Multiscale models for shapes and images

Pedro F Felzenszwalb
Brown University

Joint work with Joshua Schwartz, John Oberlin

Priors for Vision

- Markov models are widely used in computer vision
  - Natural model for curves and images
  - Tractable learning and inference
- Markov models capture local properties/regularities
  - often not enough
- Multiscale representations
  - Can capture local properties at multiple resolutions
  - Lead to rich models with a “low-dimensional” parameterization
  - (sometimes) Manageable computation

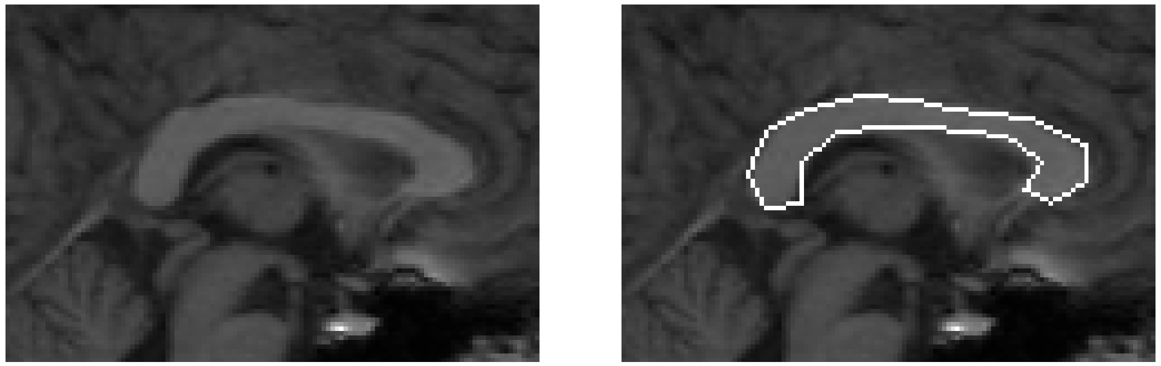
Bayesian Framework

- We observe y (curve, image)
- Hidden variables x (curve, image, class)
- Inference using Bayes rule
  - $p(x|y) \sim p(x)\, p(y|x)$
- Challenges
  - x, y are a high-dimensional objects (curve, image)
  - Efficient inference and learning

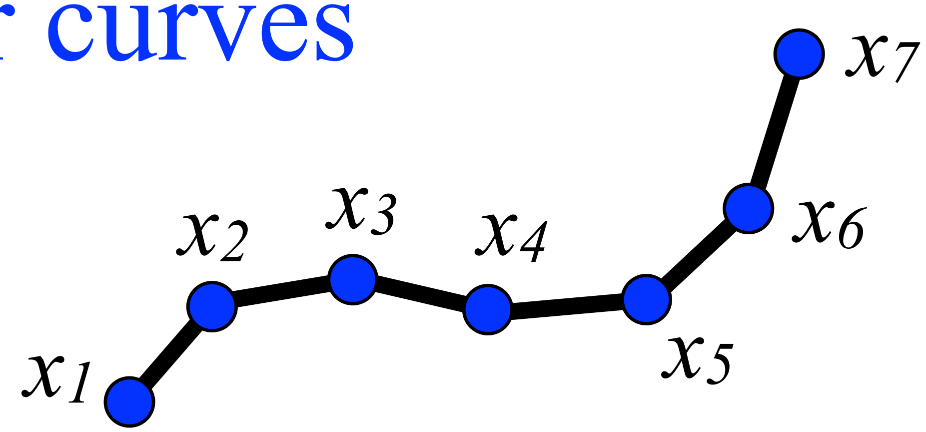
Shapes / Curves

- x = 2D curve
- classification
  - $p(x|c)$ for each class c
  - $p(c|x) \sim p(x|c)p(c)$

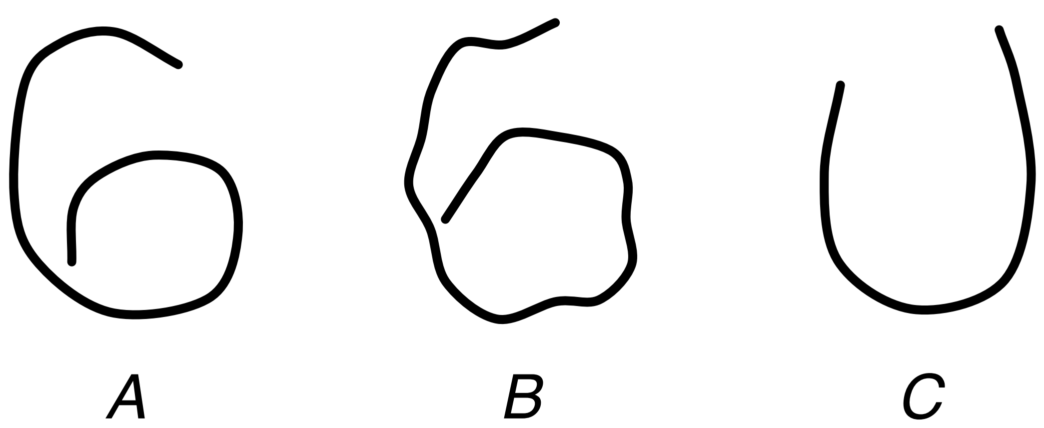
- localization/detection
  - image y
  - $p(x|y) \sim p(x)\, p(y|x)$



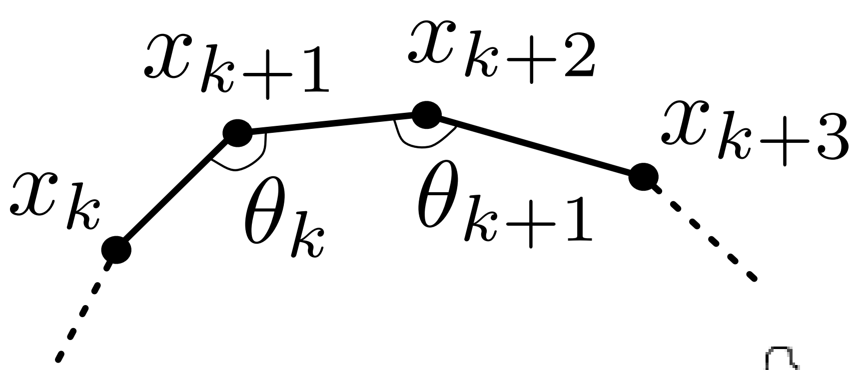
Markov models for curves

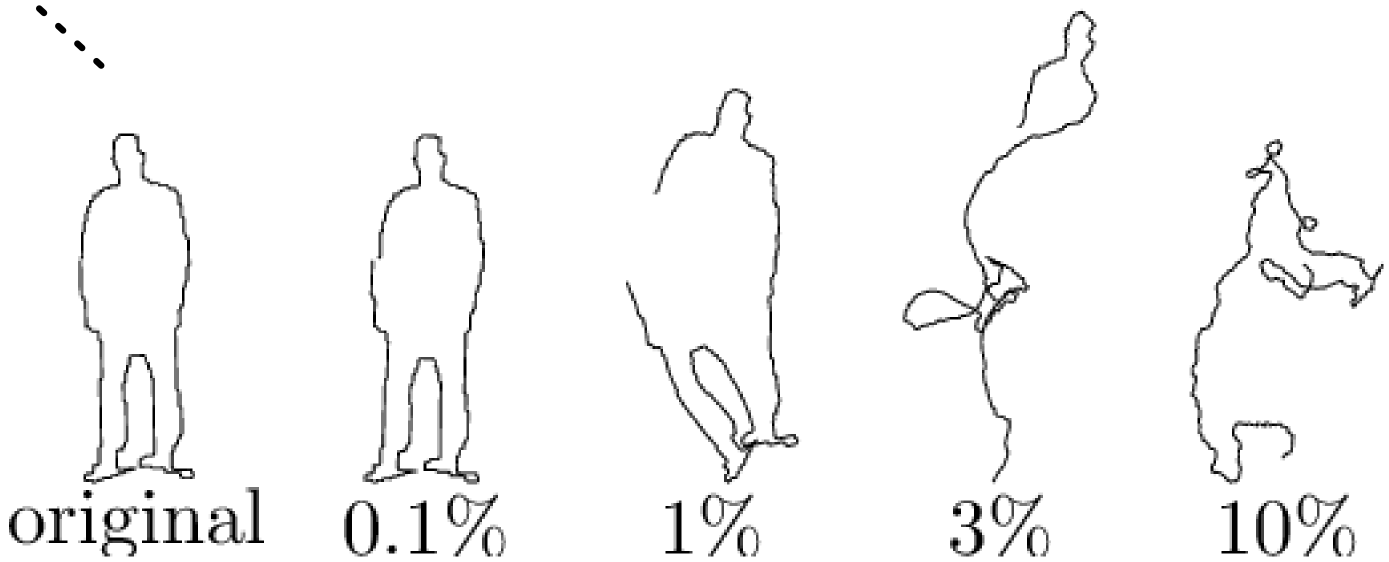


- Sequence of control points *x*
- Markov model captures local geometric properties
  - smooth, tends to curve to the left, etc.
- Often fails to capture important global geometric properties



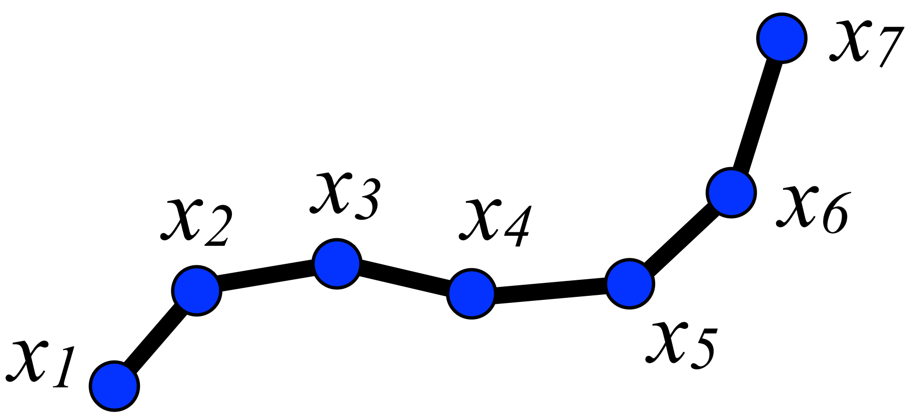
Random deformations with Markov model





- Small local changes lead to large global change
- Markov models suffer from drift
- Give up long range dependencies to allow for local variation

