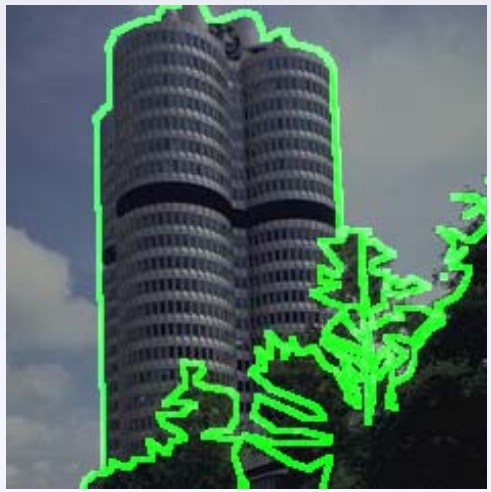
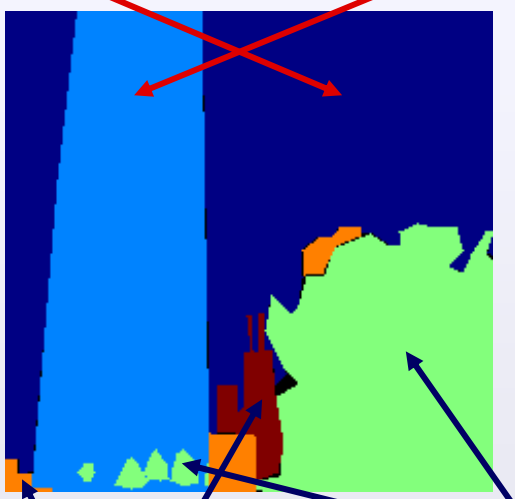
Shared Segmentation of Natural Scenes using Dependent Pitman-Yor Processes

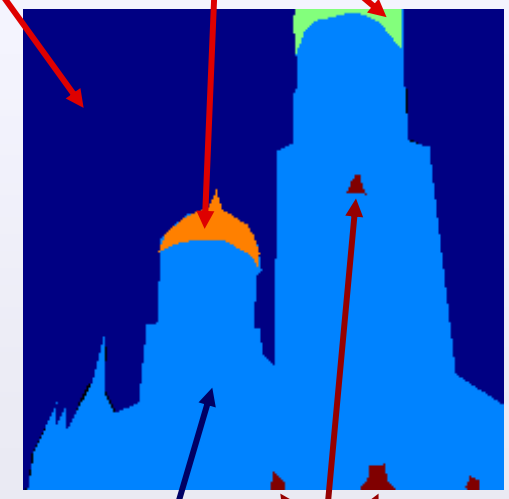
Erik Sudderth & Michael Jordan

University of California, Berkeley

Parsing Visual Scenes





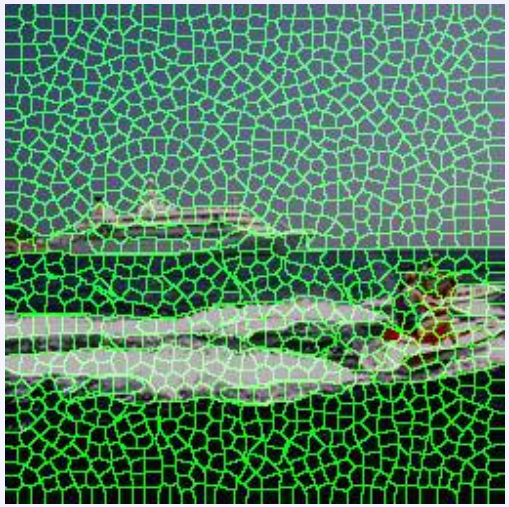


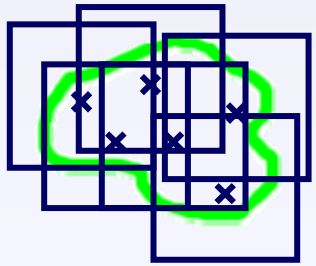
Are Images Bags of Features?

Inspired by the successes of *topic models* for text data, some have proposed learning from *local image features*

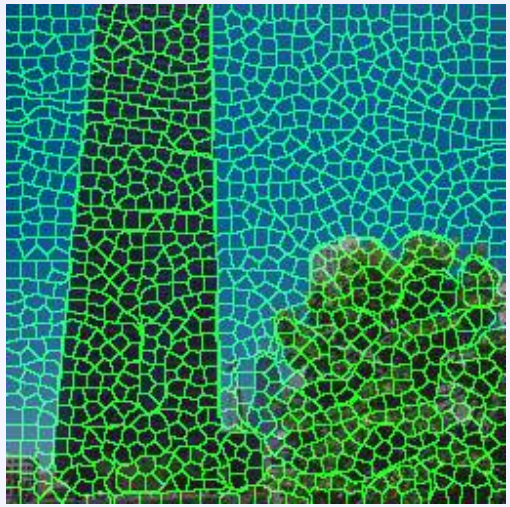
Are Images Bags of Features?

Inspired by the successes of *topic models* for text data, some have proposed learning from *local image features*





Compute color & texture descriptors for each superpixel



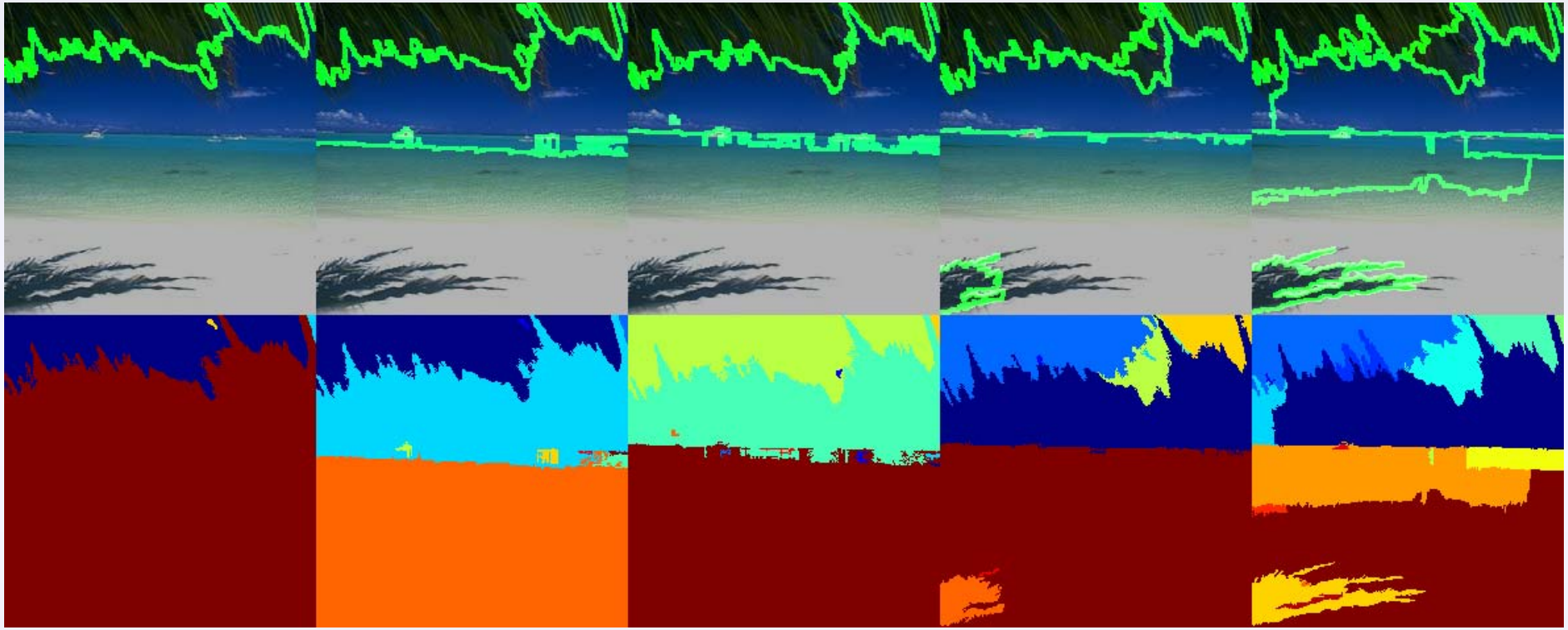
First Approach: *Fei-Fei & Perona 2005, Sivic et. al. 2005*

- Ignore spatial structure entirely (bag of "*visual words*")

Second Approach: *Russell et. al. 2006, Todorovic & Ahuja 2007*

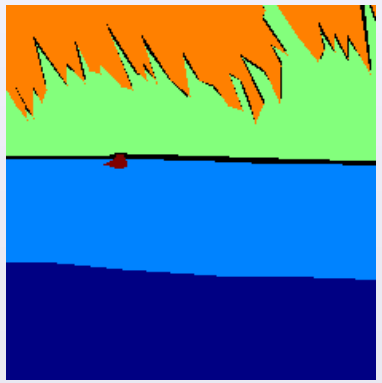
- Cluster features via one or more *bottom-up segmentations*

Segmentation: Mean Shift

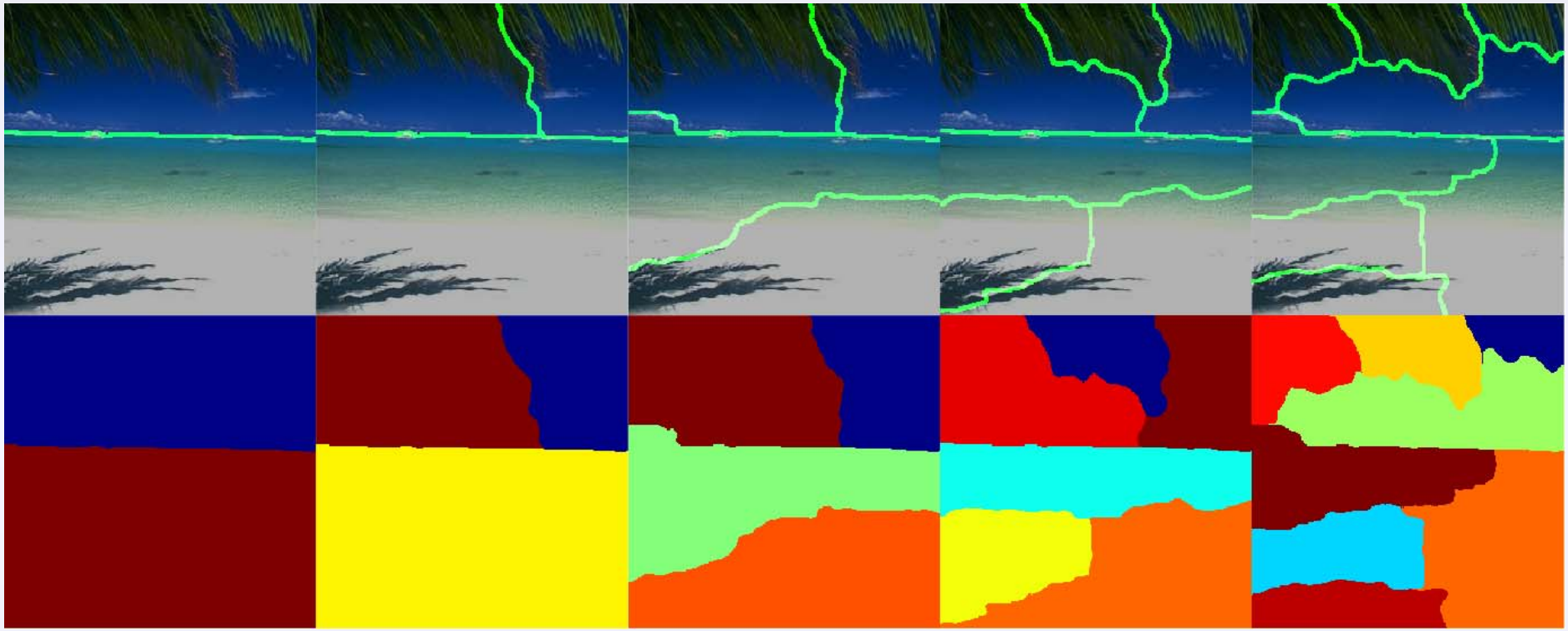


EDISON: *Comaniciu & Meer, 2002*

- Cluster by modes of appearance features
- Often *sensitive* to bandwidth parameter

