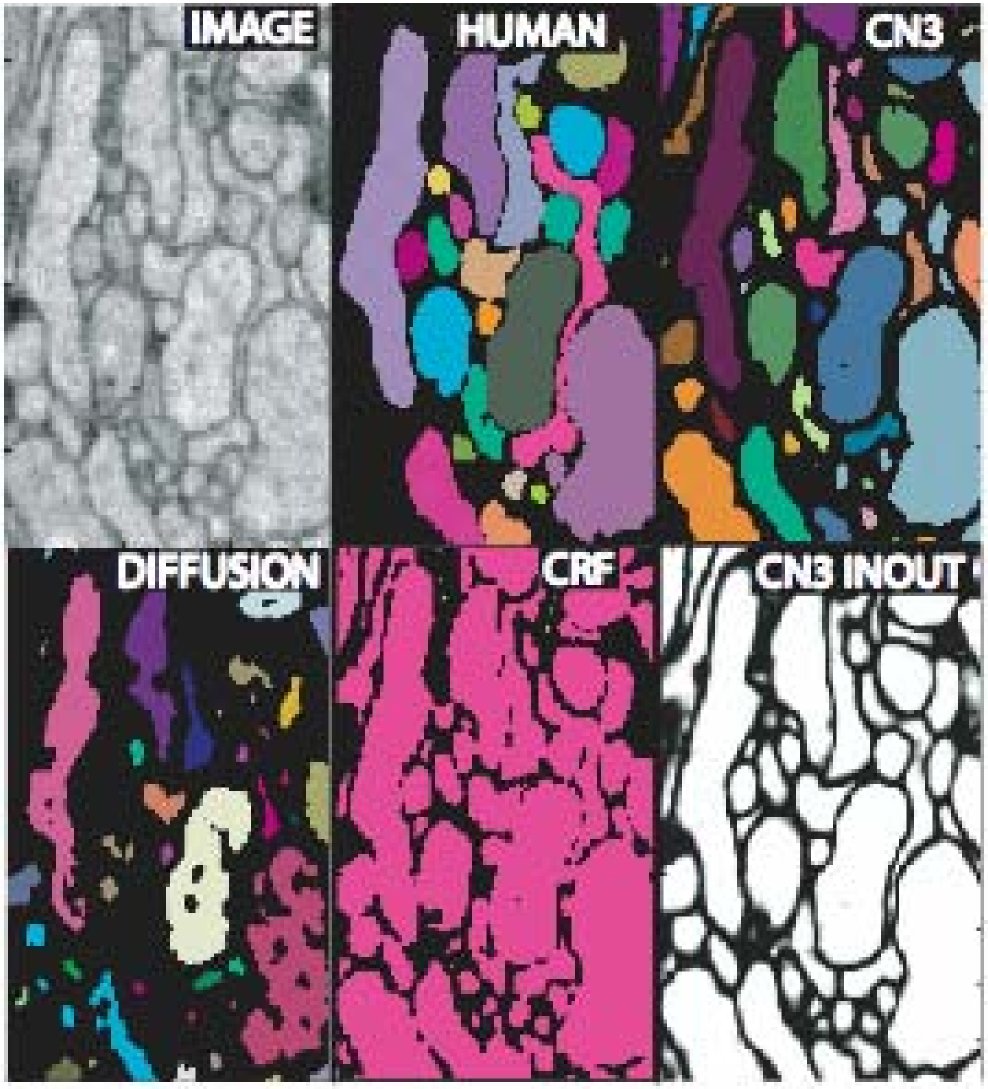
- Two objects (“in” regions) may touch.
- Not separated by an (“out” boundary).



Supersampling



Segmented images



Outline

- Convolutional networks
- Connectomics
- Binary image restoration
- Markov random fields
- Image segmentation
- Lessons

The cost of convexity is representational power.

- MAP inference for an CRF with nonnegative interactions is a convex optimization.
- The CRF was worse than CN2, and no better than thresholding.
- This was due to the nonnegativity constraint.

Bayesian methods have technical difficulties.

- MCMC sampling is slow
- Pseudolikelihood
  - trains the CRF to predict one output voxel from all the other output voxels.
  - This is evidently irrelevant for predicting the output from the input.
- Other approximations may have problems too.

Discriminative training may not be better.

- A discriminatively trained CRF was about the same as a generatively trained MRF.

Convolutional networks avoid Bayesian difficulties

- Their representational power is greater than or equal to that of MRFs.
- The gradient of the objective function for learning can be calculated exactly.
- Theoretical foundation is empirical error minimization.