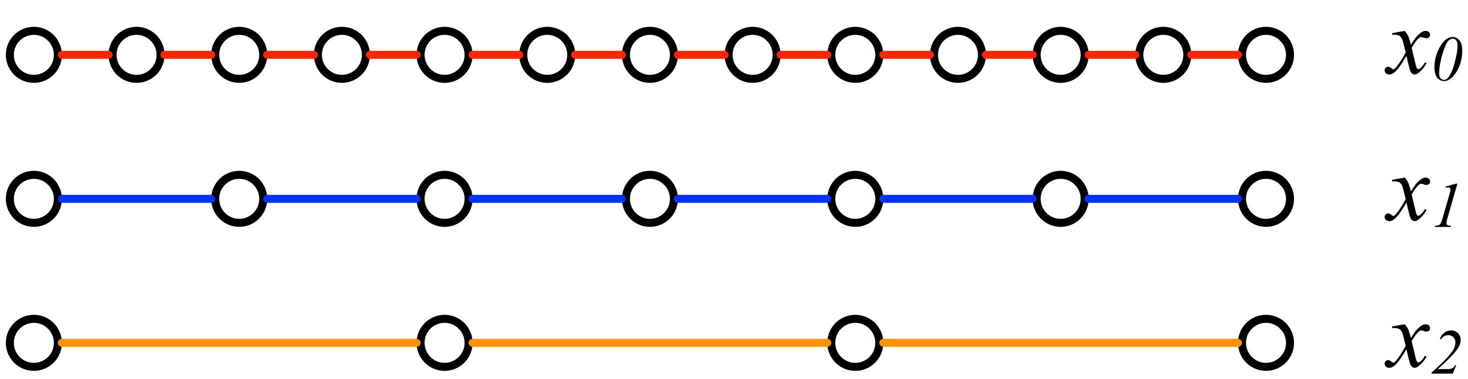
High order Markov models

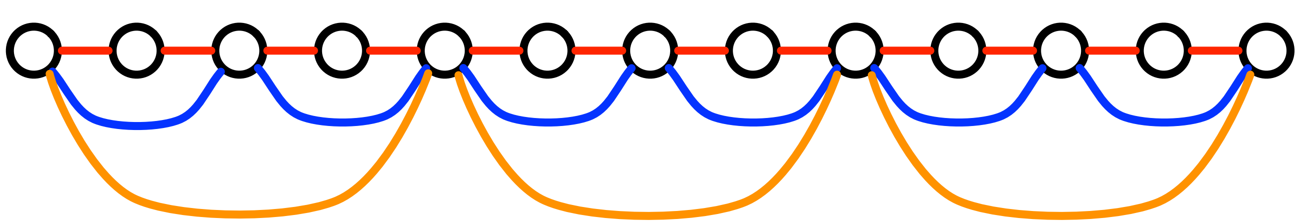


- k-th order model
  - $p(x_k|x_1,\ldots,x_{k-1})$
  - Number of parameters ~ $O(|X|^k)$
  - Complexity of inference ~ $O(|X|^k)$
  - Still suffers from drift, even with reasonably large k

Multiscale sequence model

- Capture local properties at multiple resolutions
  - Original sequence x_0
  - Subsample x_0 to get x_1, x_2 ...
  - local property of x_2 = non-local property of x_0

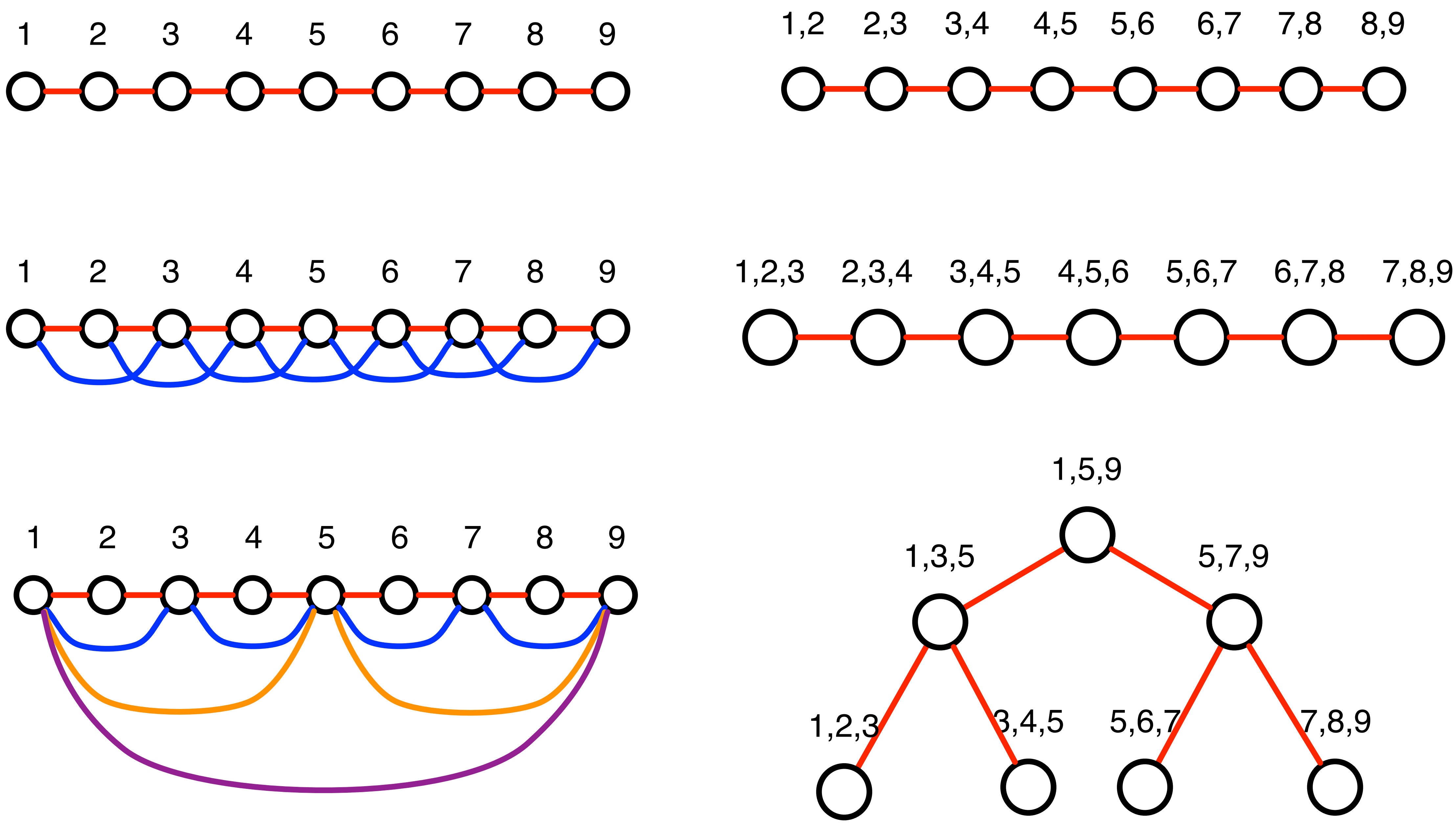




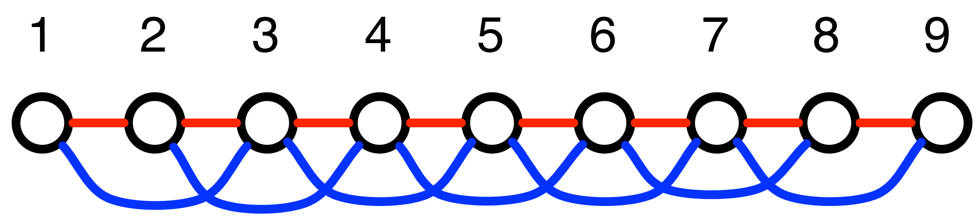
full model is tractable
tree-width = 2

Comparing sequence models

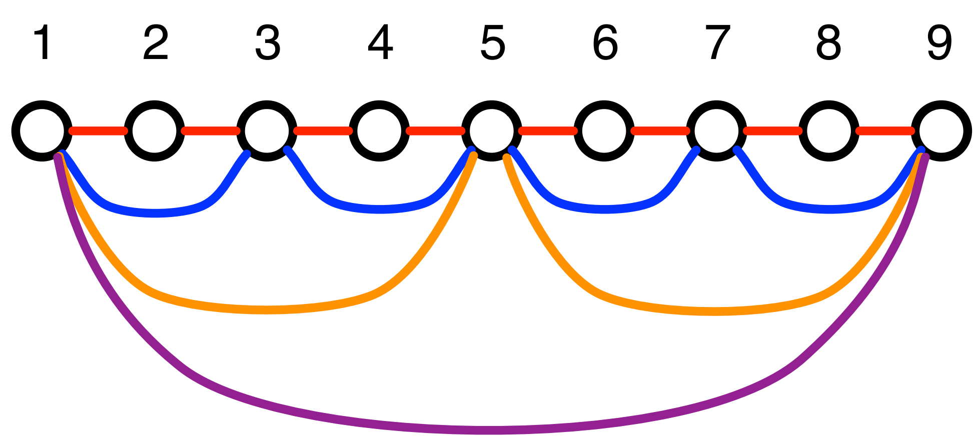
Graphical model (MRF) | Junction Tree

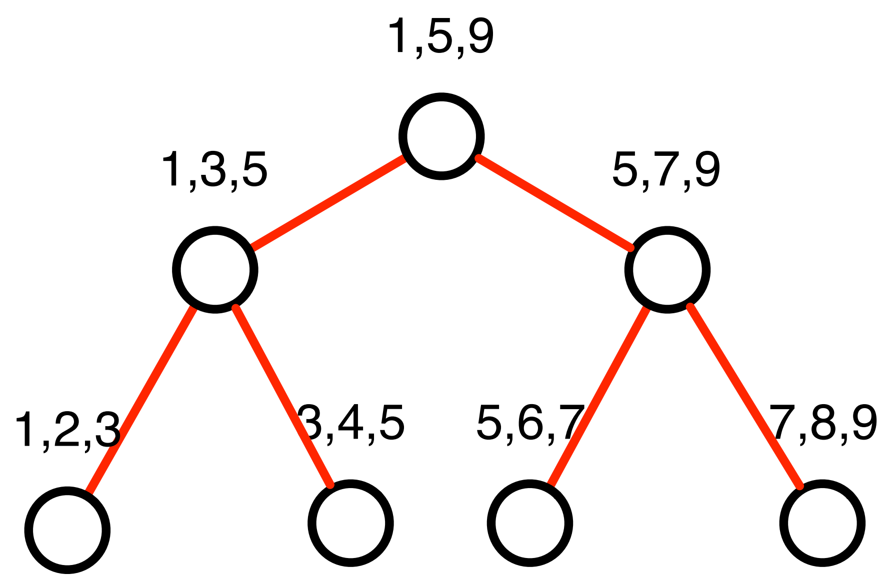


Factorizations



$$p(x_1, \ldots, x_9) = p(x_1, x_2) p(x_3|x_1, x_2) p(x_4|x_2, x_3) p(x_5|x_3, x_4) p(x_6|x_4, x_5) \ldots$$

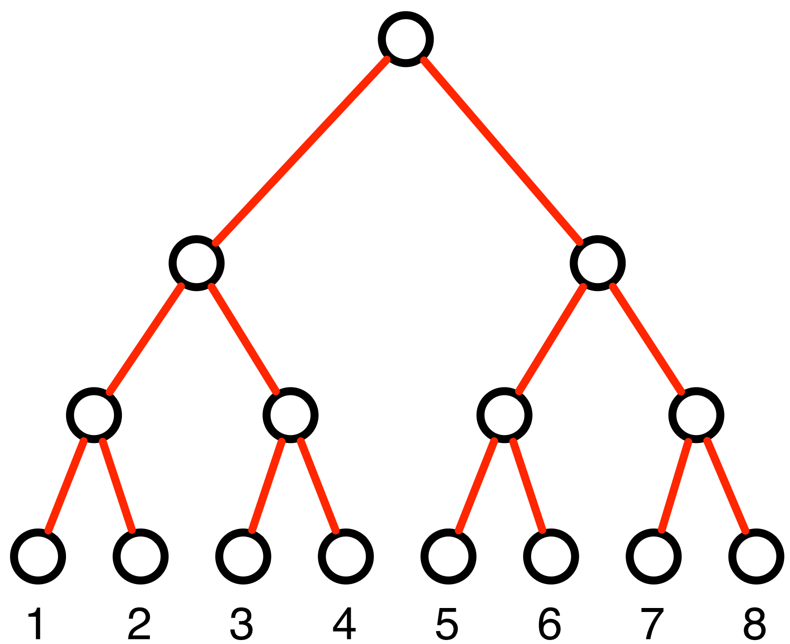




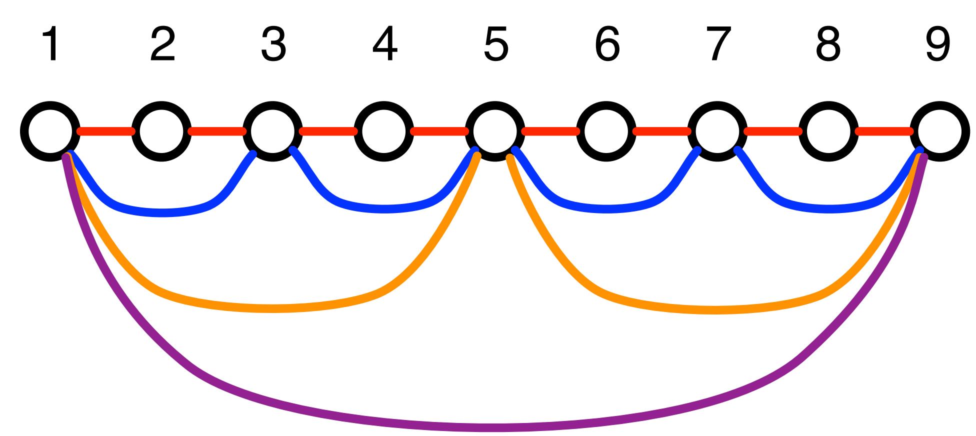
$$p(x_1, \ldots, x_9) = p(x_1, x_9) p(x_5|x_1, x_9) p(x_3|x_1, x_5) p(x_2|x_1, x_3) p(x_4|x_3, x_5) \ldots$$

