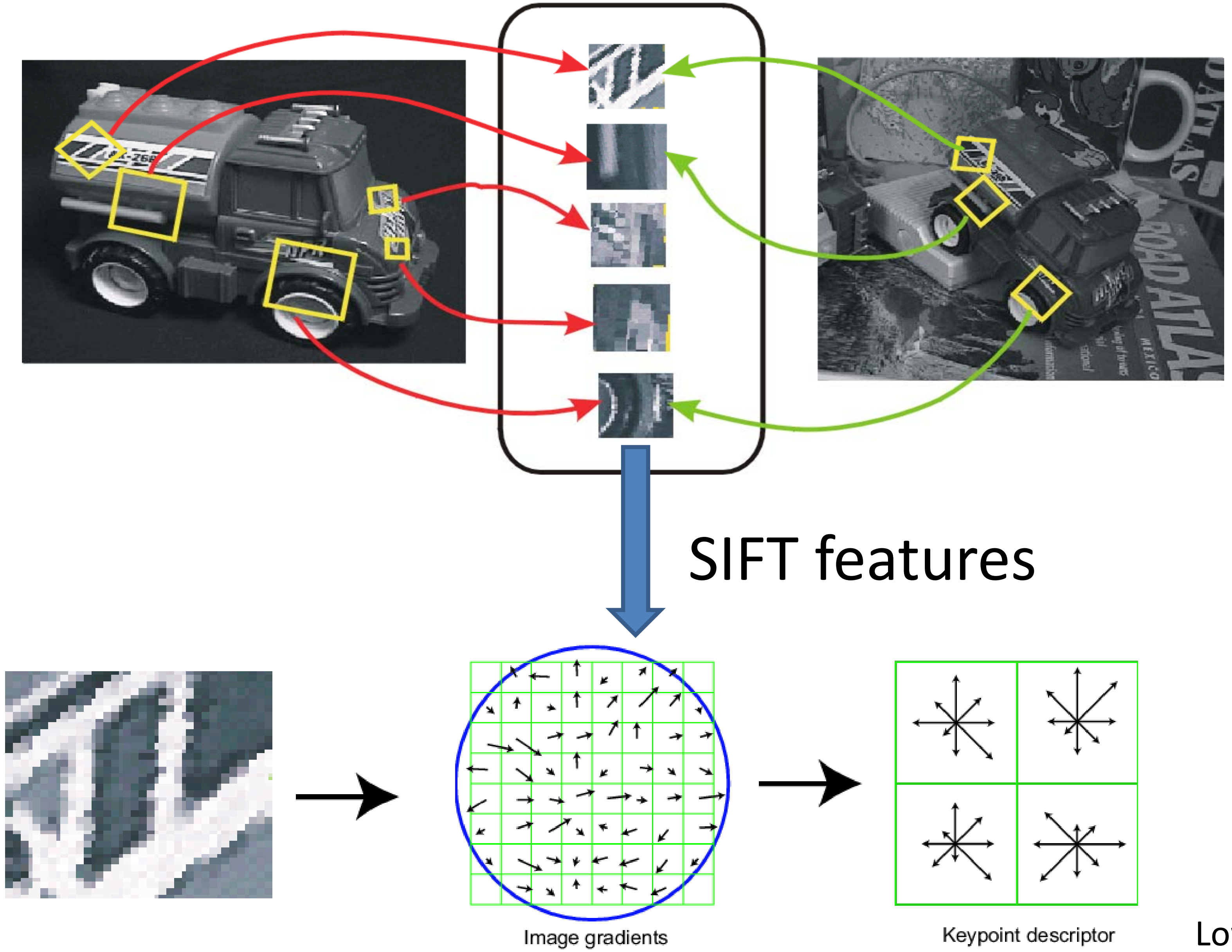
Matching single objects...