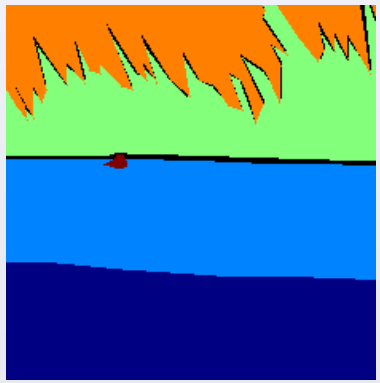


Segmentation: Normalized Cuts

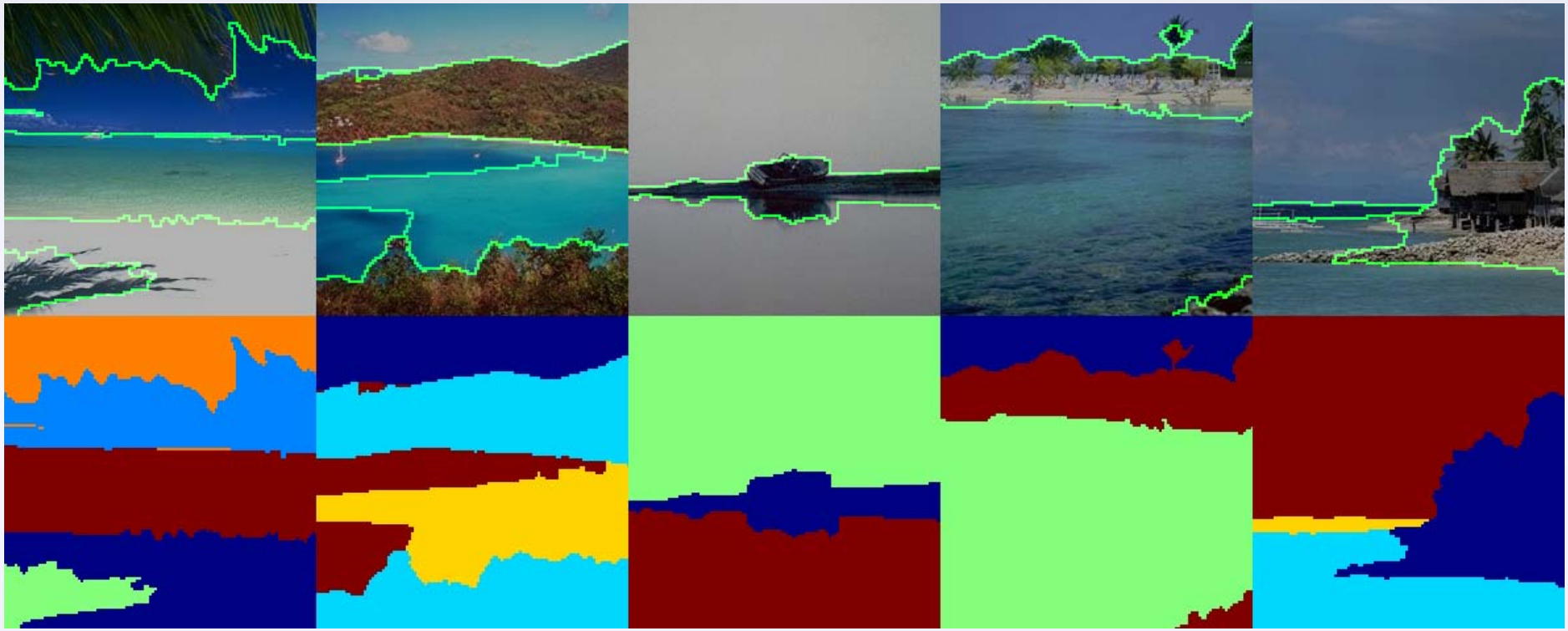


Shi & Malik 2000; Fowlkes, Martin, & Malik 2003

- Implicit bias towards *equal-sized regions*
- Is this a good model for real scenes?



Segmentation: New Approach

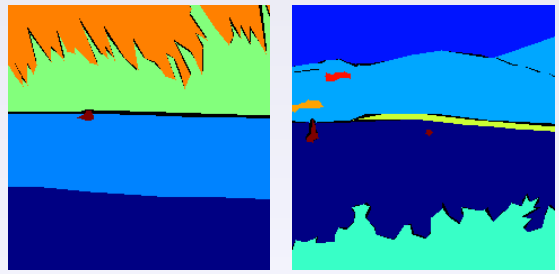


Spatially Dependent Pitman-Yor Processes

- Automatically infers the *number* of segments
- Handles regions of widely varying *size* and *appearance*
- Statistical framework for discovering *shared* categories

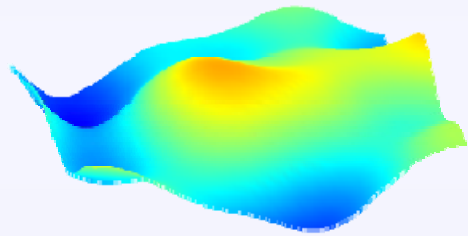
Outline

Natural Scene Statistics



- Counts, partitions, and power laws
- Hierarchical *Pitman-Yor* processes

Spatial Priors for Image Partitions



- What's wrong with Potts models?
- Spatial dependence via *Gaussian processes*

Unsupervised Image Analysis

- Image *segmentation*
- Visual *category discovery*

Priors on Counts & Partitions

Segmentation as Partitioning

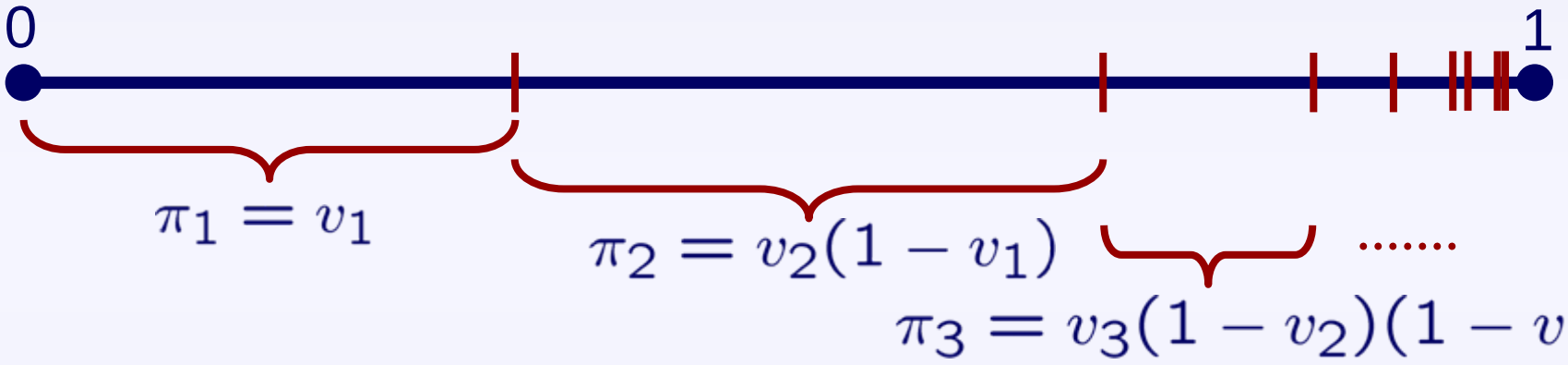
- How many regions does this image contain?
- What are the sizes of these regions?

Unsupervised Object Category Discovery

- How many object categories have I observed?
- How frequently does each category appear?

Pitman-Yor Processes

The *Pitman-Yor process* defines a distribution on infinite discrete measures, or *partitions*



$$\pi_k = v_k \left(1 - \sum_{\ell=1}^{k-1} \pi_\ell\right) = v_k \prod_{\ell=1}^{k-1} (1 - v_\ell)$$

$$v_k \sim \text{Beta}(1 - a, b + ka)$$

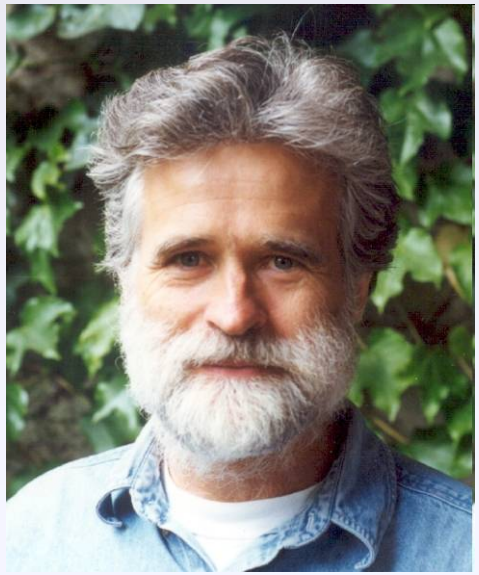
Dirichlet process:

$$a = 0$$

Why Pitman-Yor?

Generalizing the Dirichlet Process

- Distribution on partitions leads to a generalized *Chinese restaurant process*
- Special cases arise as excursion lengths for Markov chains, Brownian motions, …



Jim Pitman

Power Law Distributions

	DP	PY
Number of unique clusters in N observations	$\mathcal{O}(b \log N)$	$\mathcal{O}(bN^a)$
Size of sorted cluster weight k	$\mathcal{O}\left(\alpha_b \left(\frac{1+b}{b}\right)^{-k}\right)$	$\mathcal{O}\left(\alpha_{ab} k^{-\frac{1}{a}}\right)$

Marc Yor

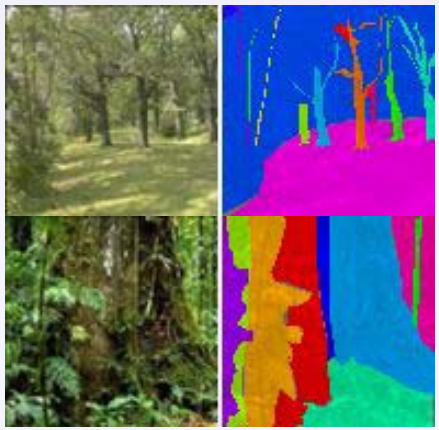
Natural Language Statistics

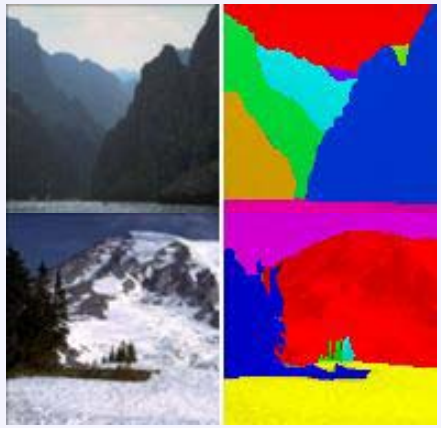
Goldwater, Griffiths, & Johnson, 2005
Teh, 2006

Natural Scene Statistics

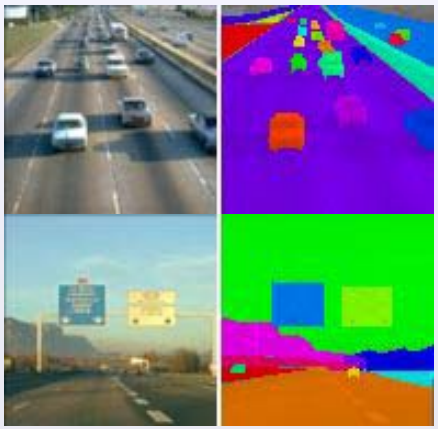
- Does Pitman-Yor prior match human segmentation?
- How do statistics vary across scene categories?