Multi-Resolution Trees

- Willsky et al.
- New variables represent sequence at coarser resolutions
- Prior defined by a tree MRF on Multi-Resolution representation

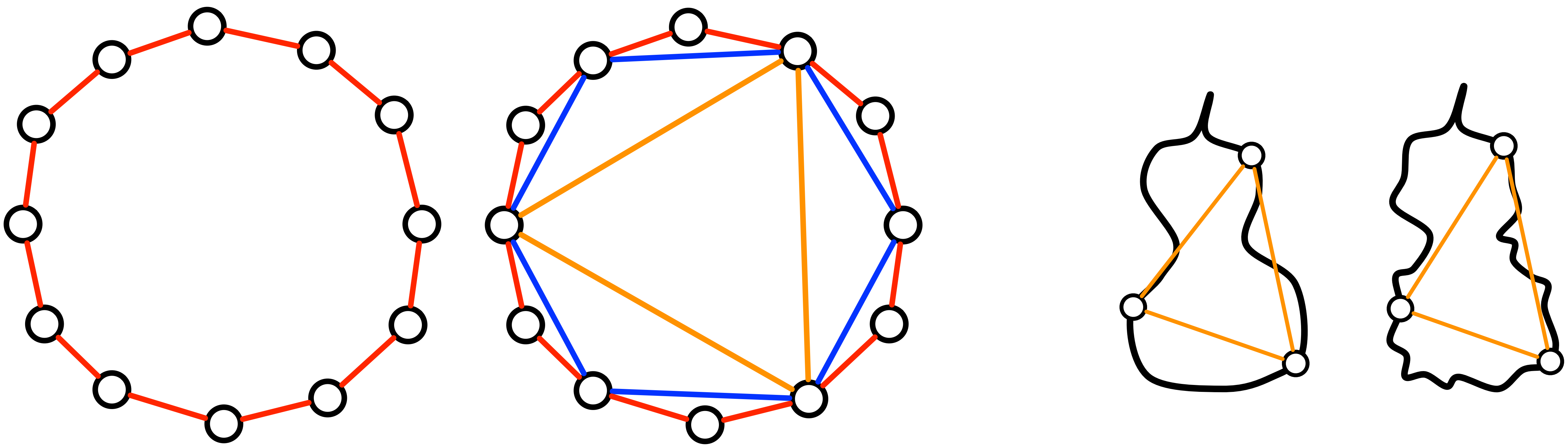


MR tree



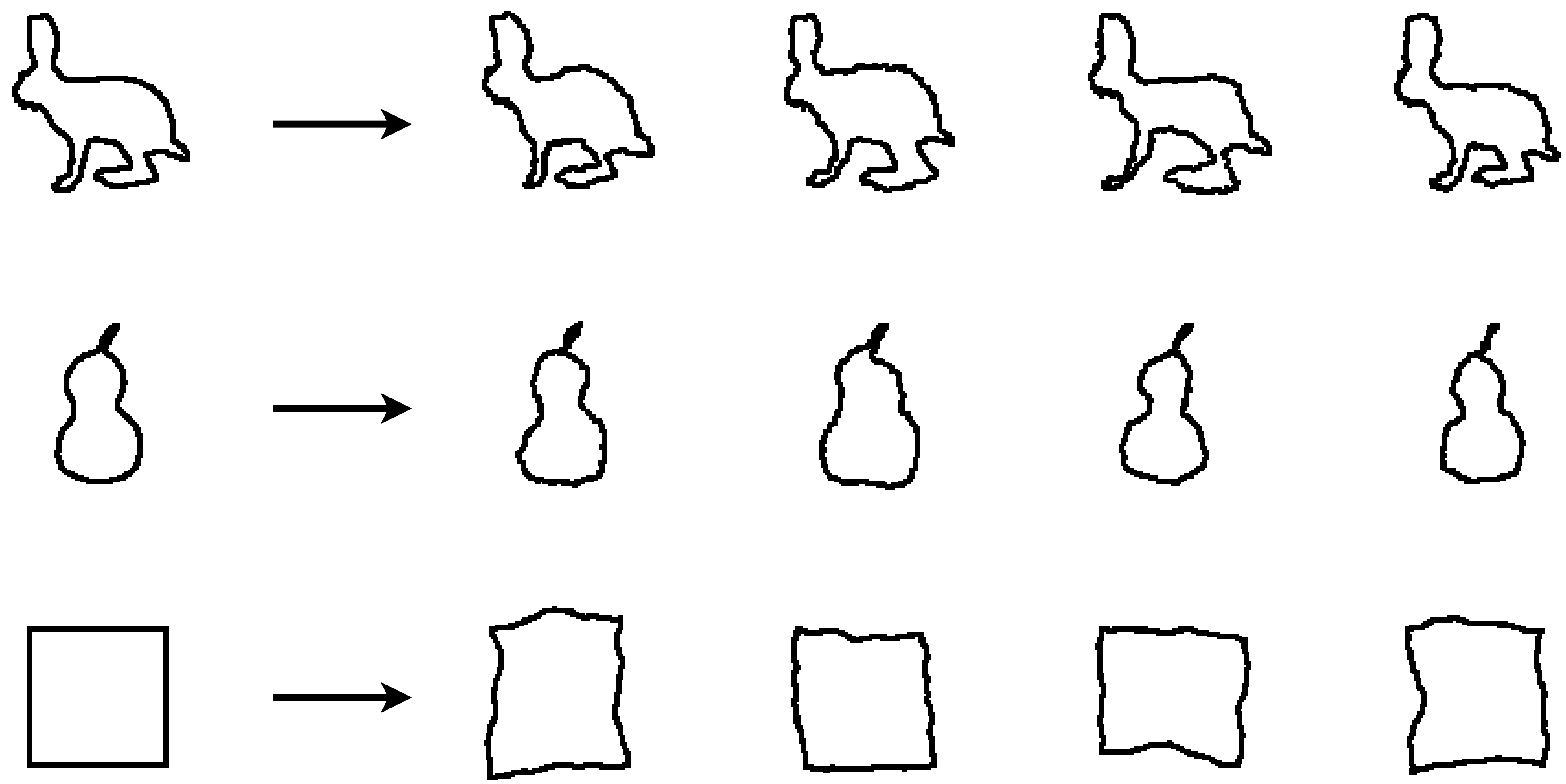
MS sequence

Closed curves / Cyclic sequences



- Both graphs have tree-width 2
  - Tractable inference ~ $O(|X|^3)$
  - Reasonable number of parameters ~ $O(|X|^3)$
- Multiscale model captures global shape properties

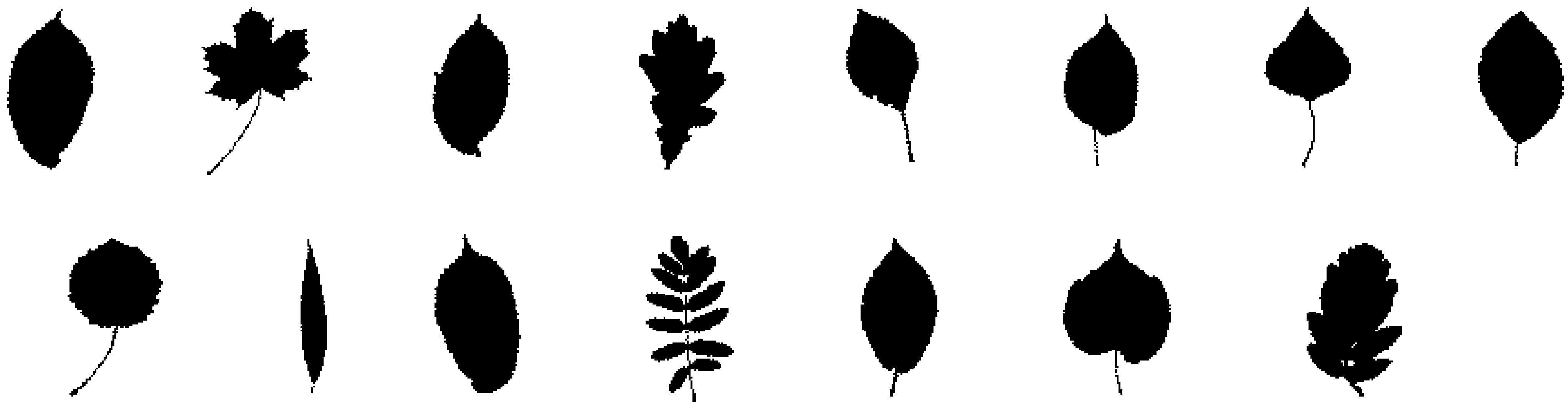
Random deformations



Multiscale model does a good job capturing global shape properties - less drift with similar deformation

Shape recognition

Swedish leaf dataset



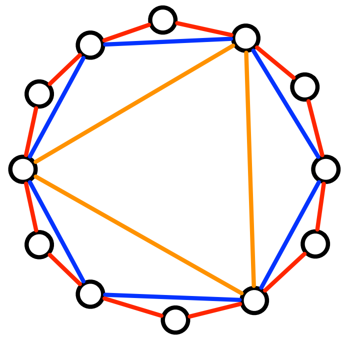
classification	
Multiscale model	96.28
Inner distance	94.13
Shape context	88.12

15 species

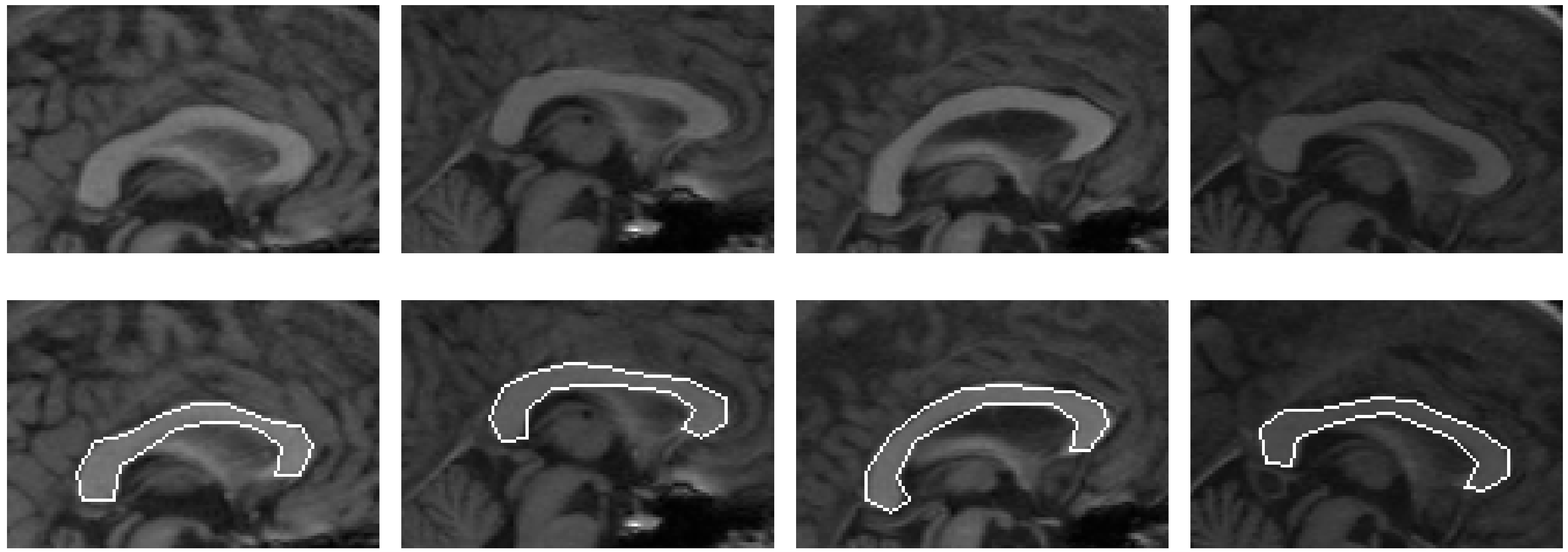
75 examples per species

(25 training, 50 test)