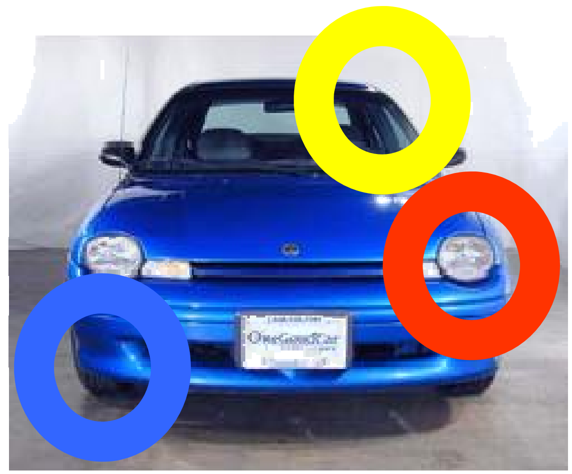
Lowe, 1999

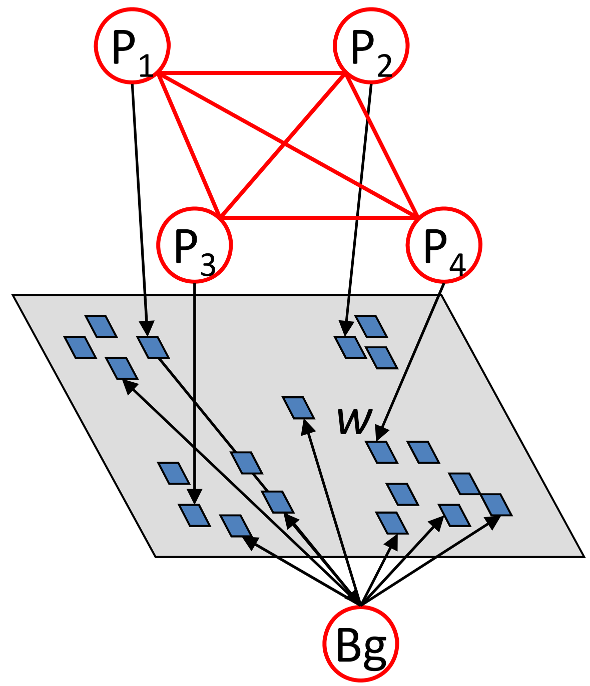
Matching a class of objects...