Coast

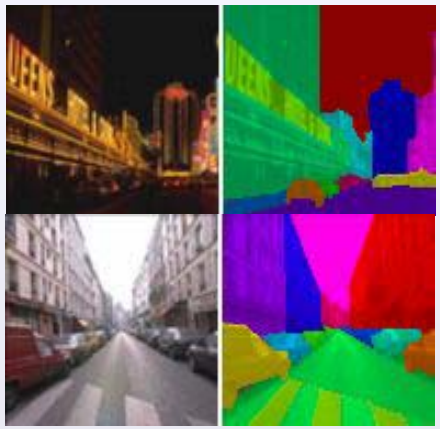

Forest

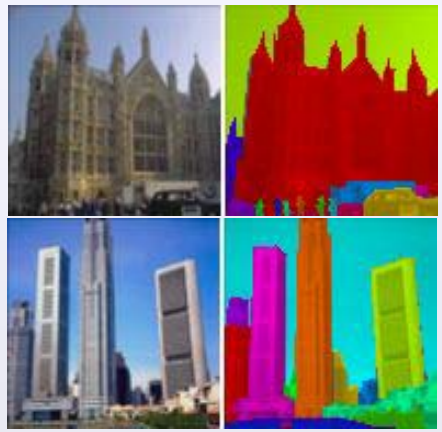

Mountain


Opencountry


Highway

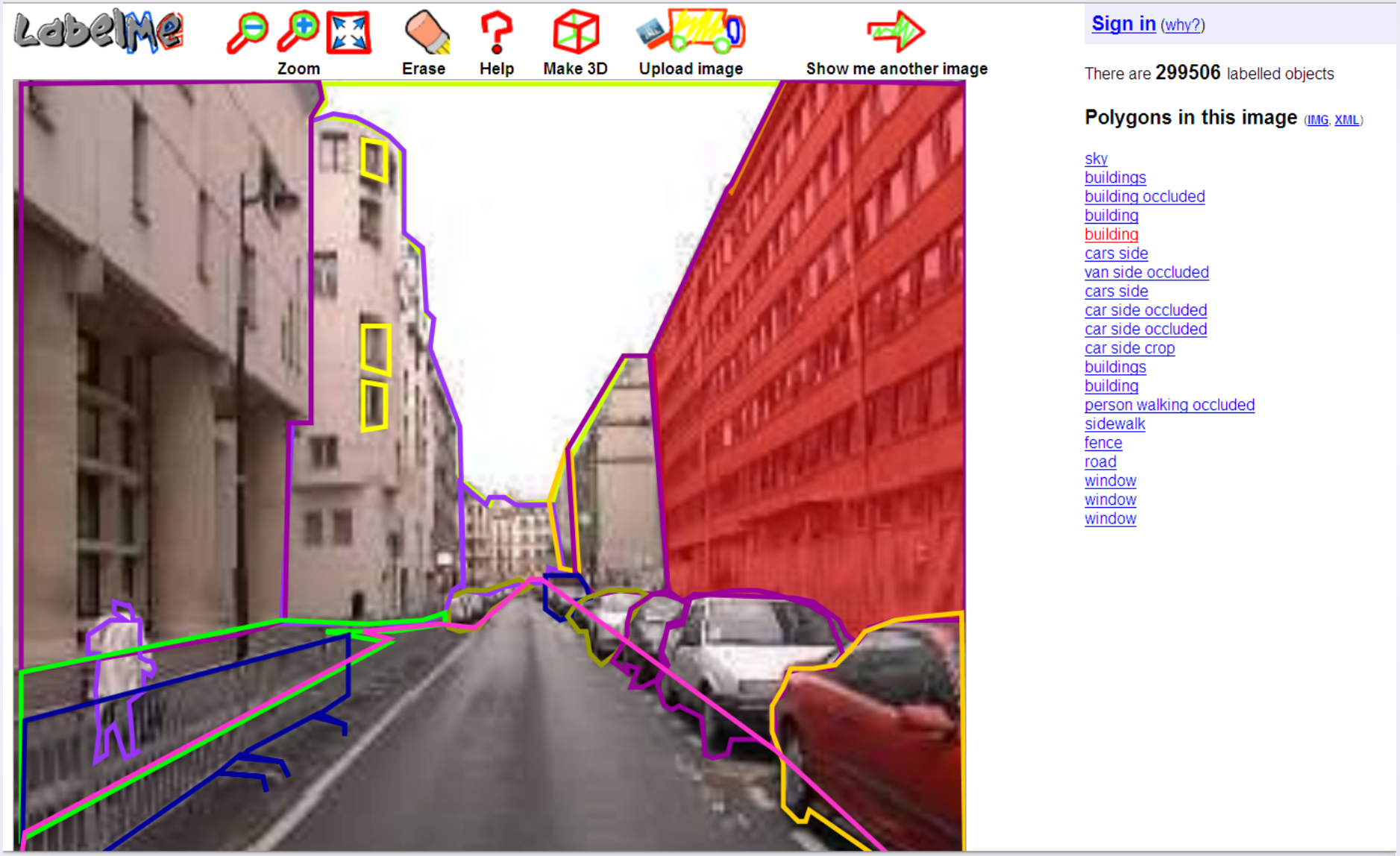

Insidecity


Street


Tallbuilding

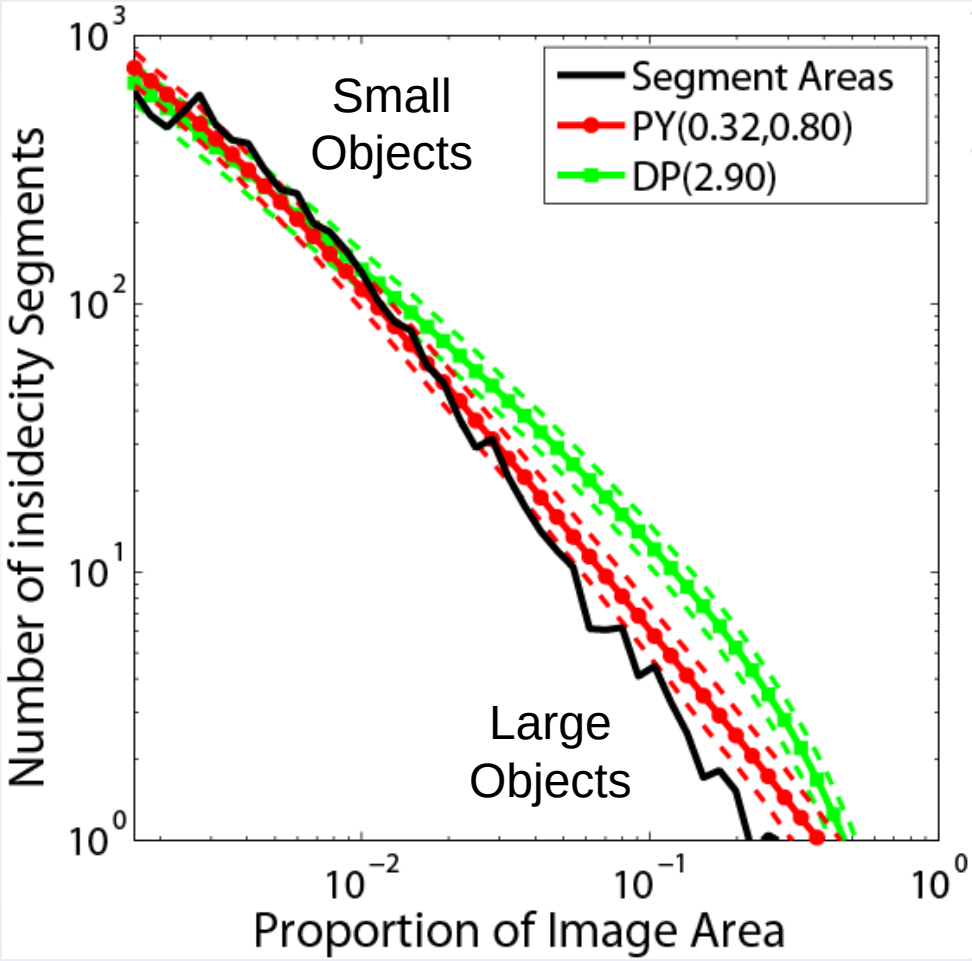
Oliva & Torralba, 2001

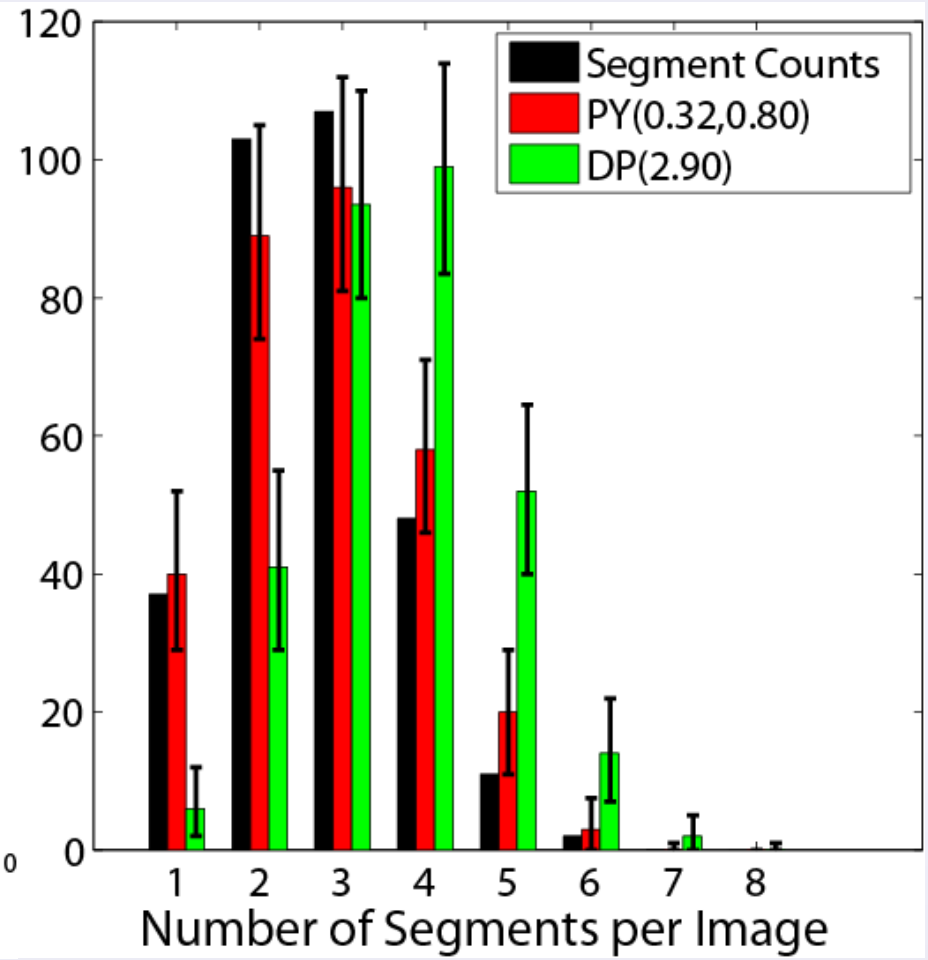
Manual Image Segmentation



Labels for more than 29,000 segments in 2,688 images of natural scenes

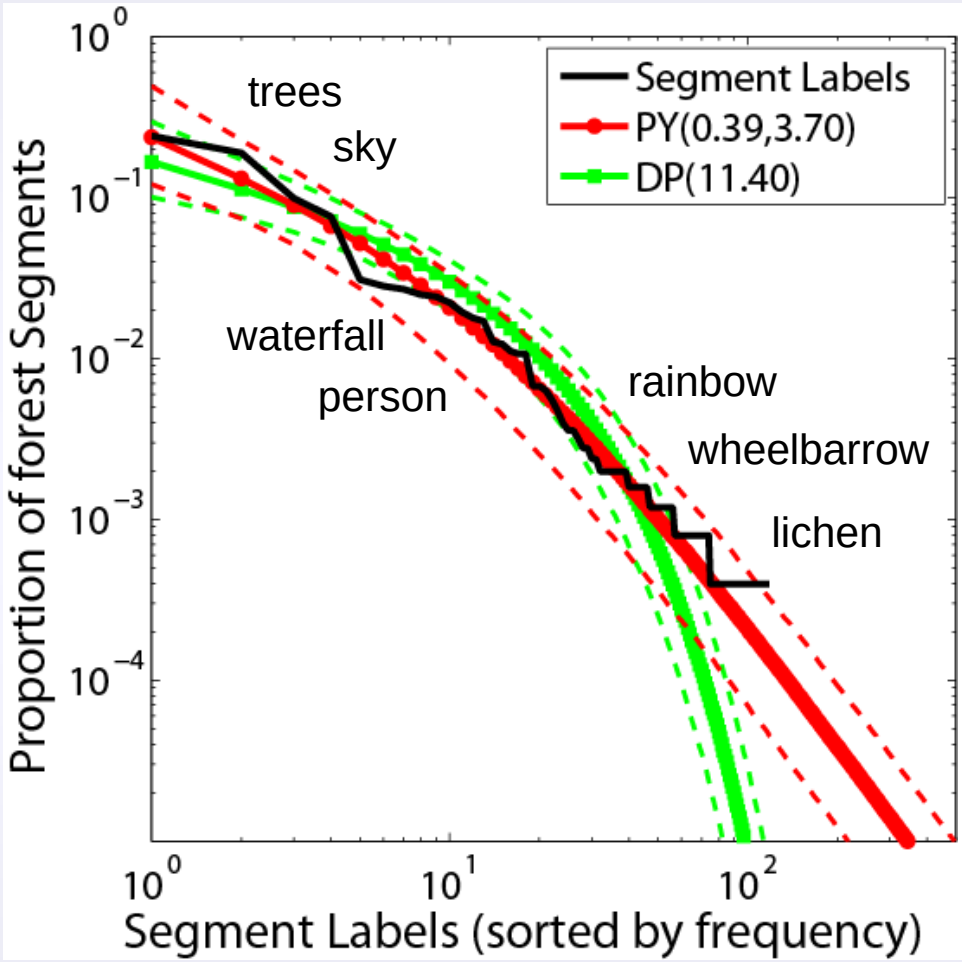
Object Sizes and Counts

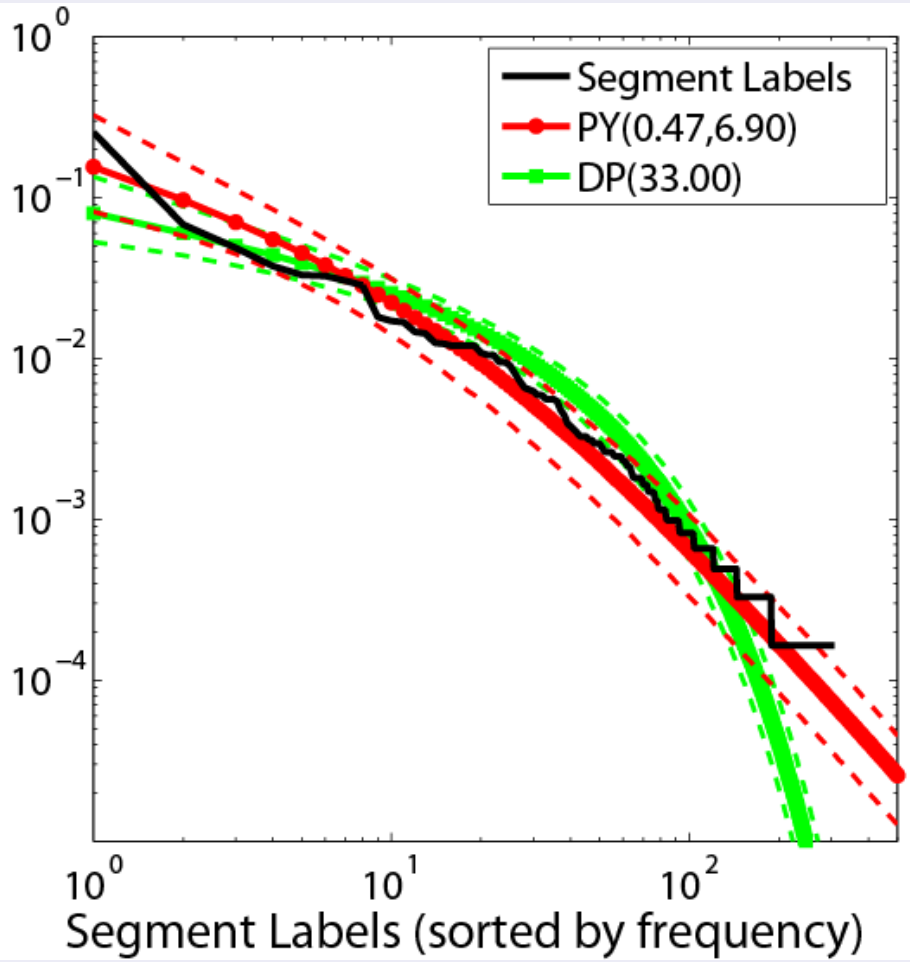




insidecity region areas ***insidecity region counts***

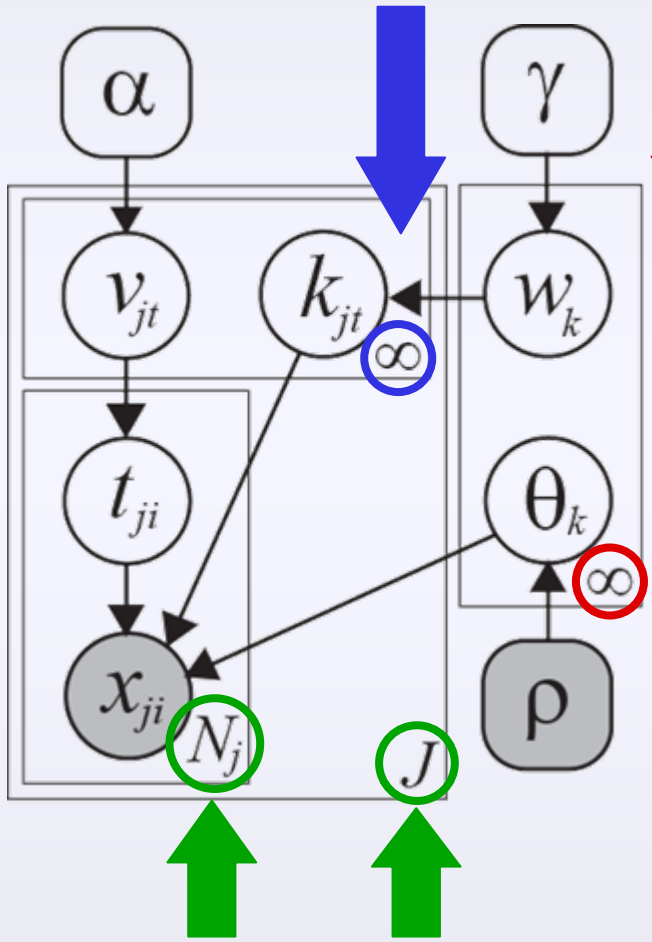
Object Name Frequencies





forest scenes ***insidecity scenes***

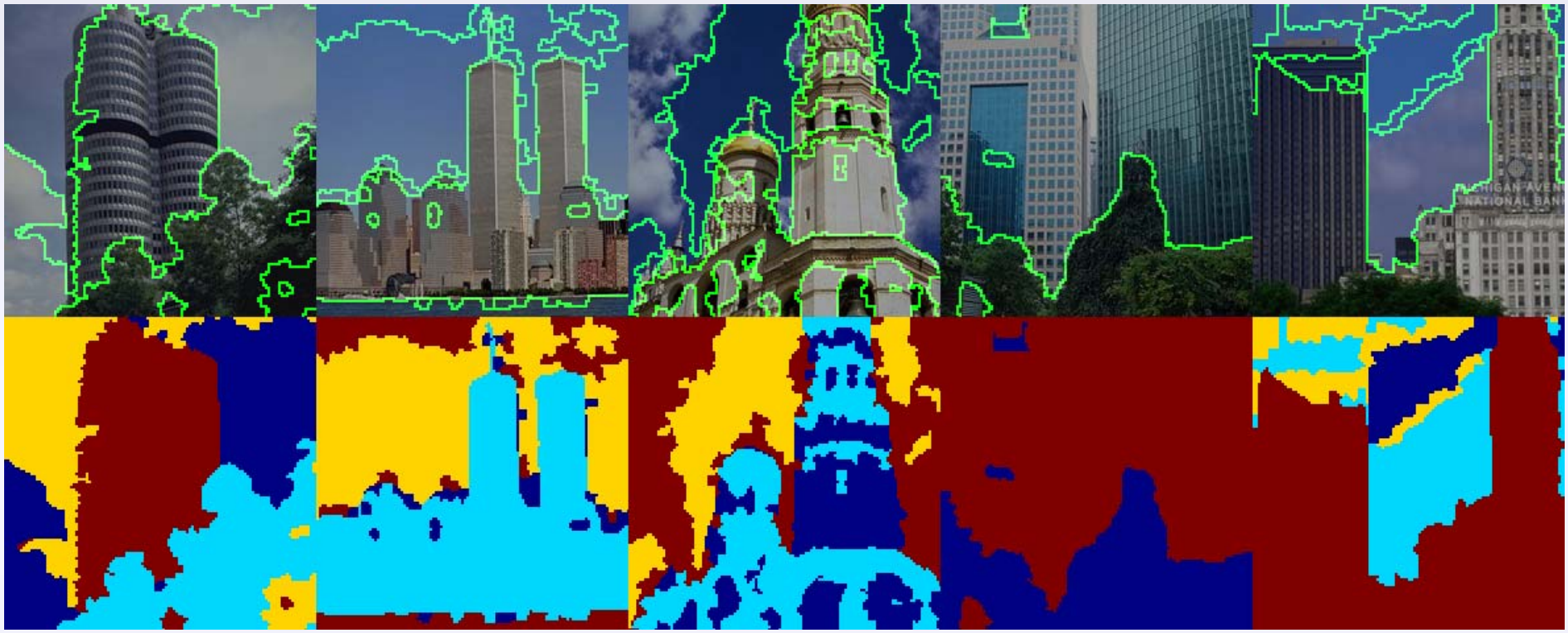
Hierarchical Pitman-Yor Model



Hierarchical DP: Teh et. al. 2004

Hierarchical PY N-gram: Teh 2006

Bag of Features Segmentation

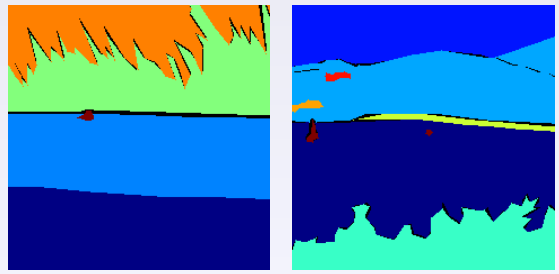