Shape Detection



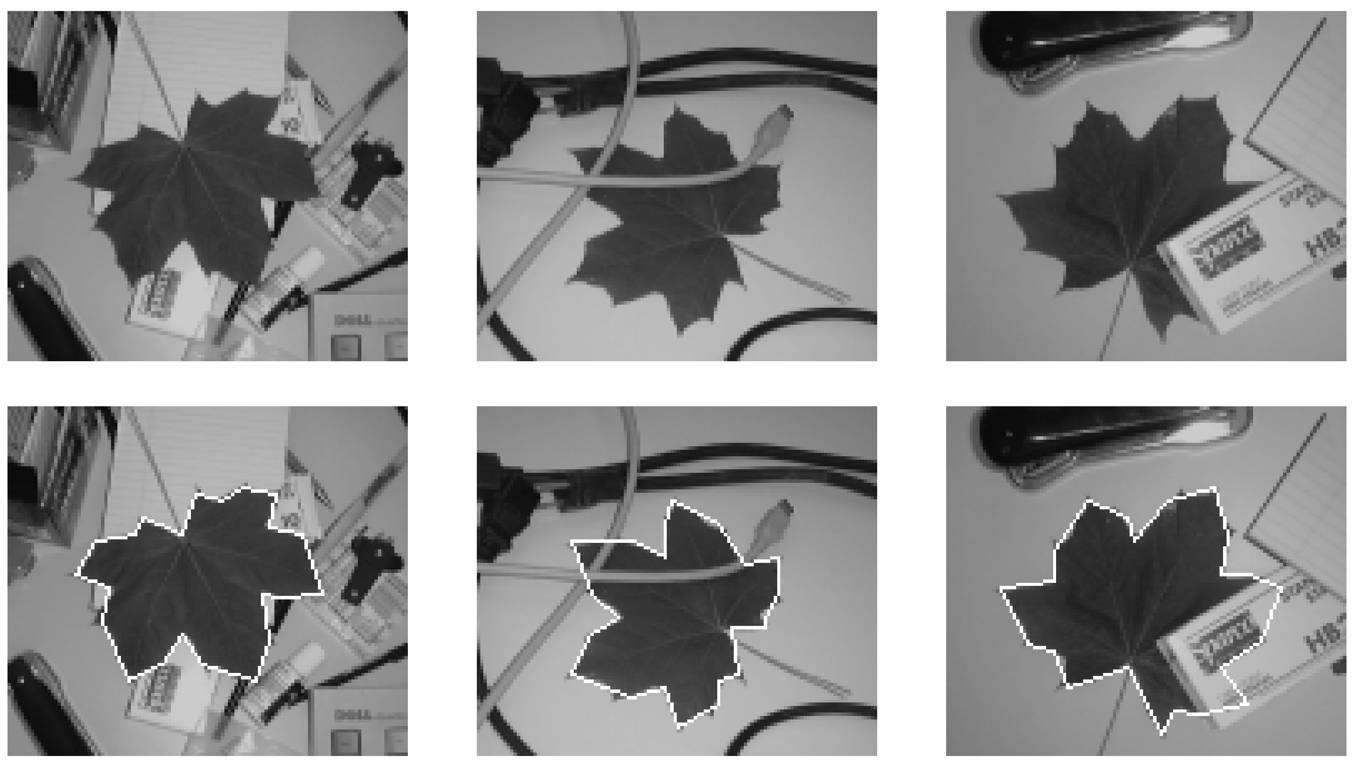
template
defining
$p(x)$



y

$p(x|y)$

template
defining
$p(x)$

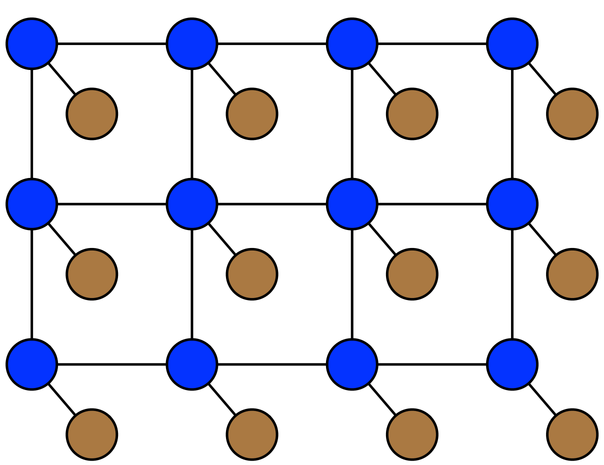


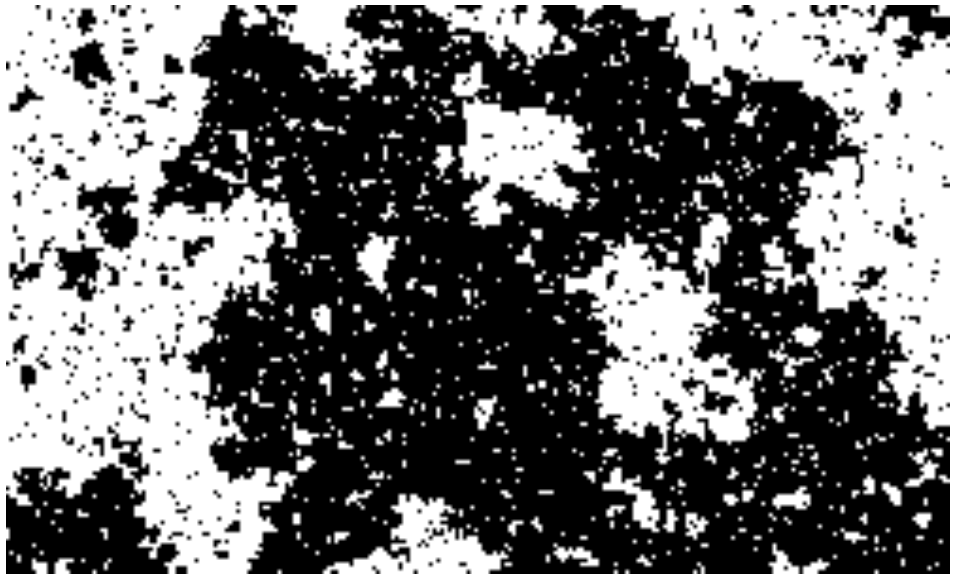
y

$p(x|y)$

Images

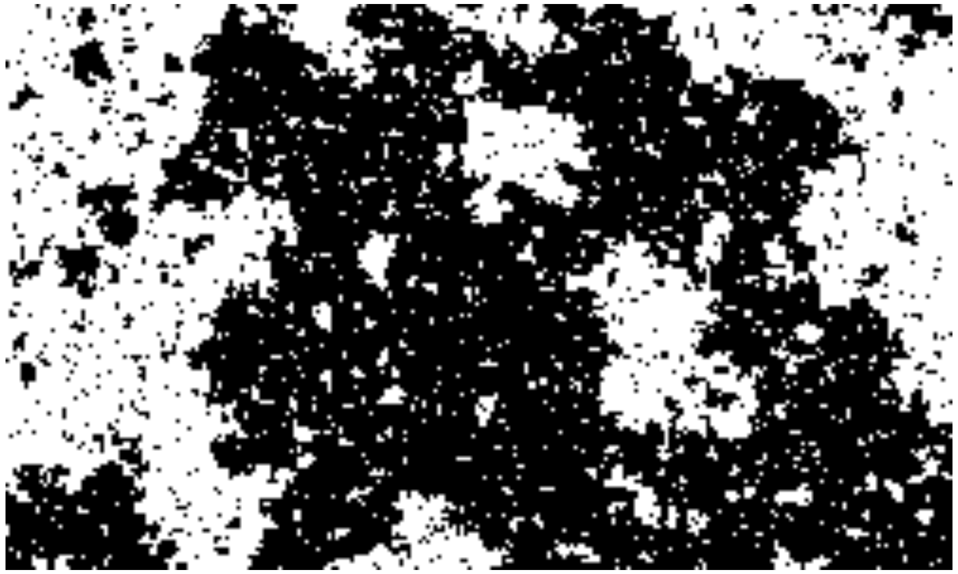
- MRF models widely used to model images
- Applications:
  - image restoration: clean picture is piecewise smooth
  - image segmentation: foreground mask is spatially coherent