Part-based representation: constellation model

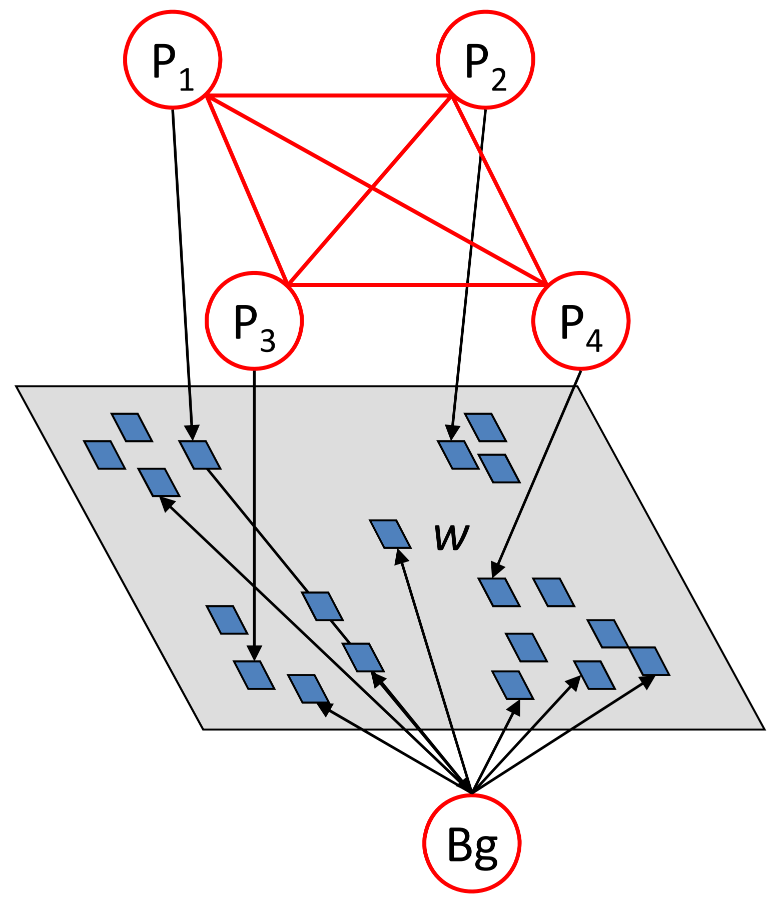
- Fischler & Elschlager 1973

- Yuille '91
- Brunelli & Poggio '93
- Lades, v.d. Malsburg et al. '93
- Cootes, Lanitis, Taylor et al. '95
- Amit & Geman '95, '99
- Perona et al. '95, '96, '98, '00, '03, '04, '05
- Felzenszwalb & Huttenlocher '00, '04
- Crandall & Huttenlocher '05, '06
- Leibe & Schiele '03, '04
- Many papers since '00