LabelMe Segments:

Outline

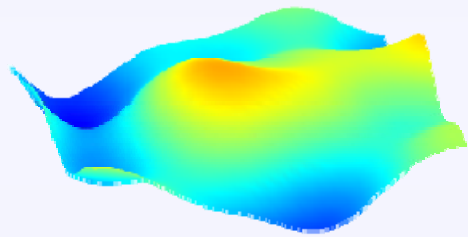
Natural Scene Statistics

- Counts, partitions, and power laws
- Hierarchical *Pitman-Yor* processes



Spatial Priors for Image Partitions

- What's wrong with Potts models?
- Spatial dependence via *Gaussian processes*



Unsupervised Image Analysis

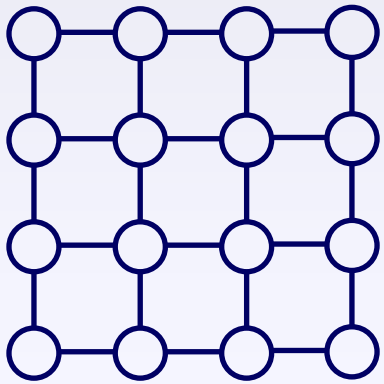
- Image *segmentation*
- Visual *category discovery*

Discrete Markov Random Fields

Ising and Potts Models

$$p(z) = \frac{1}{Z(\beta)} \prod_{(s,t)\in E} \psi_{st}(z_s, z_t)$$

$$\log \psi_{st}(z_s, z_t) = \begin{cases} \beta_{st} > 0 & z_s = z_t \\ 0 & \text{otherwise} \end{cases}$$

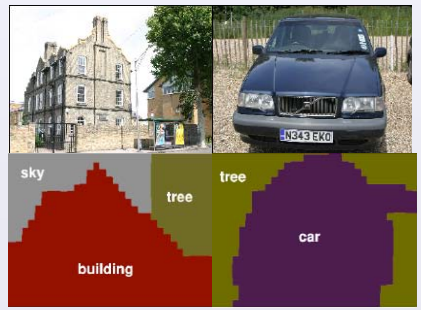


Previous Applications

- Interactive foreground segmentation
- Supervised training for known categories

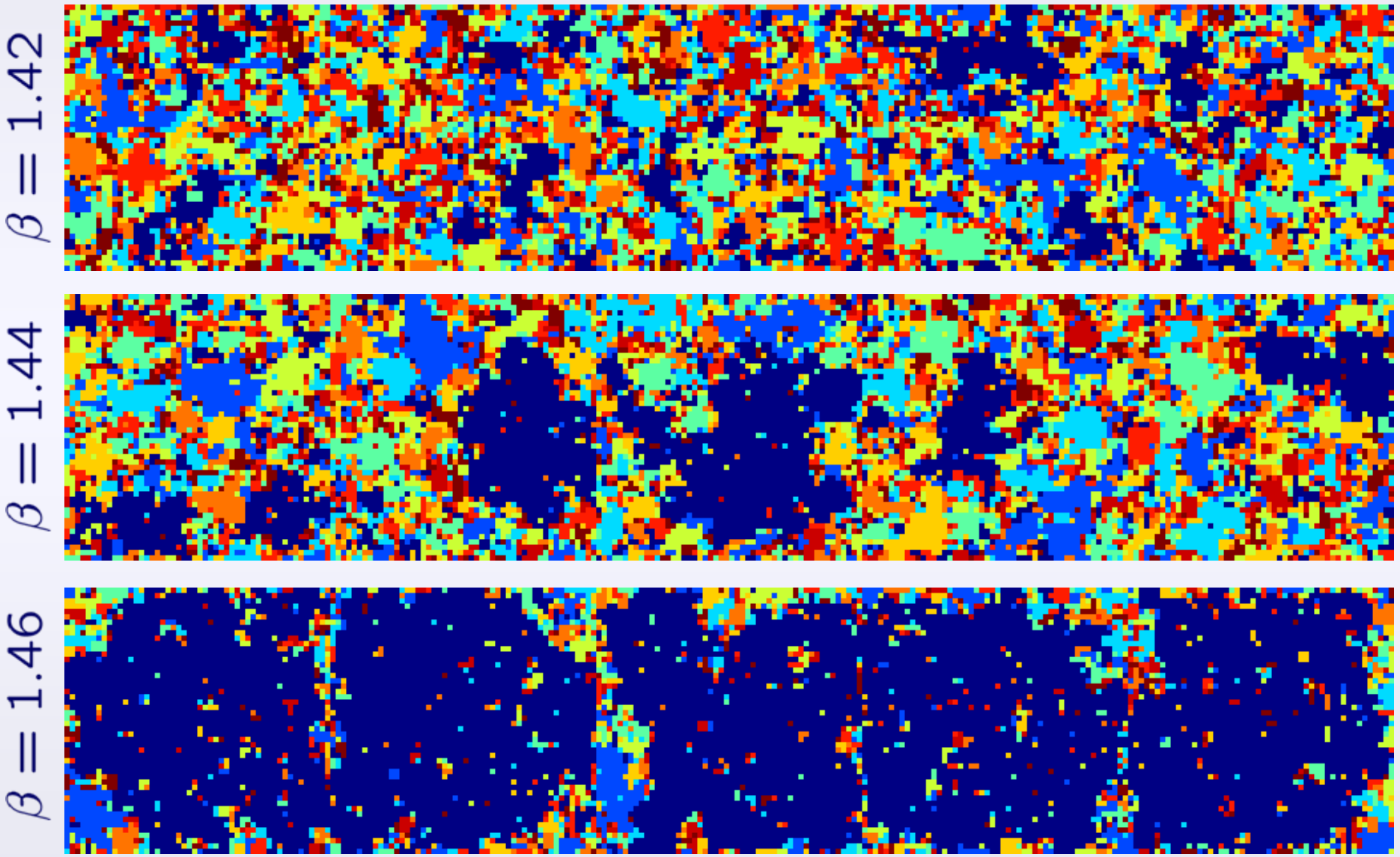
…but very little success at segmentation of unconstrained natural scenes.