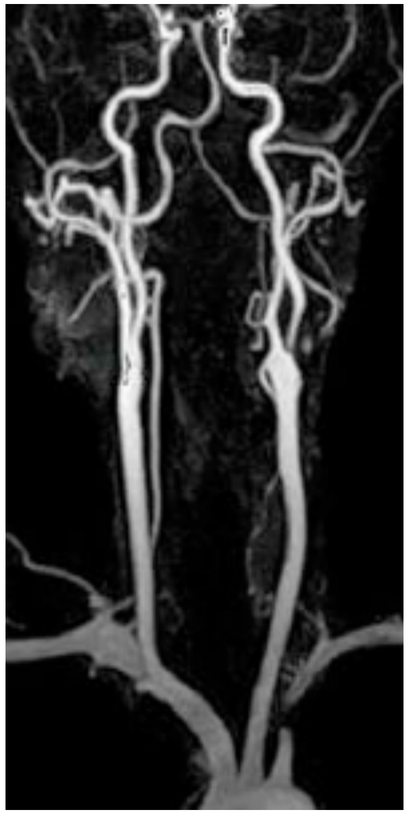


$p(x)$ = Ising model

Binary images

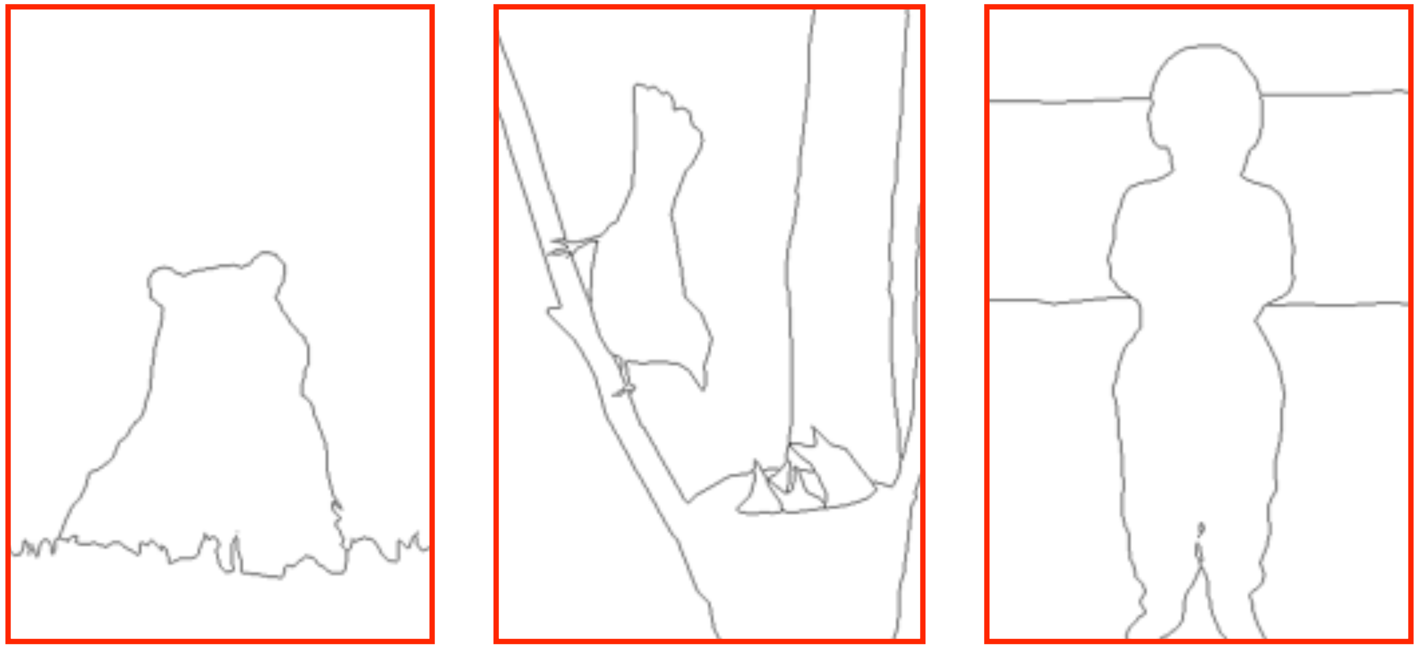


$p(x)$ = Ising model



$p(x) = ?$

Contour maps

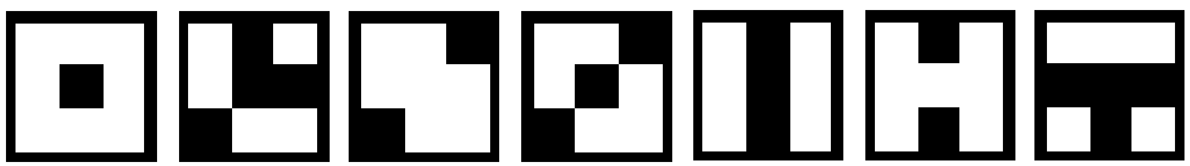


Berkeley Segmentation Dataset

- x is a binary image
  - pixel is "on" if contour goes through it
- Lots of regularities
  - Continuity, smoothness, closure, parallel lines, symmetries
- How can we build a reasonable model for $p(x)$?

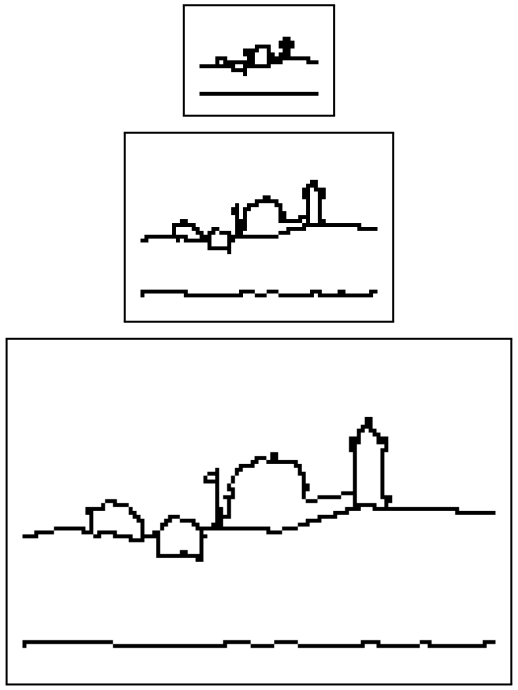
Fields-of-Patterns

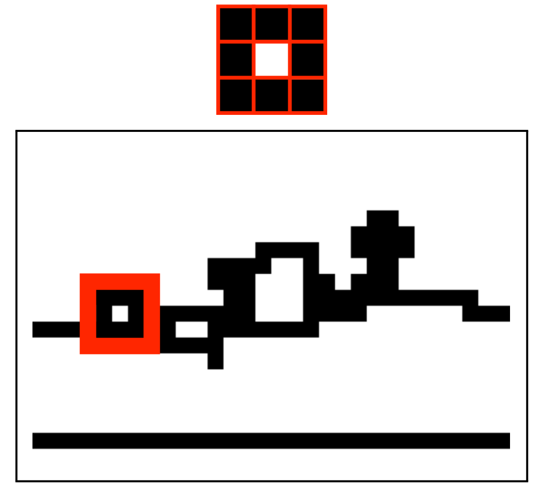
- Local property of x ~ binary pattern in 3x3 window



- Look at local properties at multiple resolutions
- Local property of coarse image reflects global properties

x



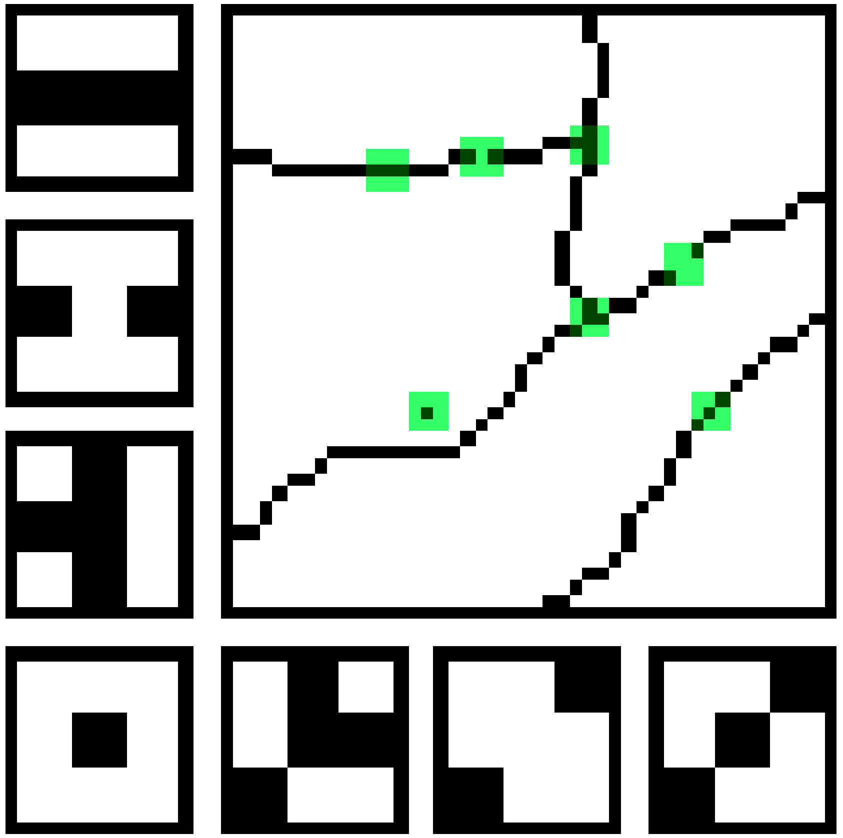


Single-scale model

- Energy model $p(x) = \frac{1}{Z} e^{-E(x)}$
  - Look at 3x3 blocks of pixels b
  - Each block has one of 512 patterns

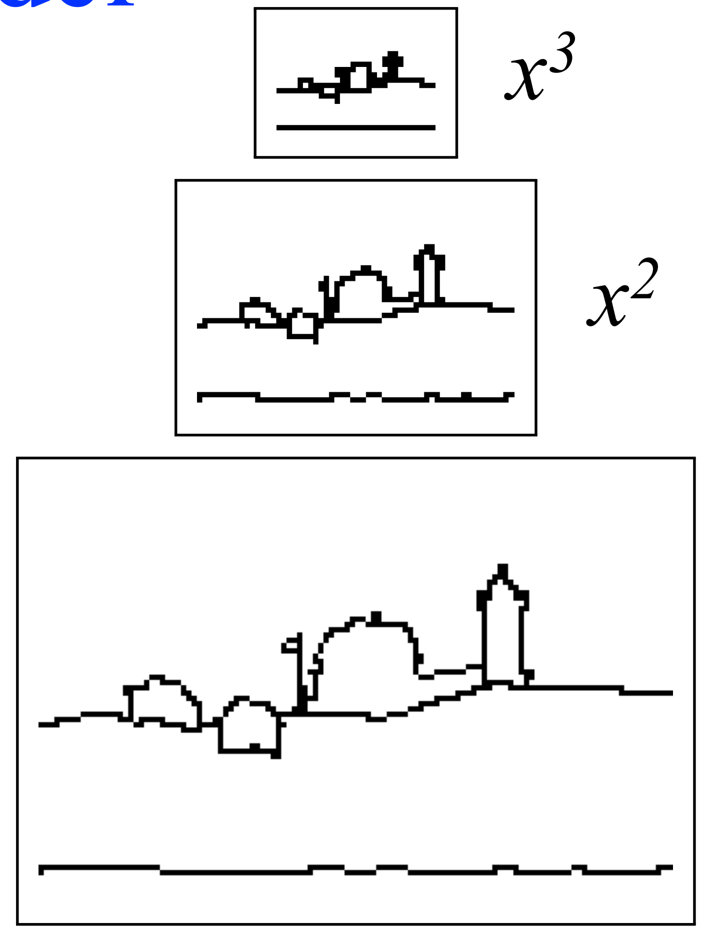
$$E(x) = \sum_b V(x_b)$$

  - V is an array of 512 costs



- Captures continuity, frequency of 1s, frequency of junctions
- But no smoothness, parallelism, closed curves, etc.