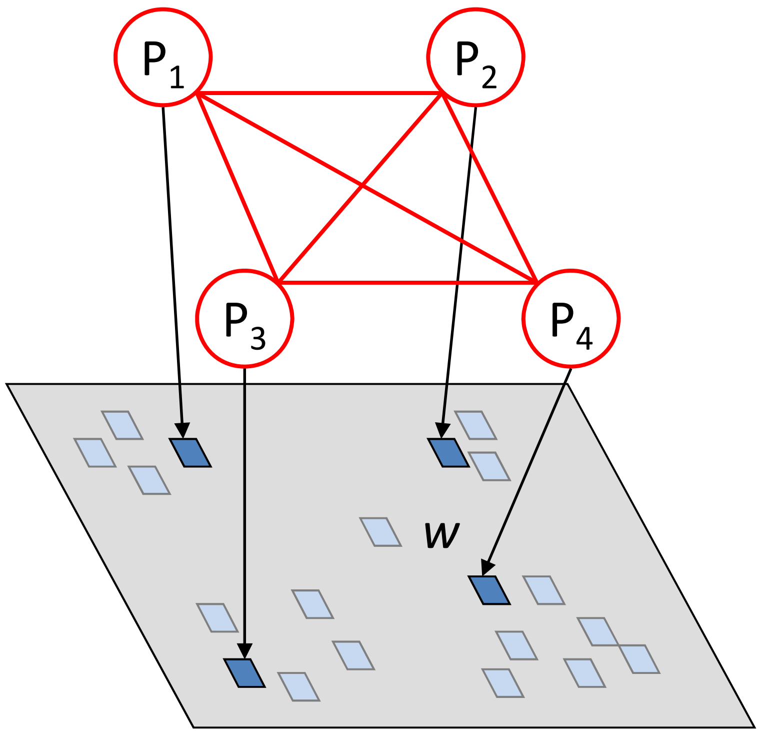




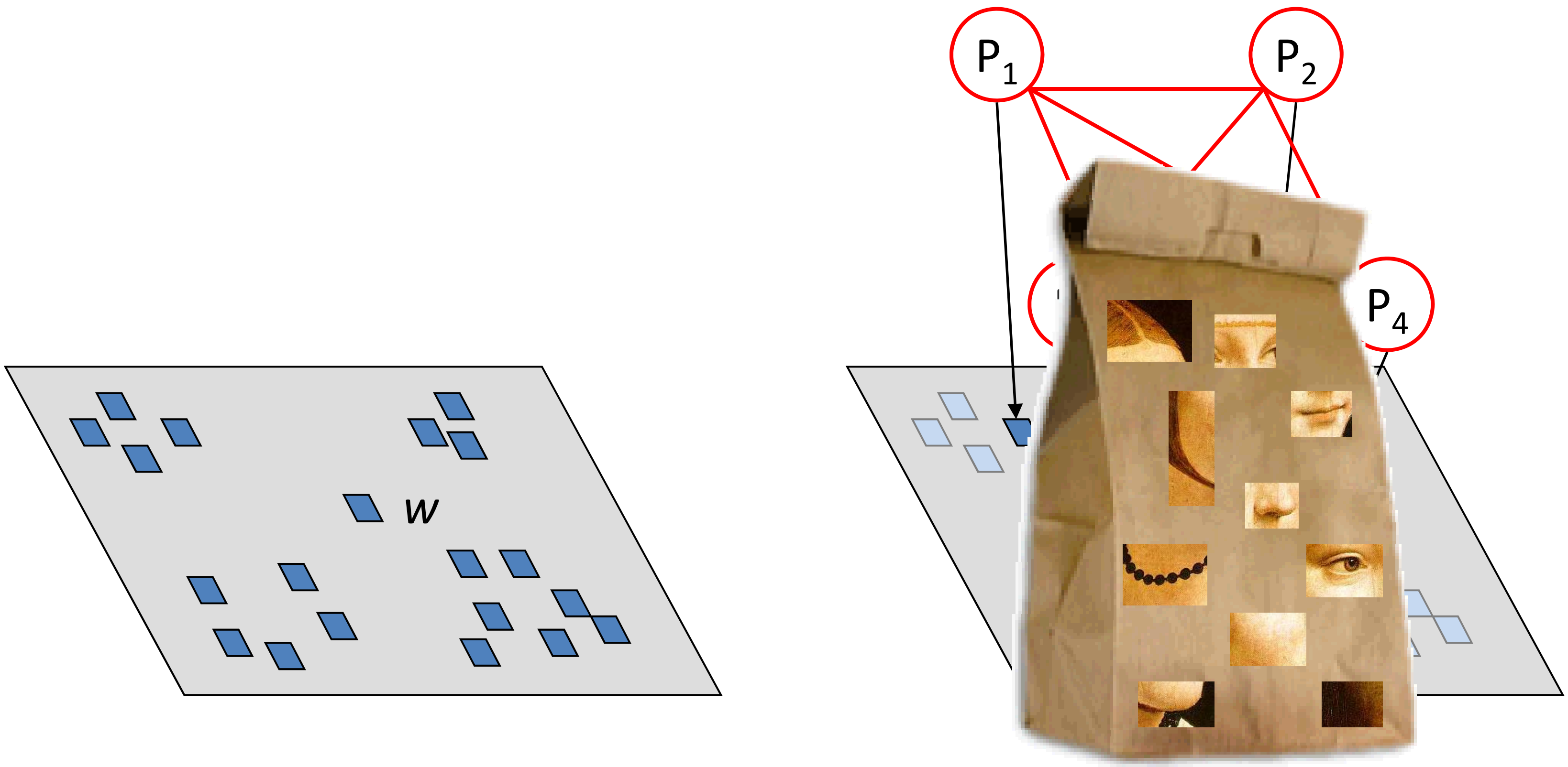
Part-based representation: constellation model