GrabCut: *Rother, Kolmogorov, & Blake 2004*



Verbeek & Triggs, 2007

10-State Potts Samples



States sorted by size: largest in blue, smallest in red

1996 IEEE DSP Workshop

The Ising/Potts model is not well suited to segmentation tasks

R.D. Morris *X. Descombes*
J. Zerubia
INRIA, 2004, route des Lucioles, BP93, Sophia Antipolis Cedex, France.

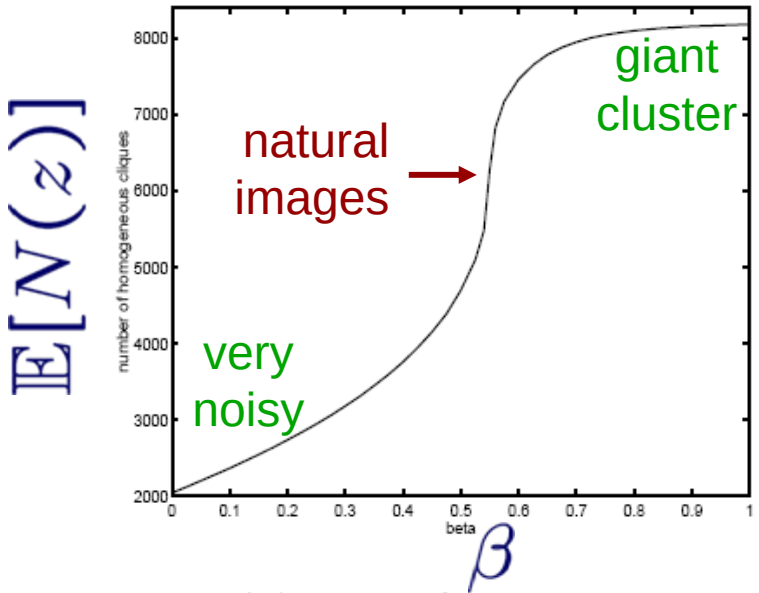


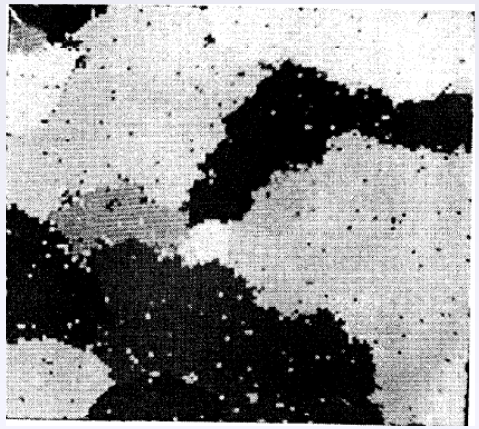
Figure 1. $< N(\mathbf{x}) >$ vs β for $64 \times 64 \times 4$-state Potts model

$N(z)$ → number of edges on which states take same value

β → edge strength

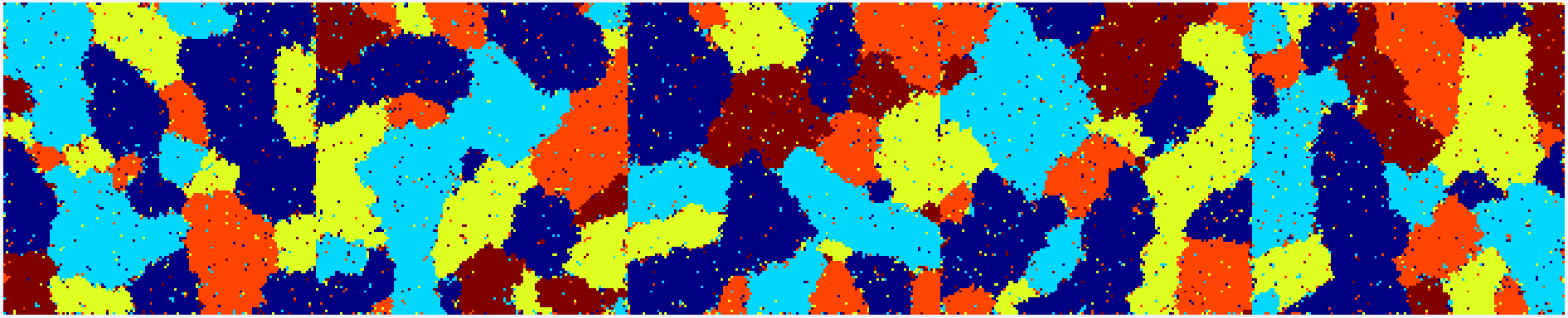
Even within the *phase transition* region, samples lack the *size distribution* and *spatial coherence* of real image segments

Geman & Geman, 1984

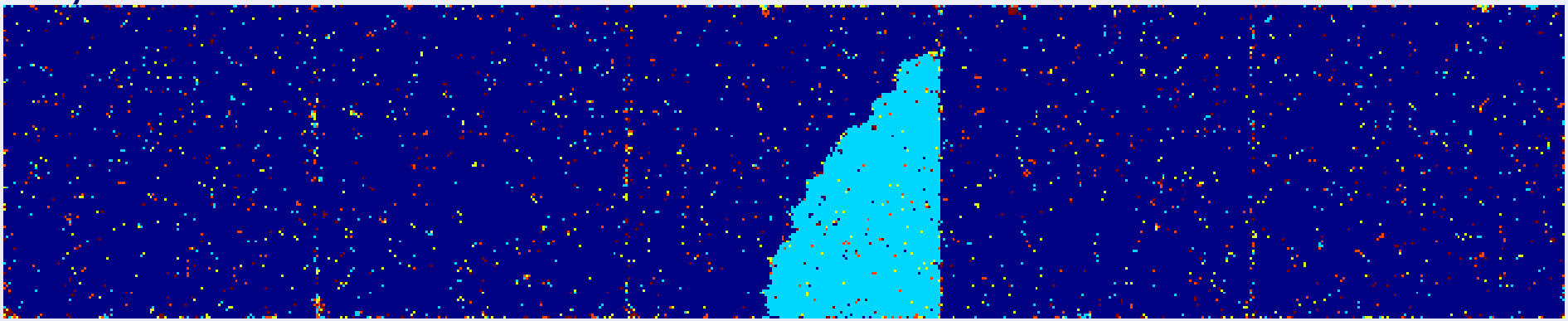


128 x128 grid
8 nearest neighbor edges
K = 5 states
Potts potentials: $\beta = 2/3$

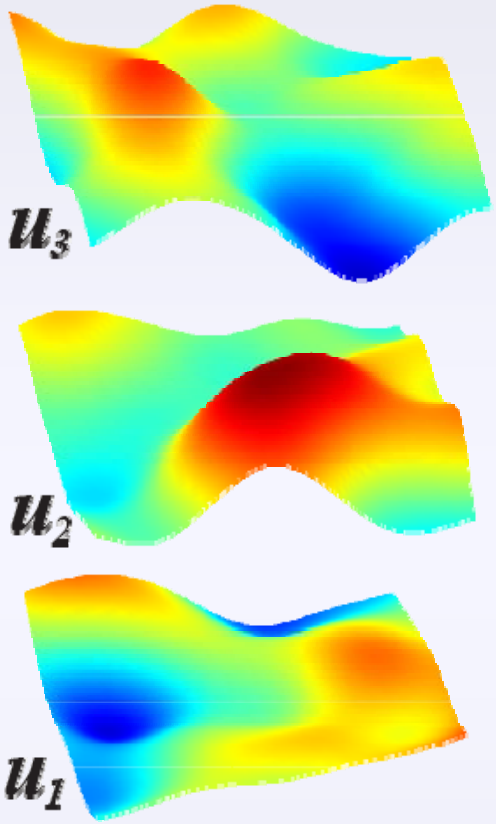
200 Iterations



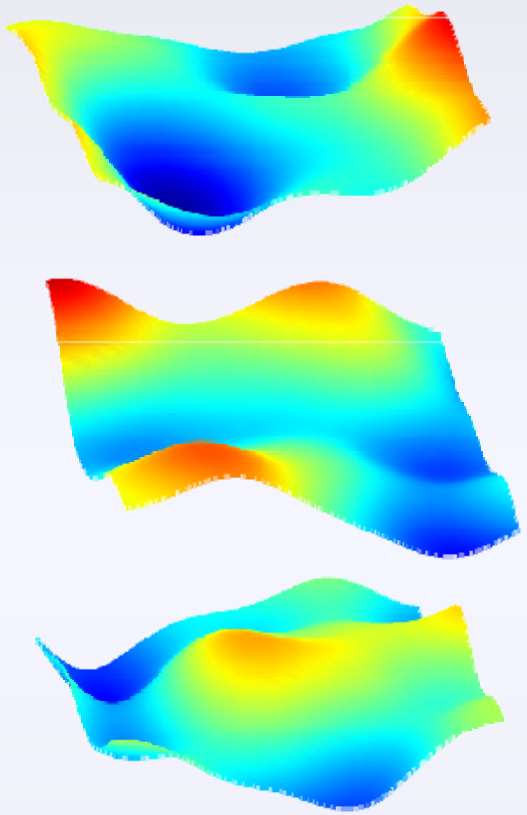
10,000 Iterations

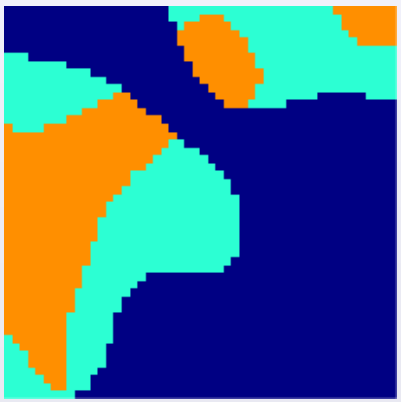


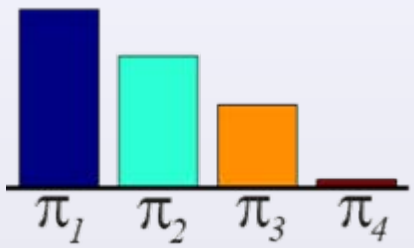
Spatially Dependent Pitman-Yor

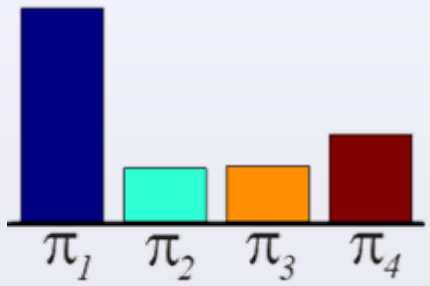


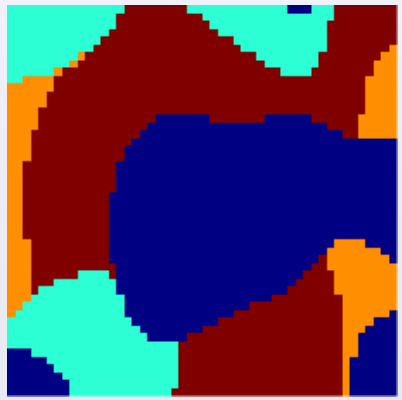
- Cut random *surfaces* (samples from a GP) with *thresholds*
  (as in Level Set Methods)
- Assign each pixel to the *first* surface which exceeds threshold
  (as in Layered Models)