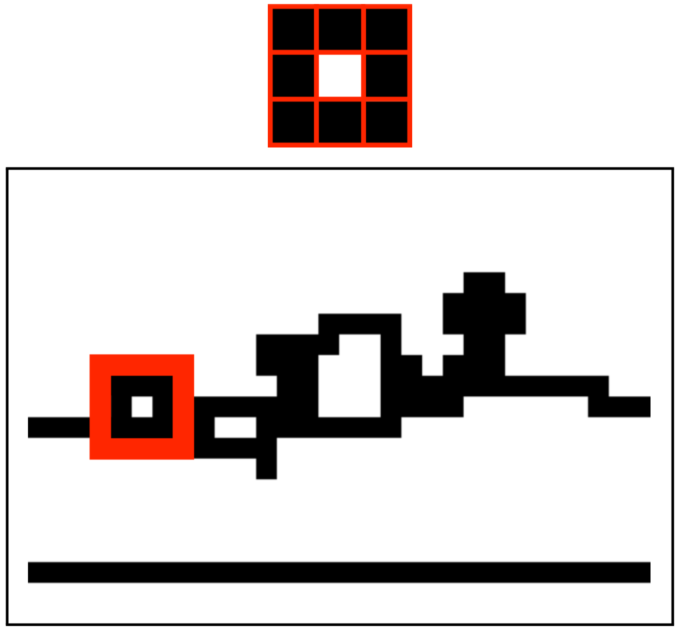
Multi-scale model



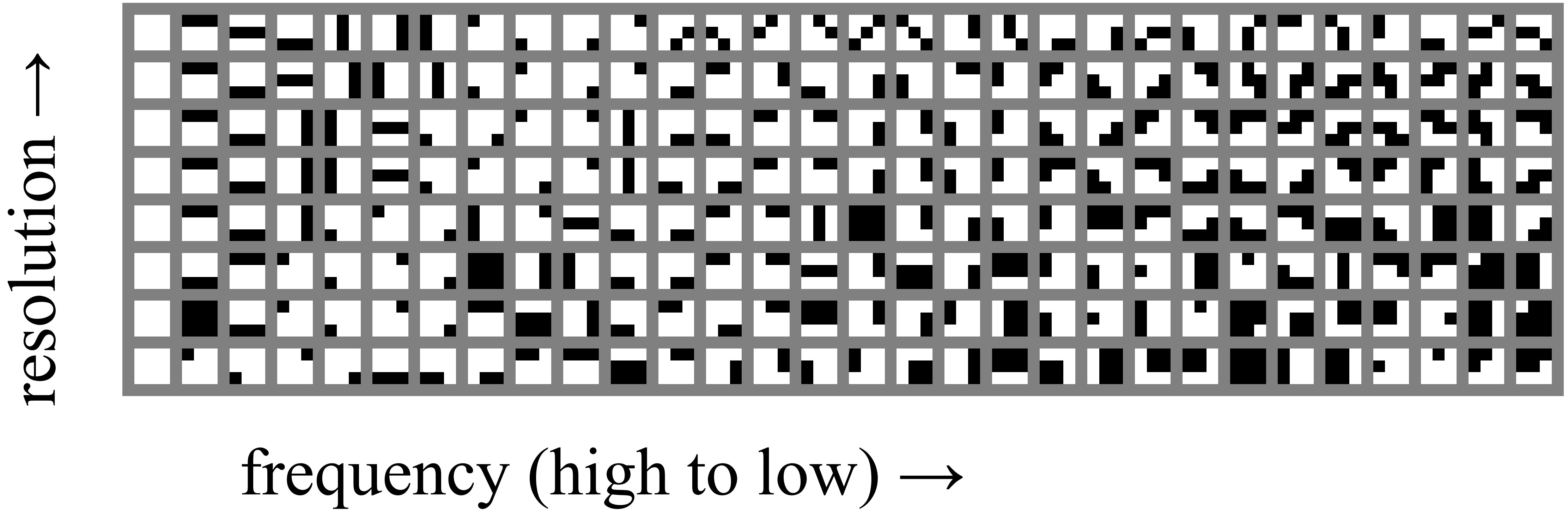
- OR pyramid
  - $x^1 \ldots x^K$
  - x^{i+1} is a coarsening of x^i
- Look at 3x3 blocks at all resolutions

$$E(x) = \sum_l \sum_b V^l(x_b^l)$$

  - $V^i \neq V^j$
  - K arrays of 512 costs

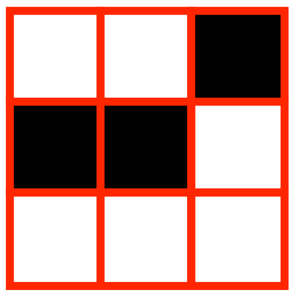


Frequency of Patterns (BSDS)

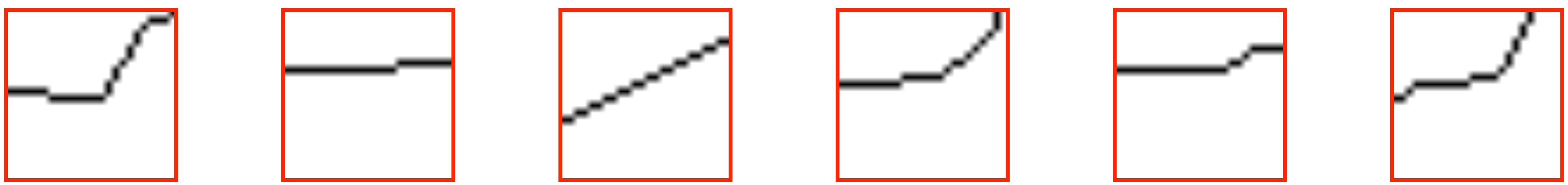


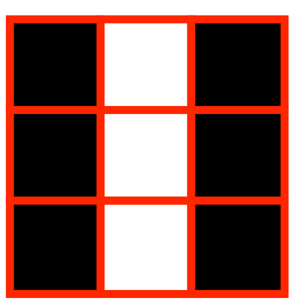
Maximum likelihood model matches frequencies of patterns

Coarse patterns (BSDS)

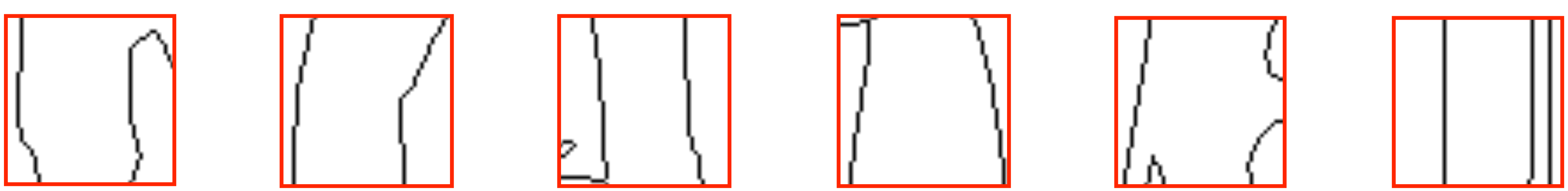


level 4 24x24

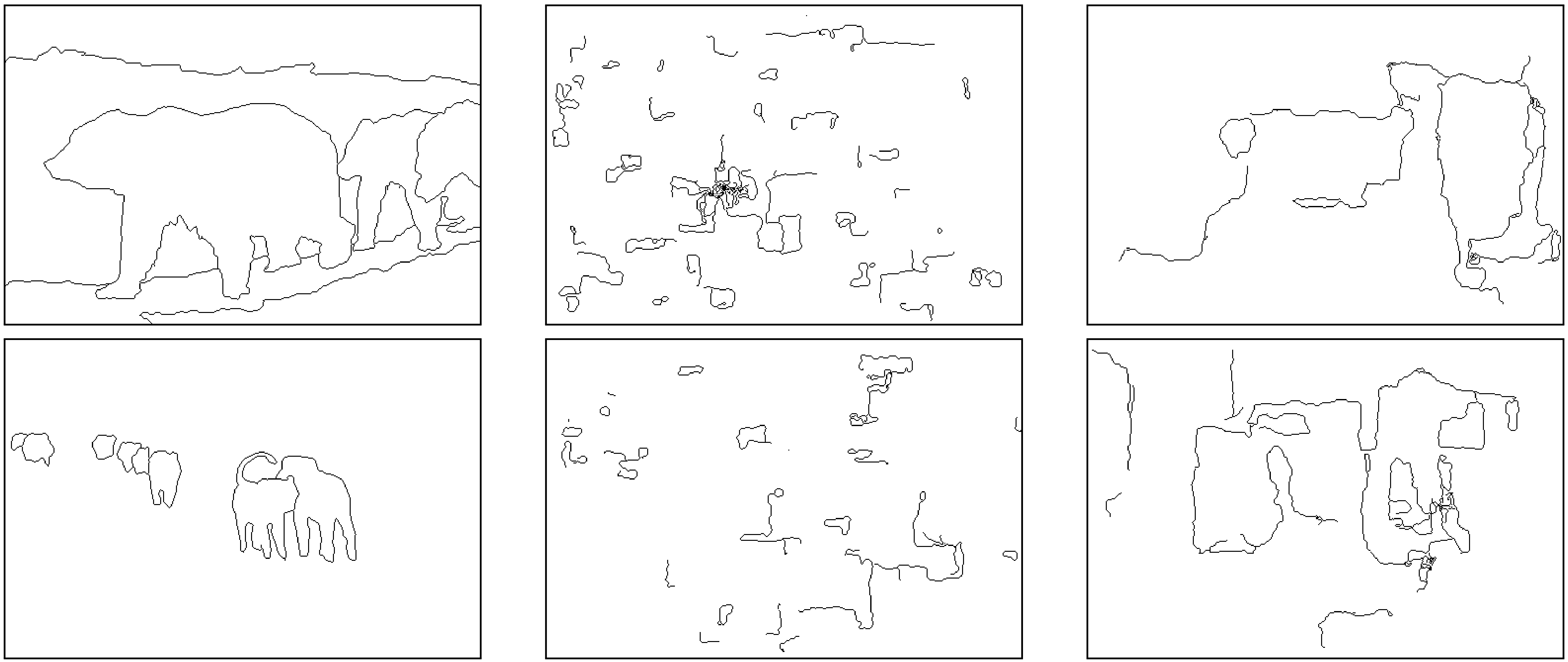




level 5 48x48



Samples from the prior $p(x)$

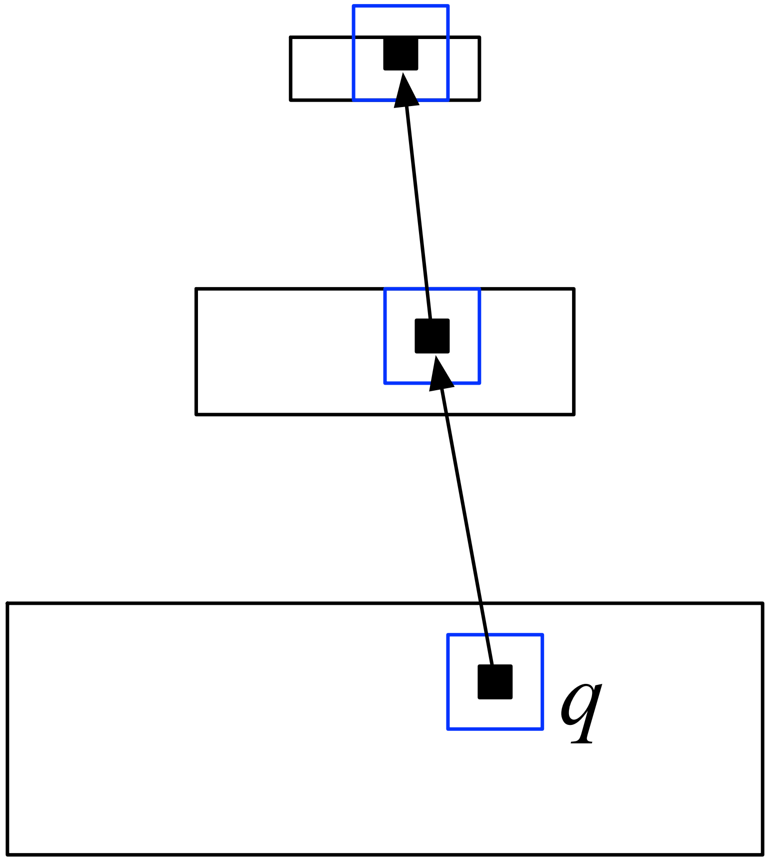


training data　　single-scale　　multi-scale

Inference

- Inference with MRF generally hard
- Singlescale FOP
  - hard but model is local
    - Gibbs sampling
    - Loopy BP
- Multiscale FOP
  - model not local on x but local on pyramid

Gibbs sampling



- Repeatedly update pixels/blocks
  - Sample new value for x_q given rest of x
  - Requires energy difference between $x_q = 0$ and $x_q = 1$
- Efficient computation using multiscale representation
  - Change in x_q affects a small number of auxiliary variables
  - Energy difference is local over $x^1 \ldots x^K$

Maximum Likelihood Estimation

$$p(x) = \frac{1}{Z} e^{-E(x)} \qquad E(x) = w \cdot \phi(x)$$

- $\phi(x)$: vector of counts of each pattern at each scale
- w : vector of costs
- Training examples $x_1, \ldots, x_n$
- Negative log-likelihood is convex
- MLE model: $E_p[\phi(x)] = \overline{\phi(x_i)}$
  - expected freq of each pattern = average freq in training data