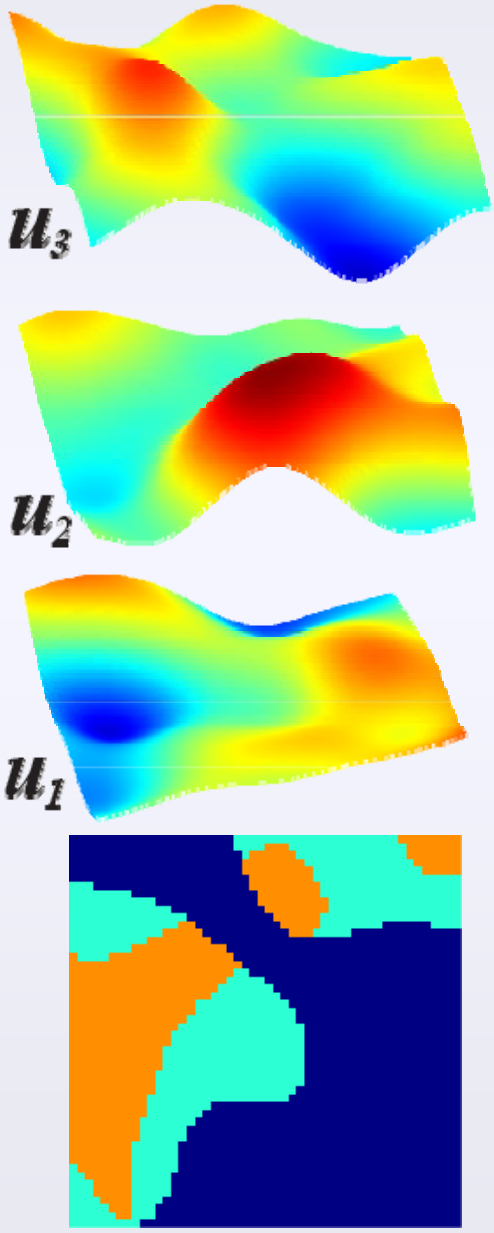


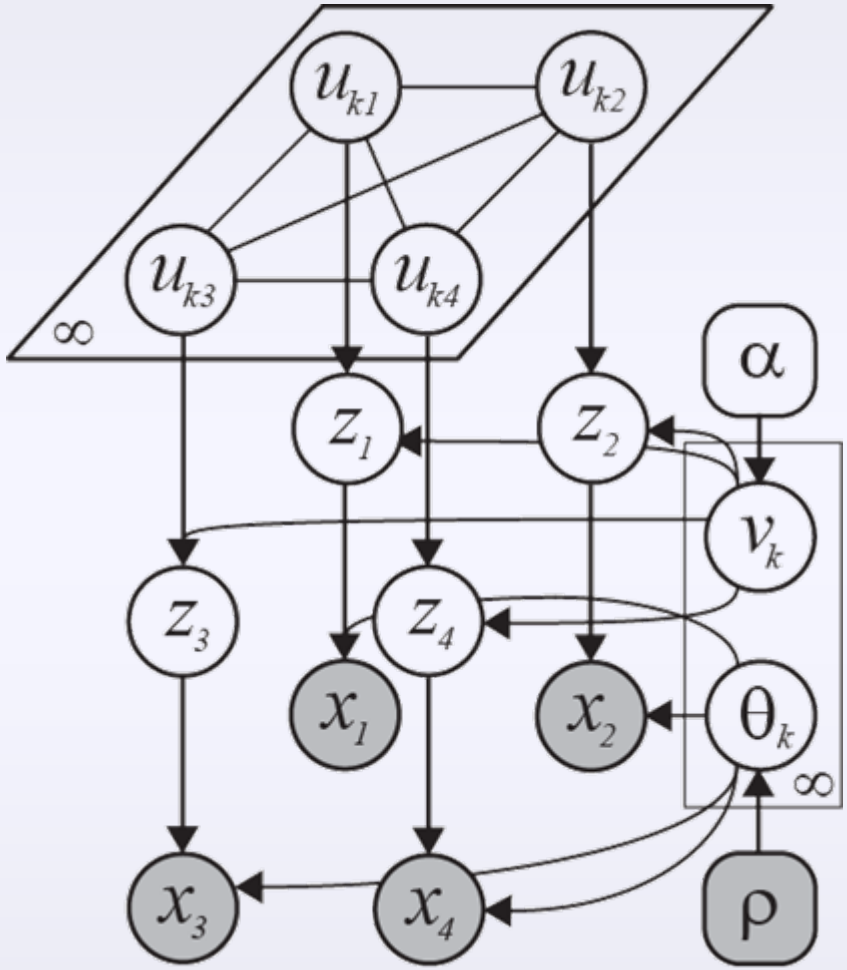




Duan, Guindani, & Gelfand,
Generalized Spatial DP, 2007

Spatially Dependent Pitman-Yor





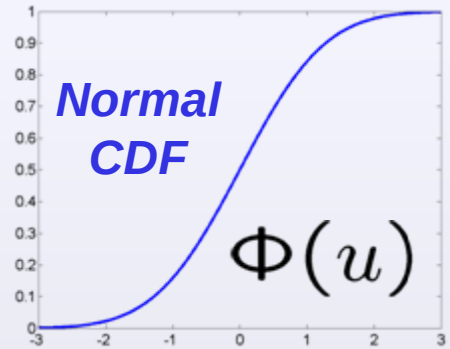
Non-Markov Gaussian Processes:

$$u_{ki} \sim \mathcal{N}(0,1)$$

$$u_{ki} \perp u_{\ell i}$$

PY prior: Segment size

$$v_k \sim \text{Beta}(1-a, b+ka)$$

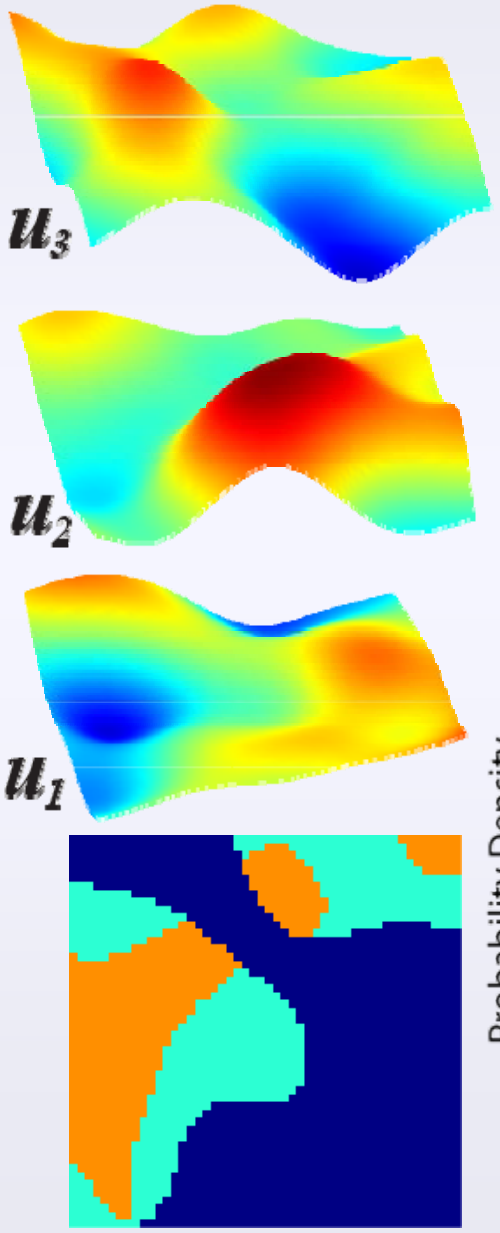


$$z_i = \min\{k \mid u_{ki} < \Phi^{-1}(v_k)\}$$

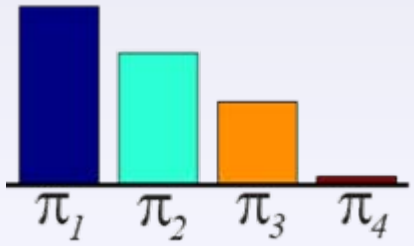
$$x_i \sim \text{Mult}(\theta_{z_i})$$

Feature Assignments

Preservation of PY Marginals



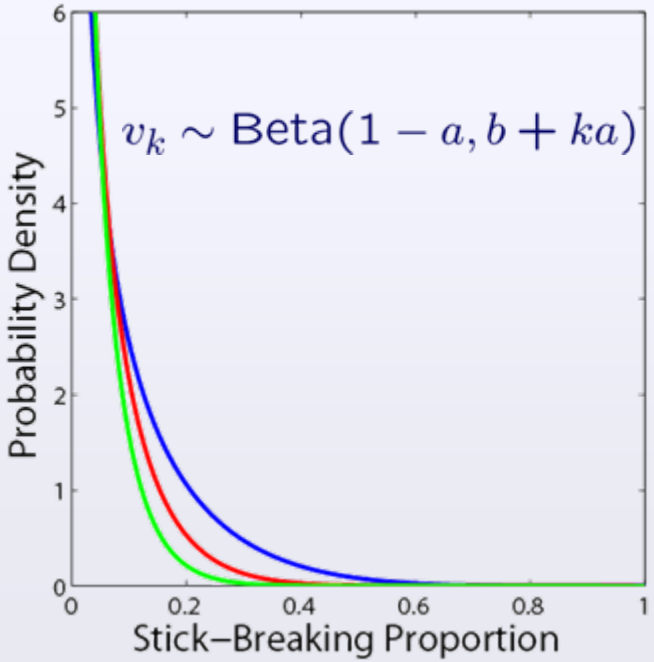
Why Ordered Layer Assignments?

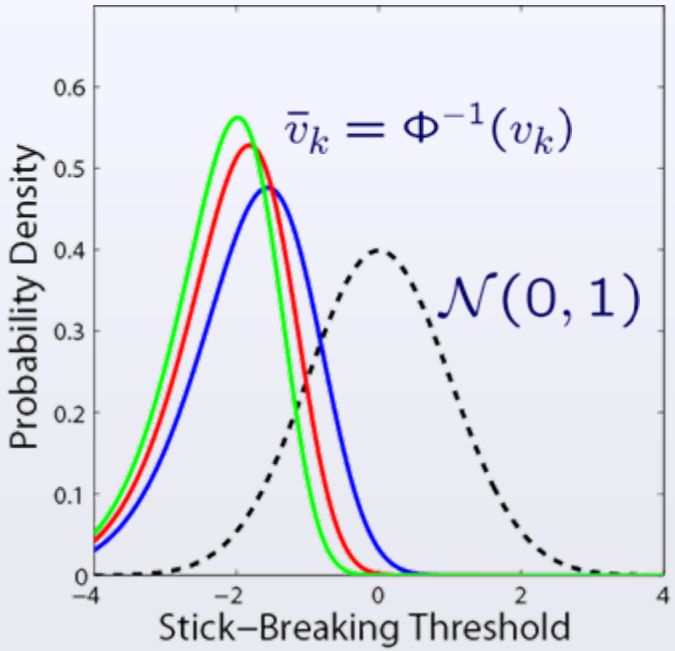


$$\pi_k = v_k \prod_{\ell=1}^{k-1} (1 - v_\ell)$$

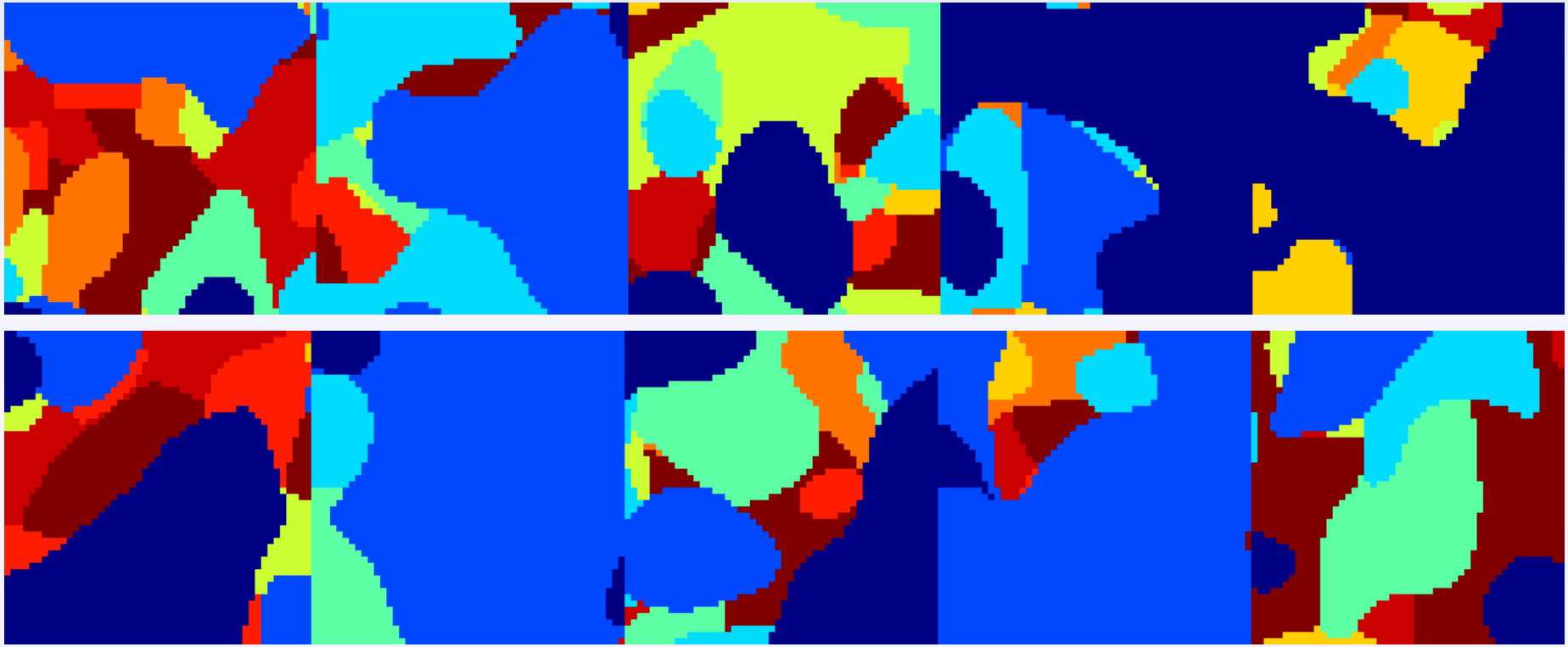
$$v_k = \mathbb{P}(z_i = k \mid z_i \neq k-1, \ldots, 1)$$

Stick Size Prior → *Random Thresholds*

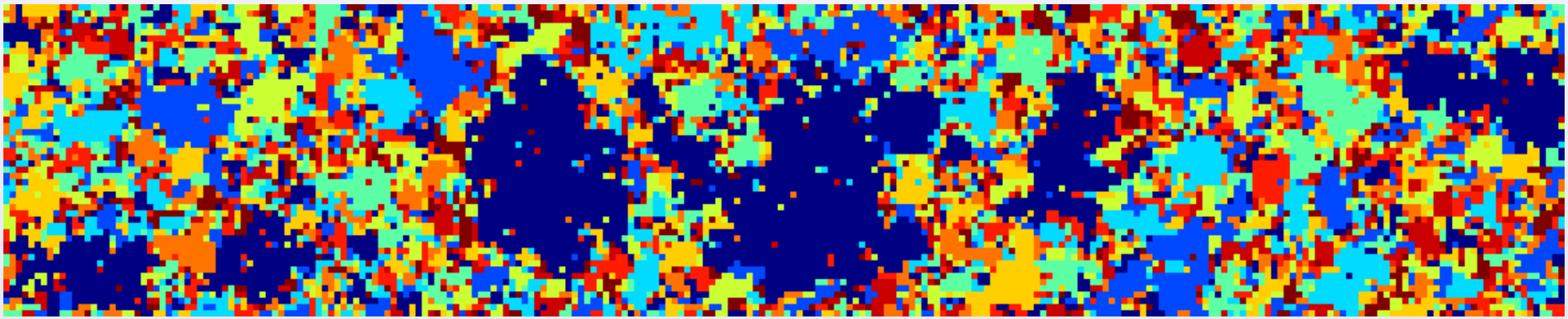




Samples from Spatial Prior



Comparison: Potts Markov Random Field



Learning & Inference

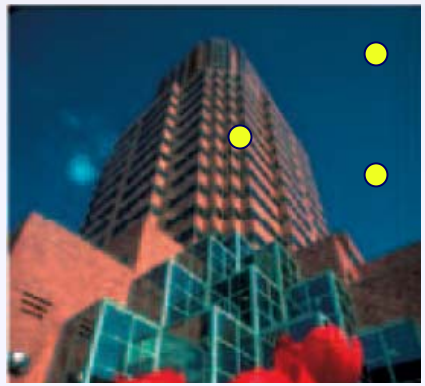
GP Covariance

$C_{ij} \Longleftrightarrow$ *probability that features at locations* (y_i, y_j) *are in the same segment*

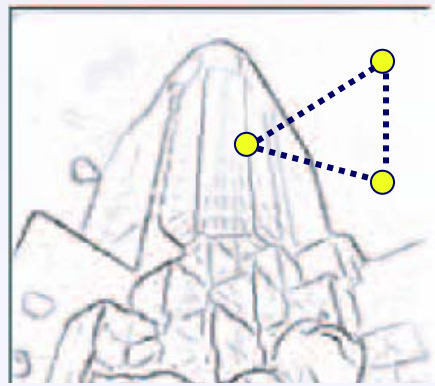
- Bag of features:

$$C_{ij} = \delta(y_i - y_j)$$

- Image distance
- Intervening countours



$$C_{ij} = e^{-\lambda(y_i - y_j)^2}$$



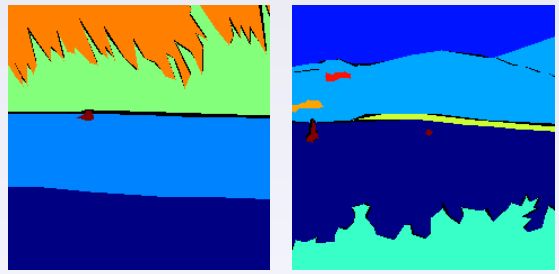
UC Berkeley Pb boundary detector

Mean Field Variational Inference

- Factorized *Gaussian* posteriors on *thresholds* & *eigenvector expansion* of dense covariance
- Jointly optimize surface & threshold via *conjugate gradient*
- Initialize by *annealing* to reduce local optima

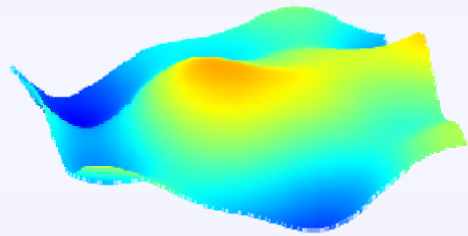
Outline

Natural Scene Statistics



- Counts, partitions, and power laws
- Hierarchical *Pitman-Yor* processes

Spatial Priors for Image Partitions

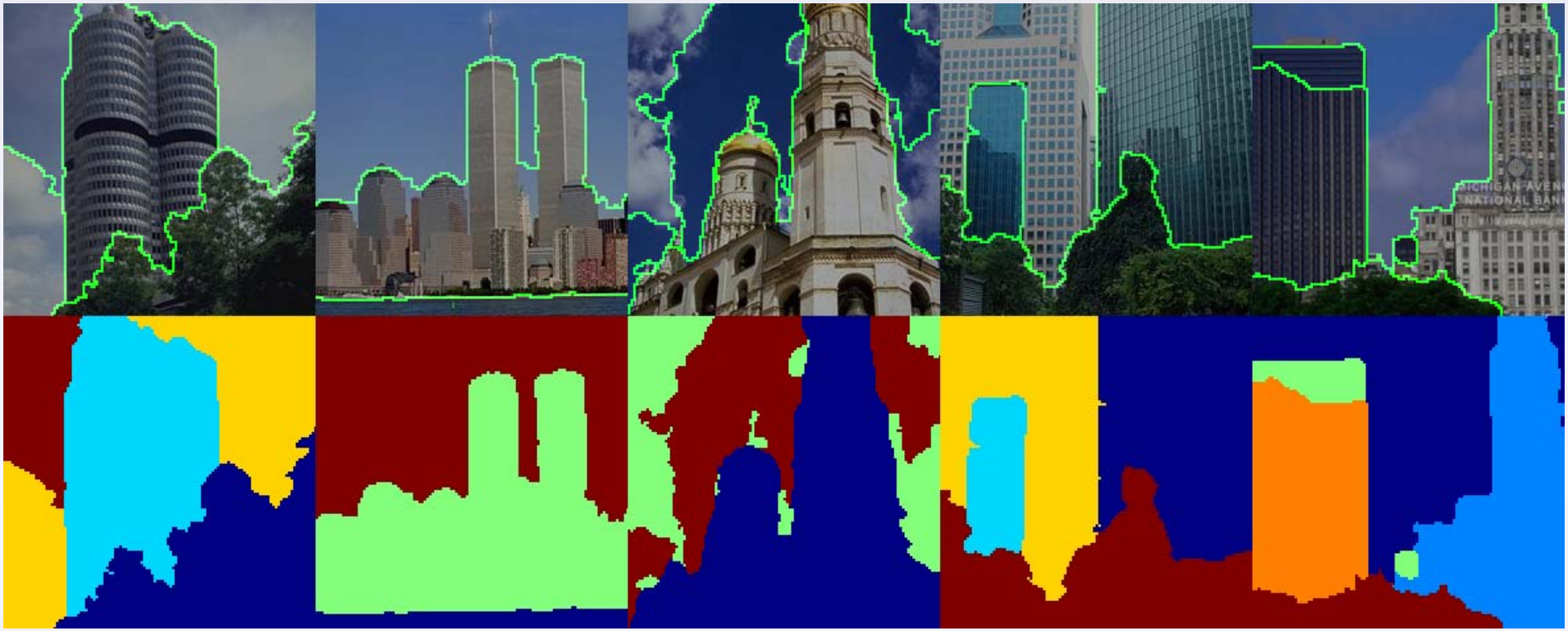


- What's wrong with Potts models?
- Spatial dependence via *Gaussian processes*

Unsupervised Image Analysis

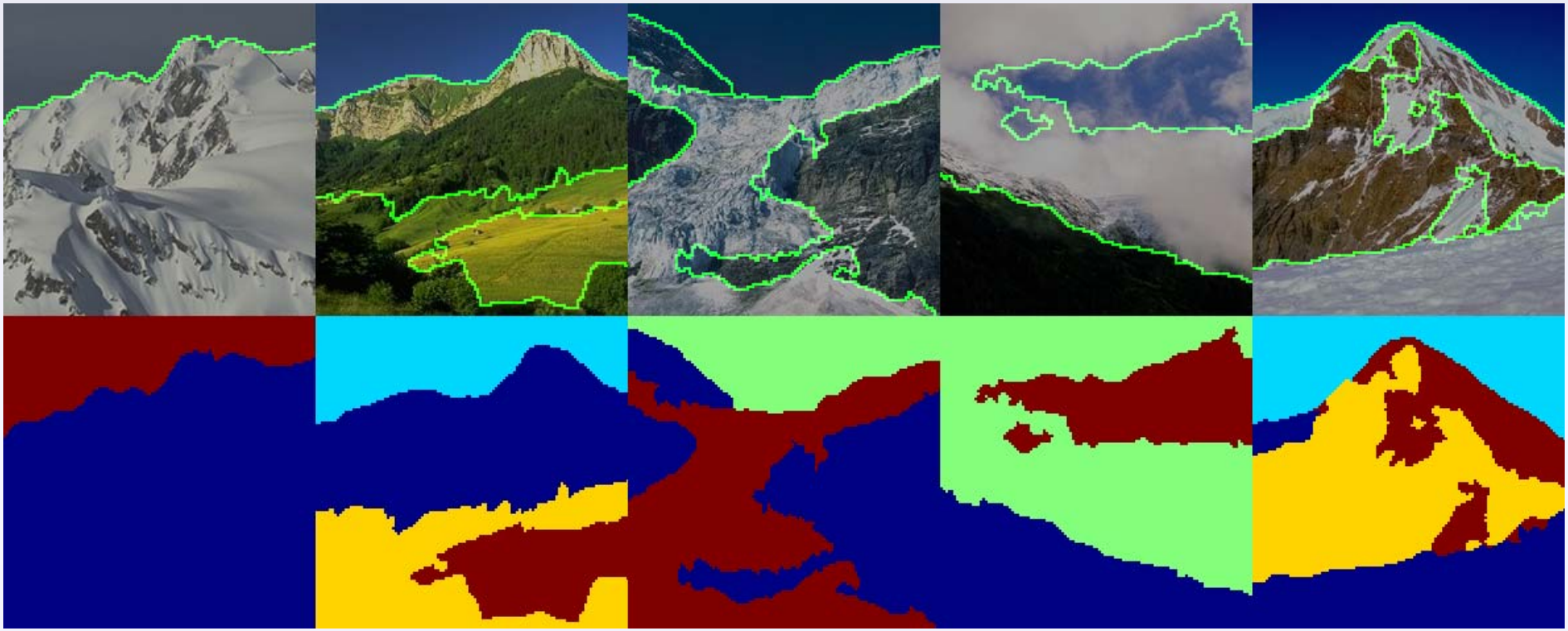
- Image *segmentation*
- Visual *category discovery*