Stochastic Gradient Descent

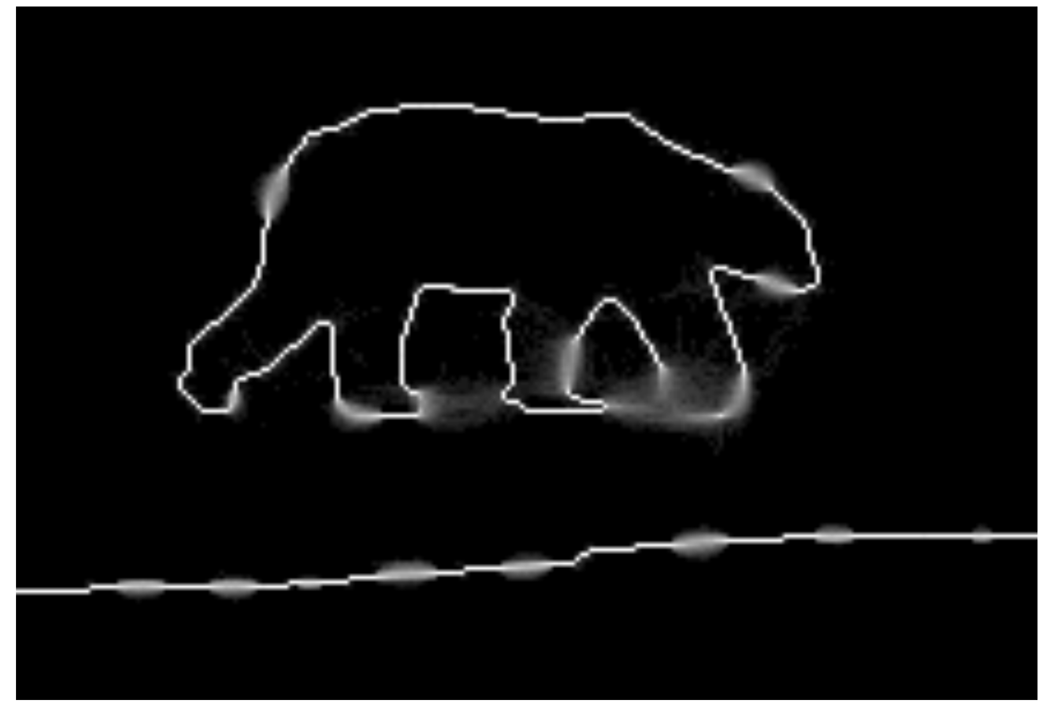
$$w' = w + \eta(E_p[\phi(x)] - \overline{\phi(x_i)})$$

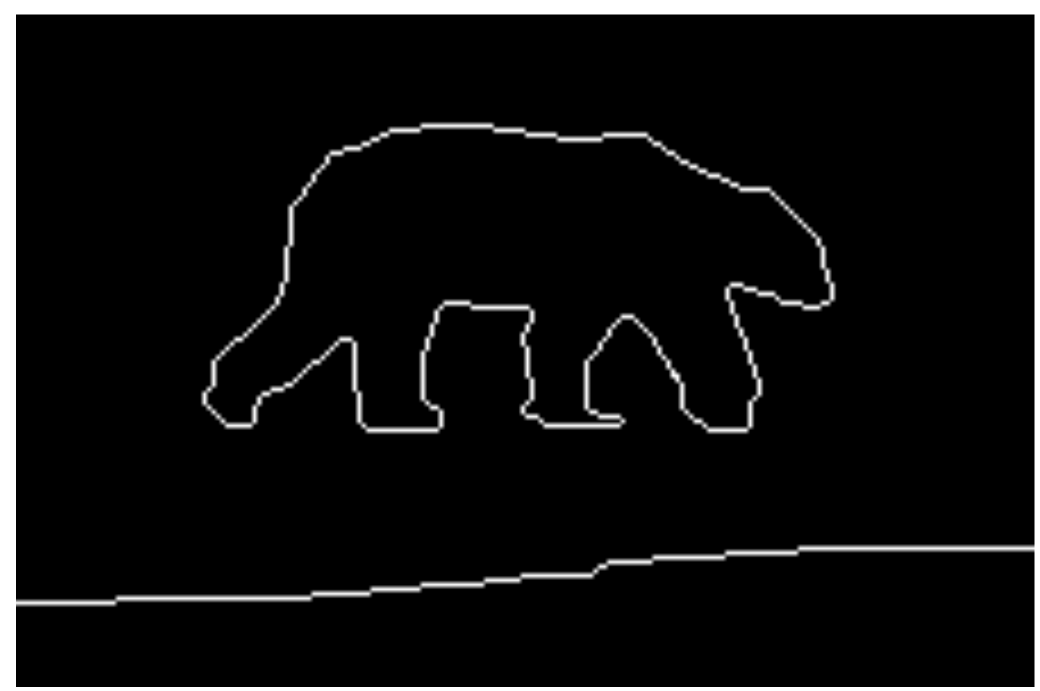
- Estimate expectation by sampling from *p*
- Mix MCMC simulation with gradient descent
  - Let *M* be Markov chain with stationary distribution *p*
  - Maintain *m* states $s_1, \ldots, s_m \sim p(x)$
    - Update model $w' = w + \eta(\overline{\phi(s_i)} - \overline{\phi(x_i)})$
    - Evolve $s_1, \ldots, s_m$ according to *M* for a *few* steps

Contour completion

$p(x_q\!=\!1|y)$

y

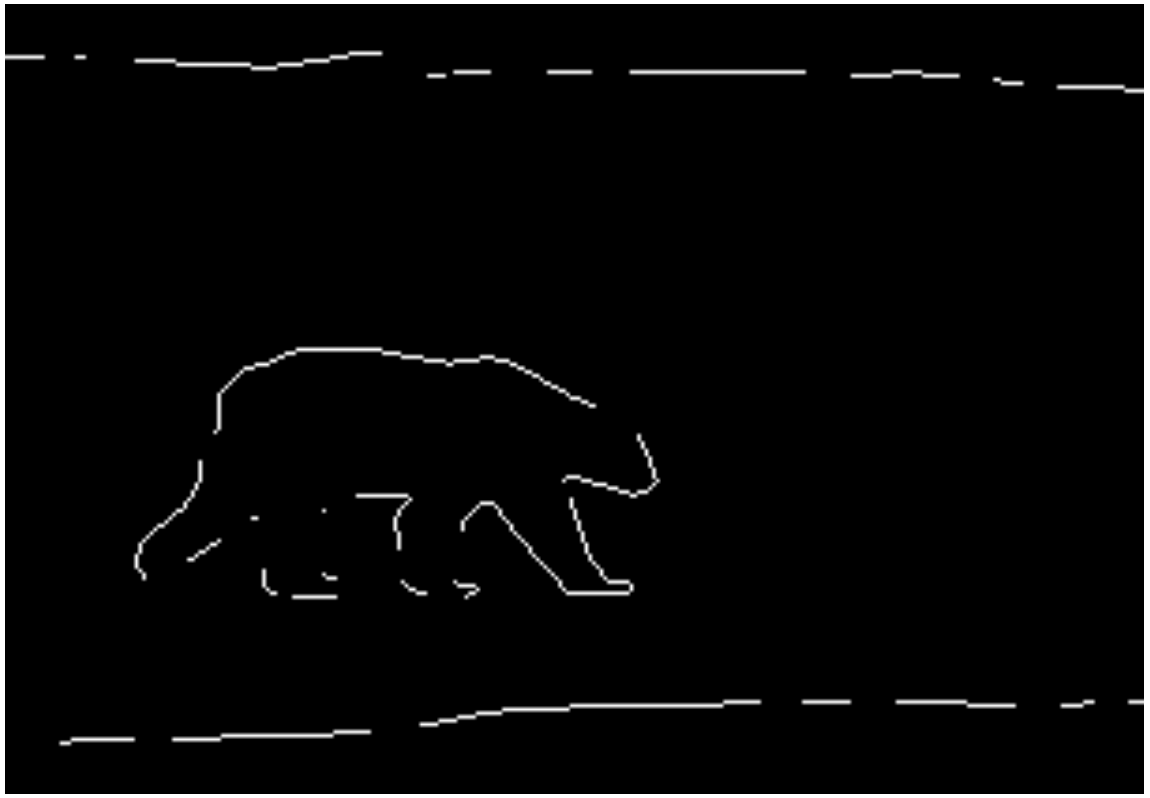


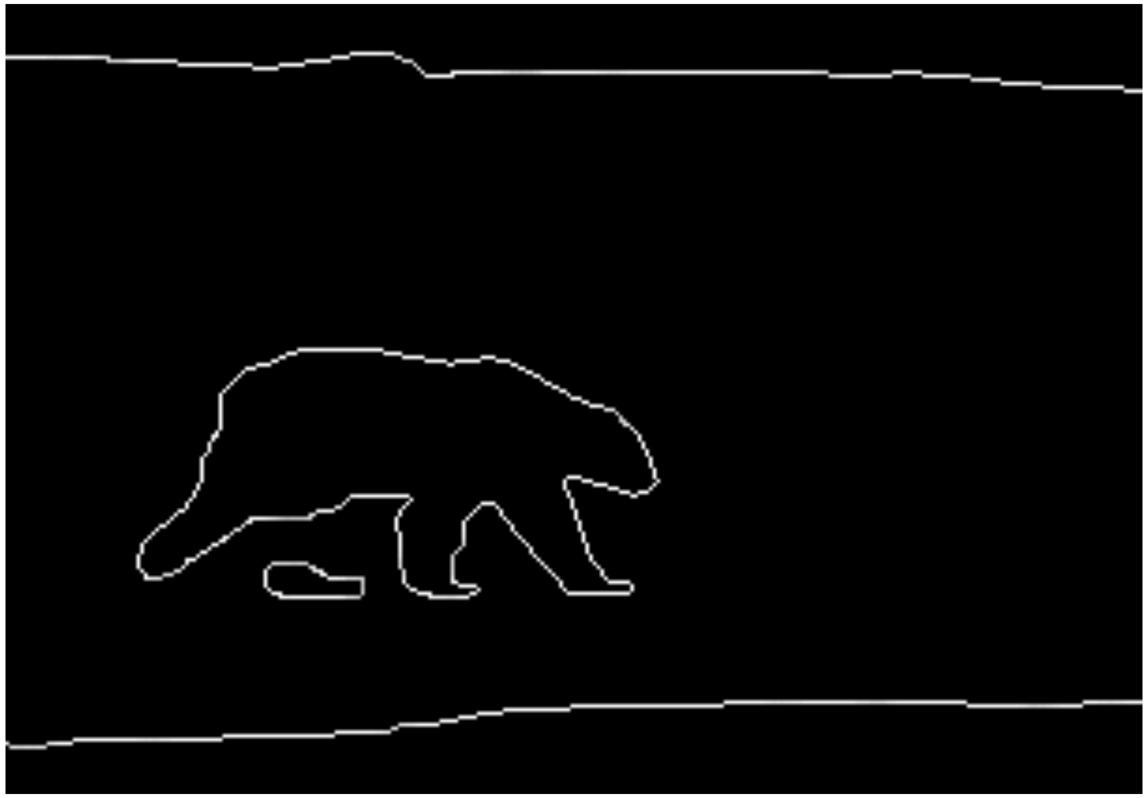


“Stochastic completion field”