Tallbuilding Segments: PY-Edge



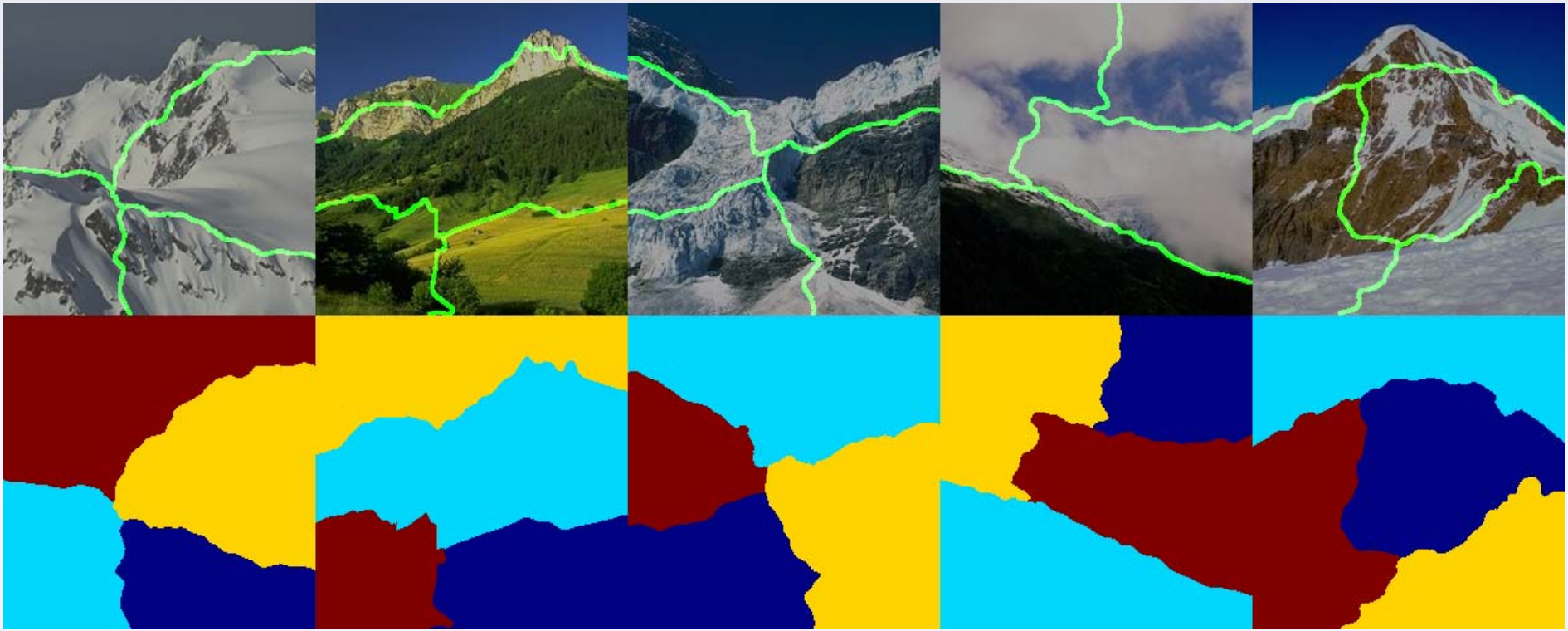
LabelMe Segments:

Mountain Segments: PY-Edge



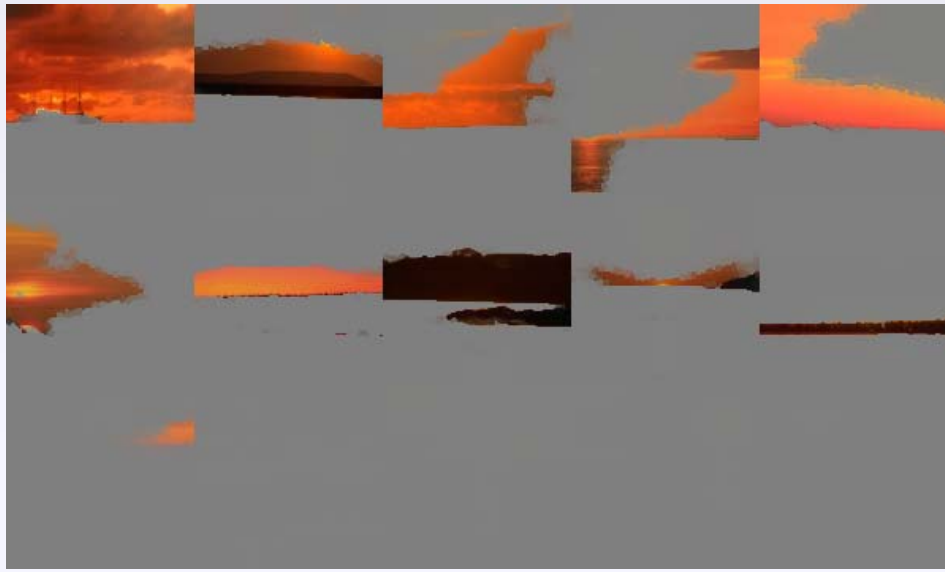
LabelMe Segments:

Mountain Baseline: NCuts

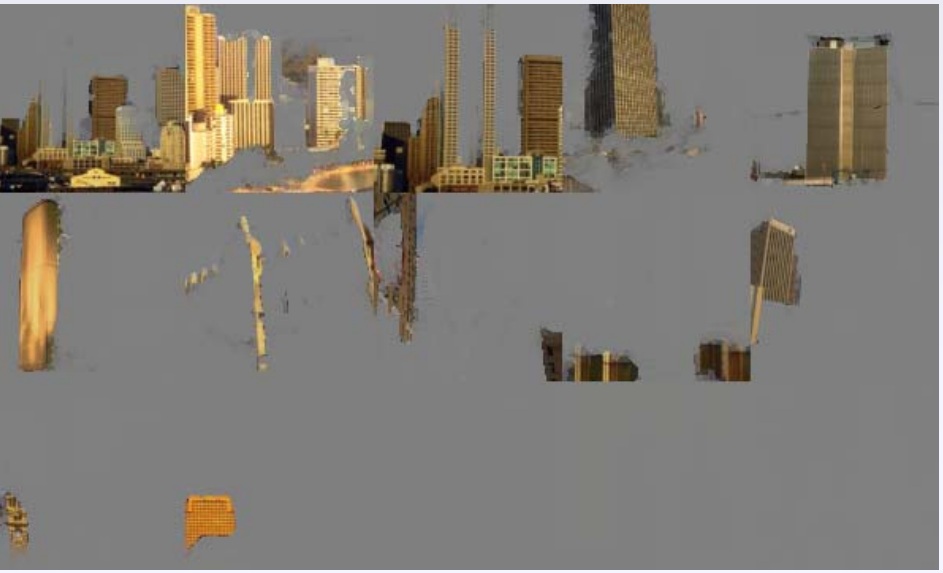


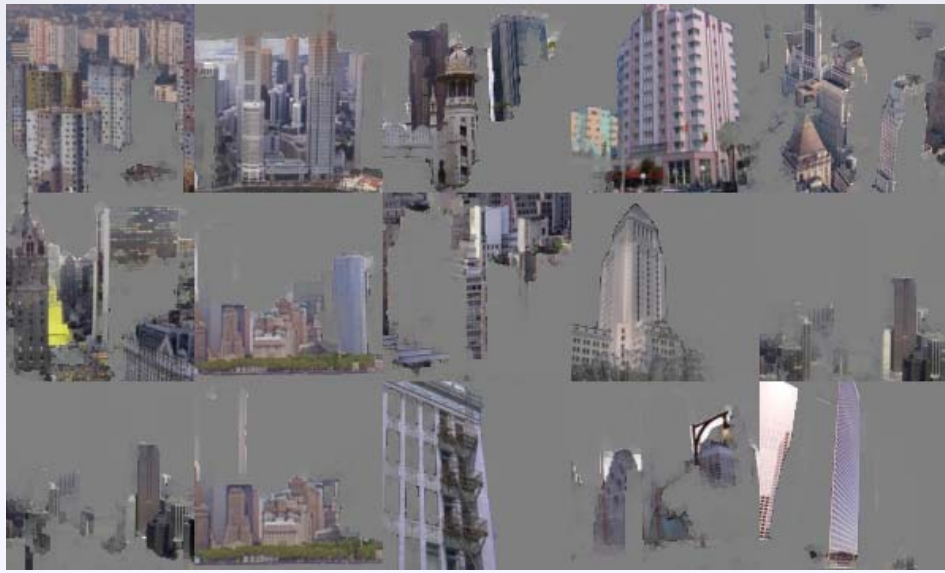
LabelMe Segments:

Visual Categories: Coast

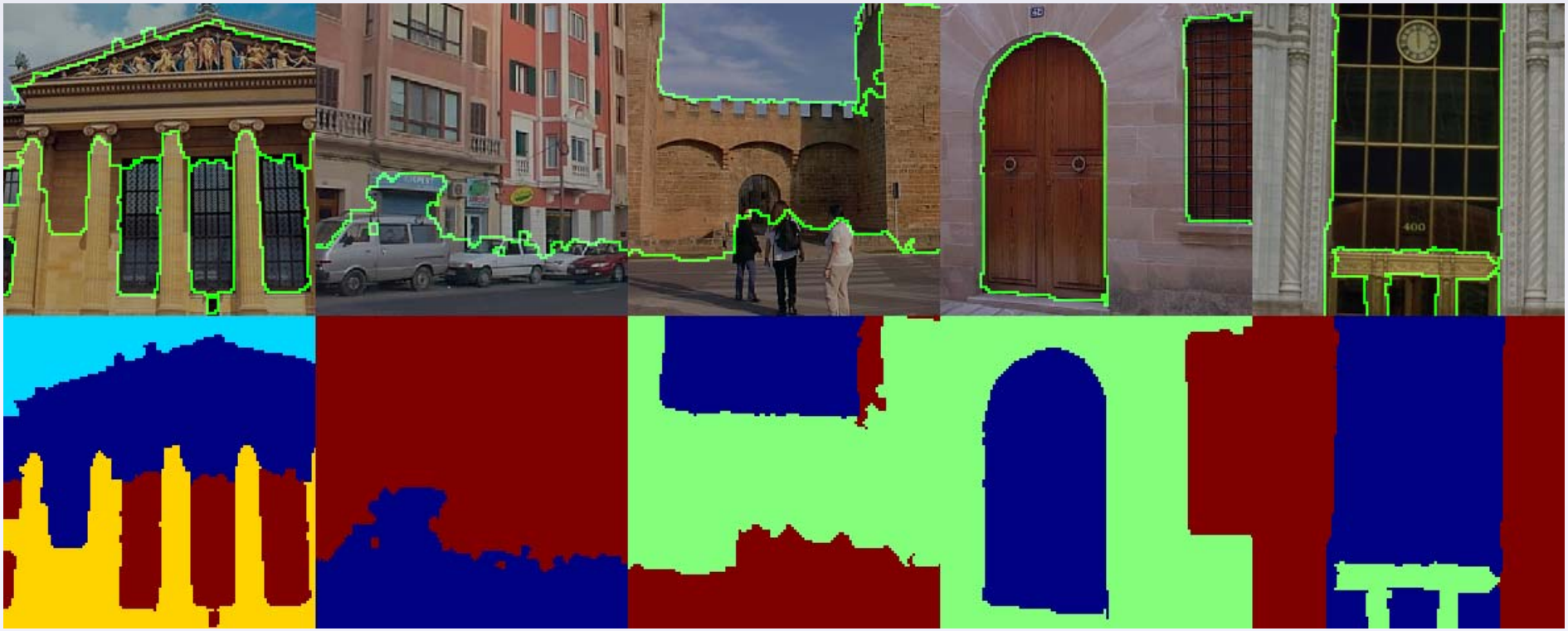


Visual Categories: Tallbuilding





Challenge: Structured Objects

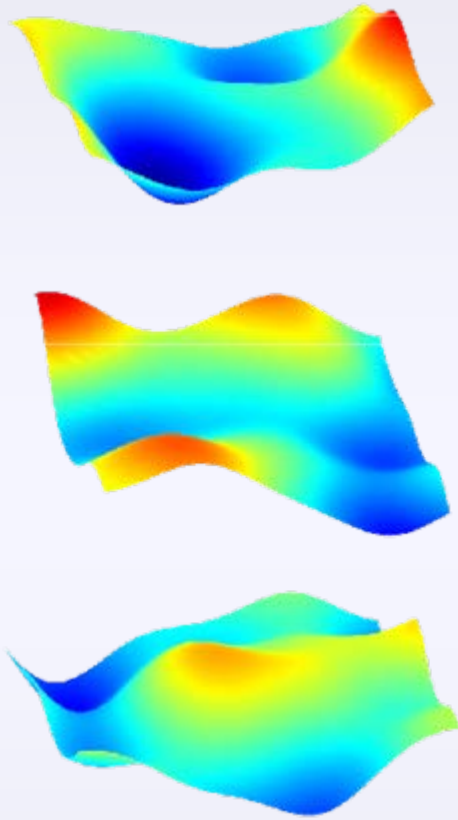


LabelMe Segments:

Conclusions

Dependent Pitman-Yor Processes allow…

- efficient variational *parsing* of scenes into unknown numbers of segments
- empirically justified *power law* priors
- learning of *shared appearance models* from related images & scenes



Future Directions

- parallelized, scalable learning from extremely *large image databases*
- nonparametric models of dependency in *other application domains*