MAP

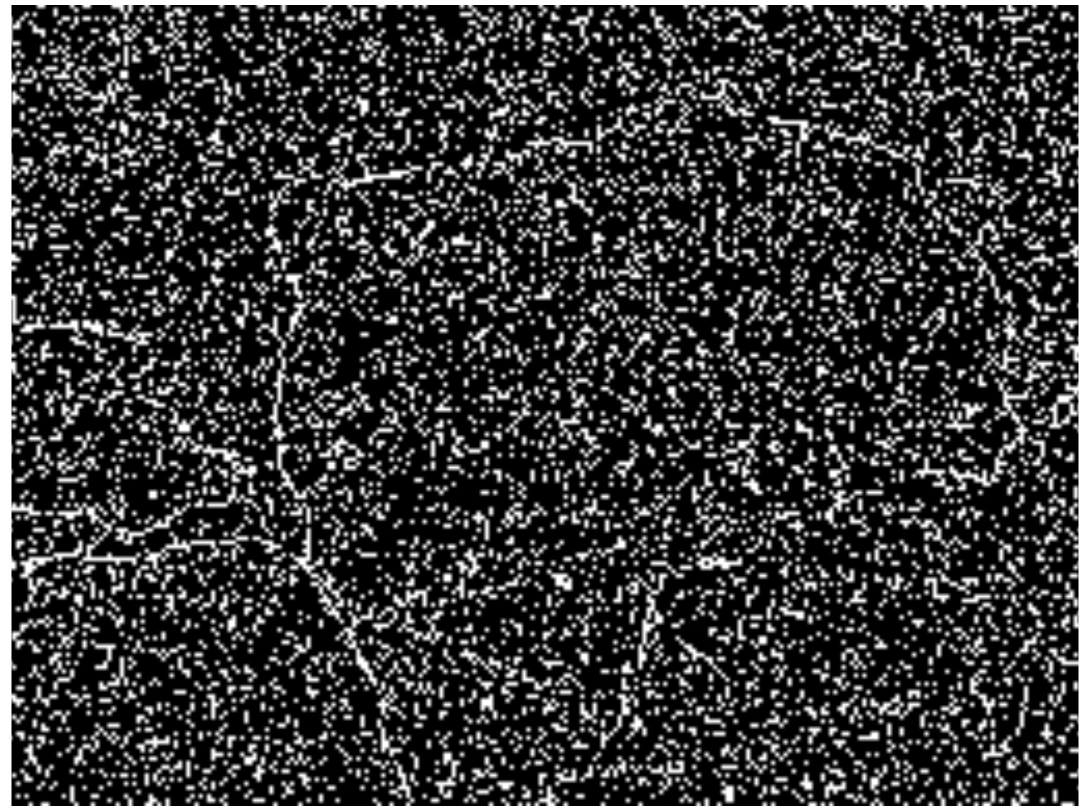
Contour completion





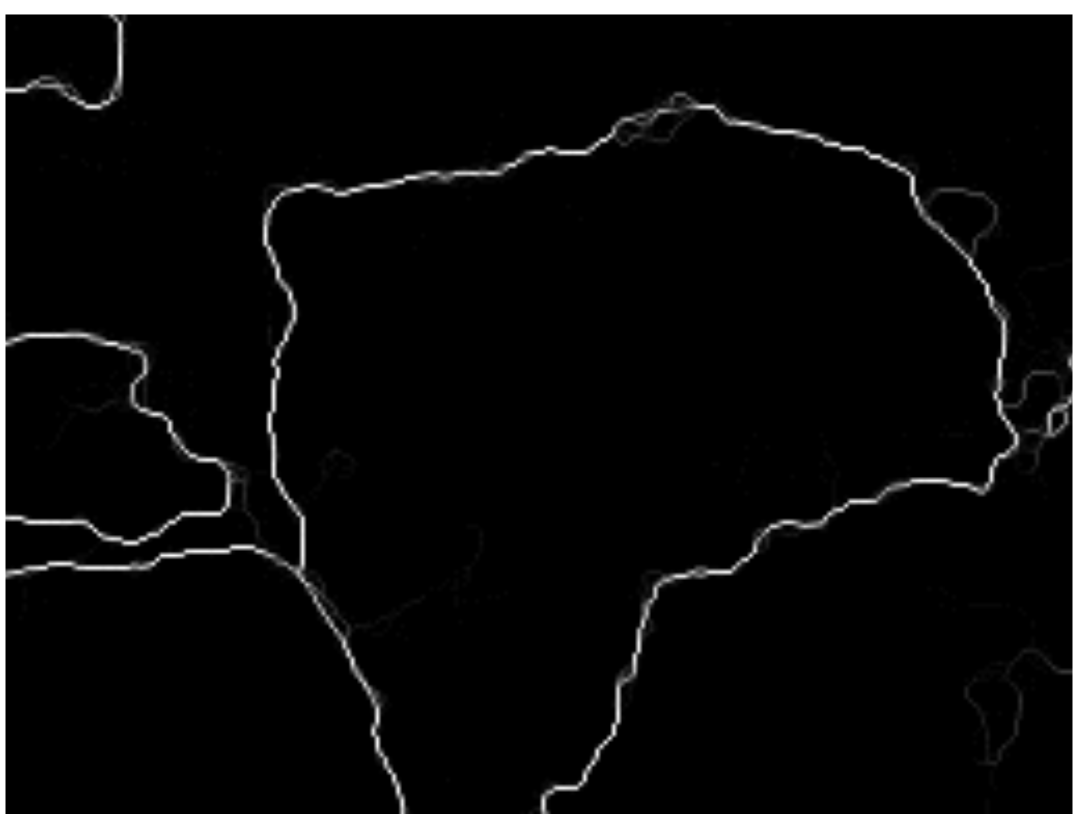
completion/restoration

iid noise
20% flipped



y

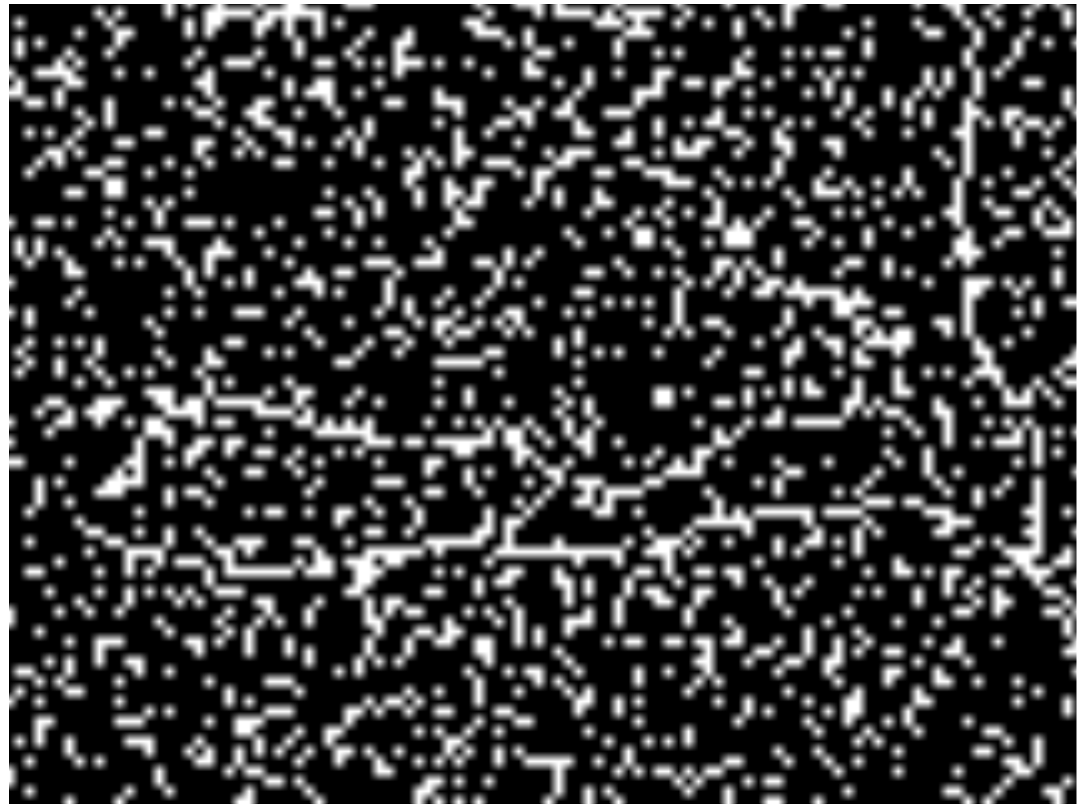
$p(x_q=1|y)$



x

completion/restoration

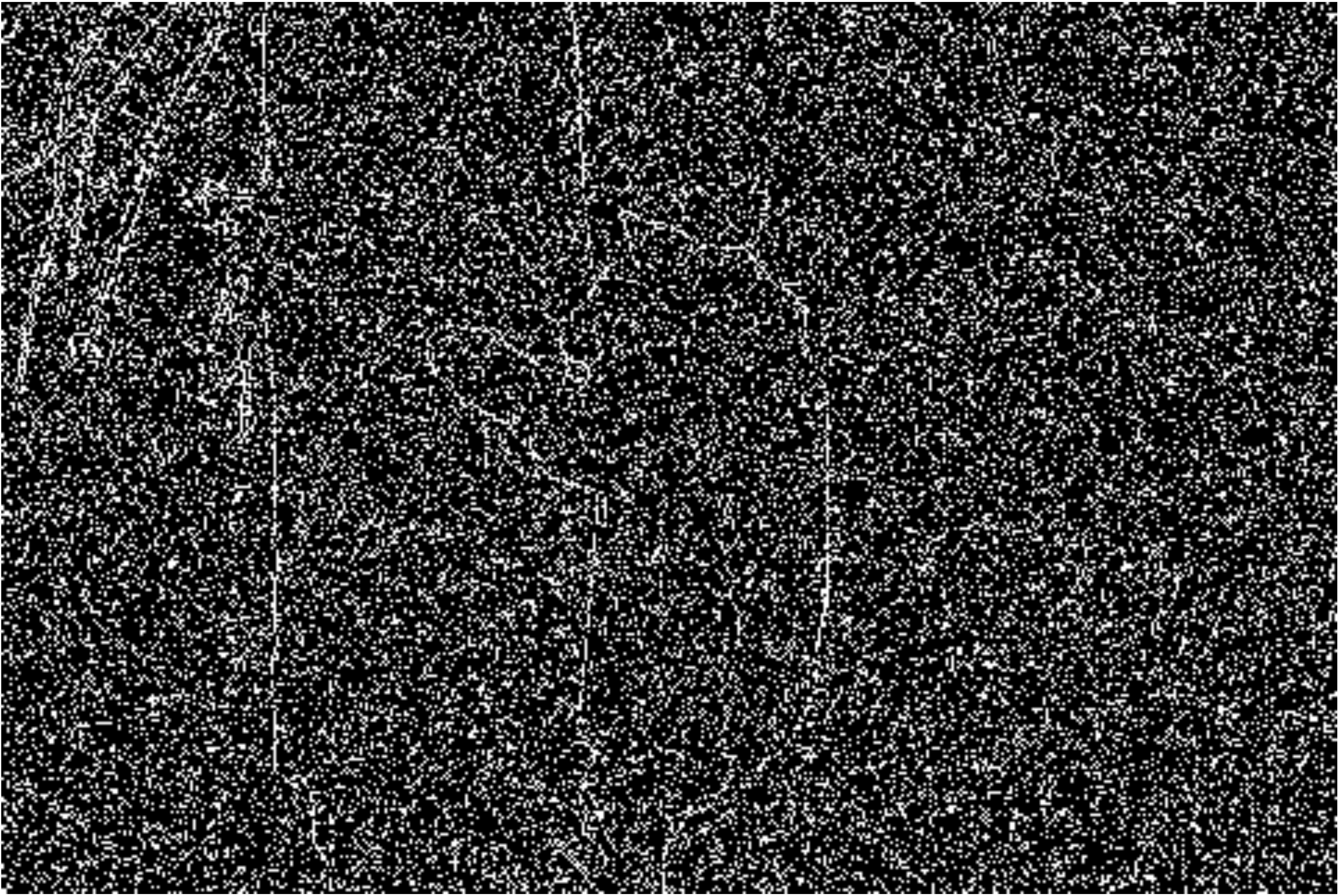
iid noise
20% flipped



y

$p(x_q=1|y)$

x



Summary

- Standard Markov models capture local properties
- MS models can capture local properties at multiple resolutions
- Local-property of coarse x = global property of x
- Future directions
  - coarse-to-fine inference
  - non-binary images
  - data models