Part-based representation: constellation model

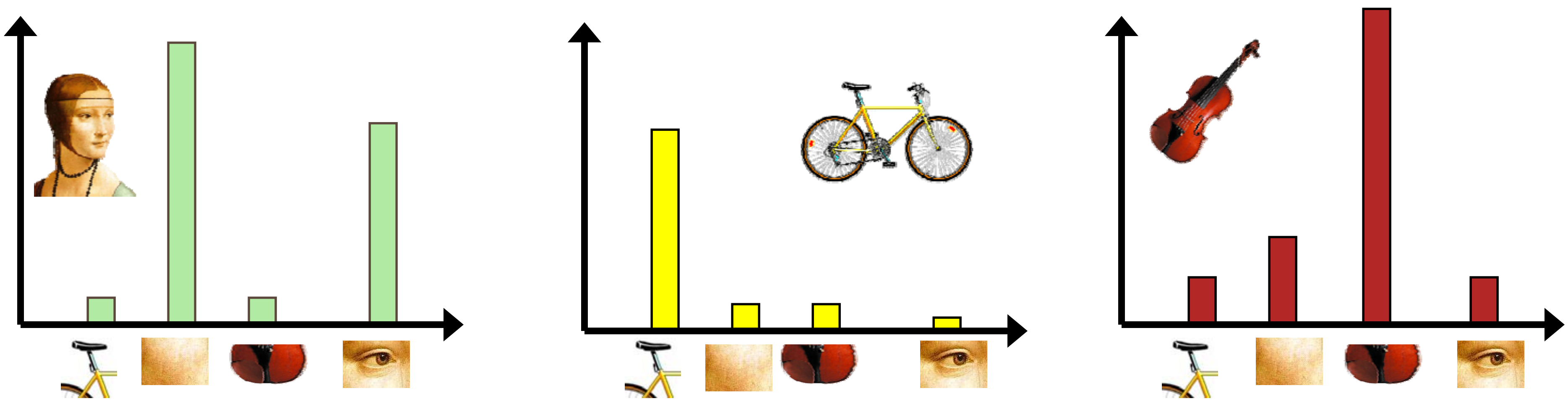


"Bag of words" representation



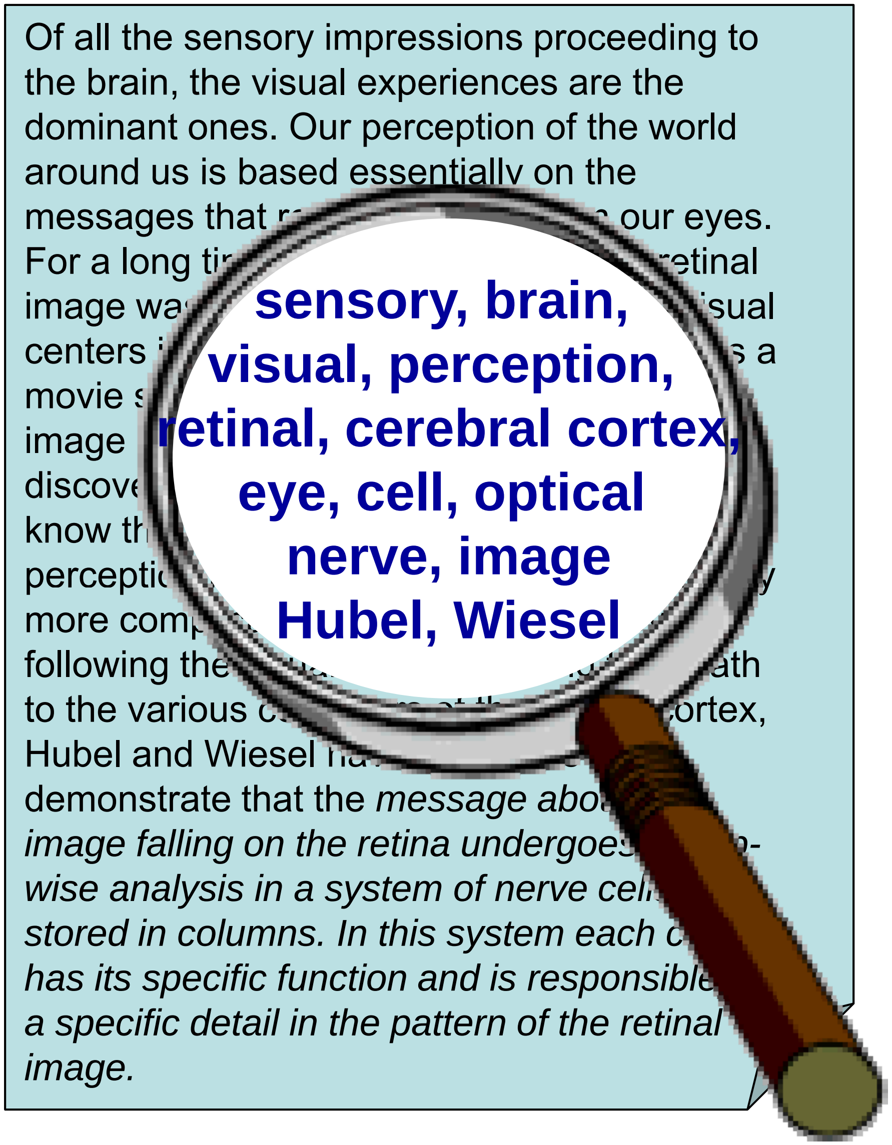
Csurka et al. 2004; earlier work in texture: Leung & Malik, 1999

“Bag of words” representation

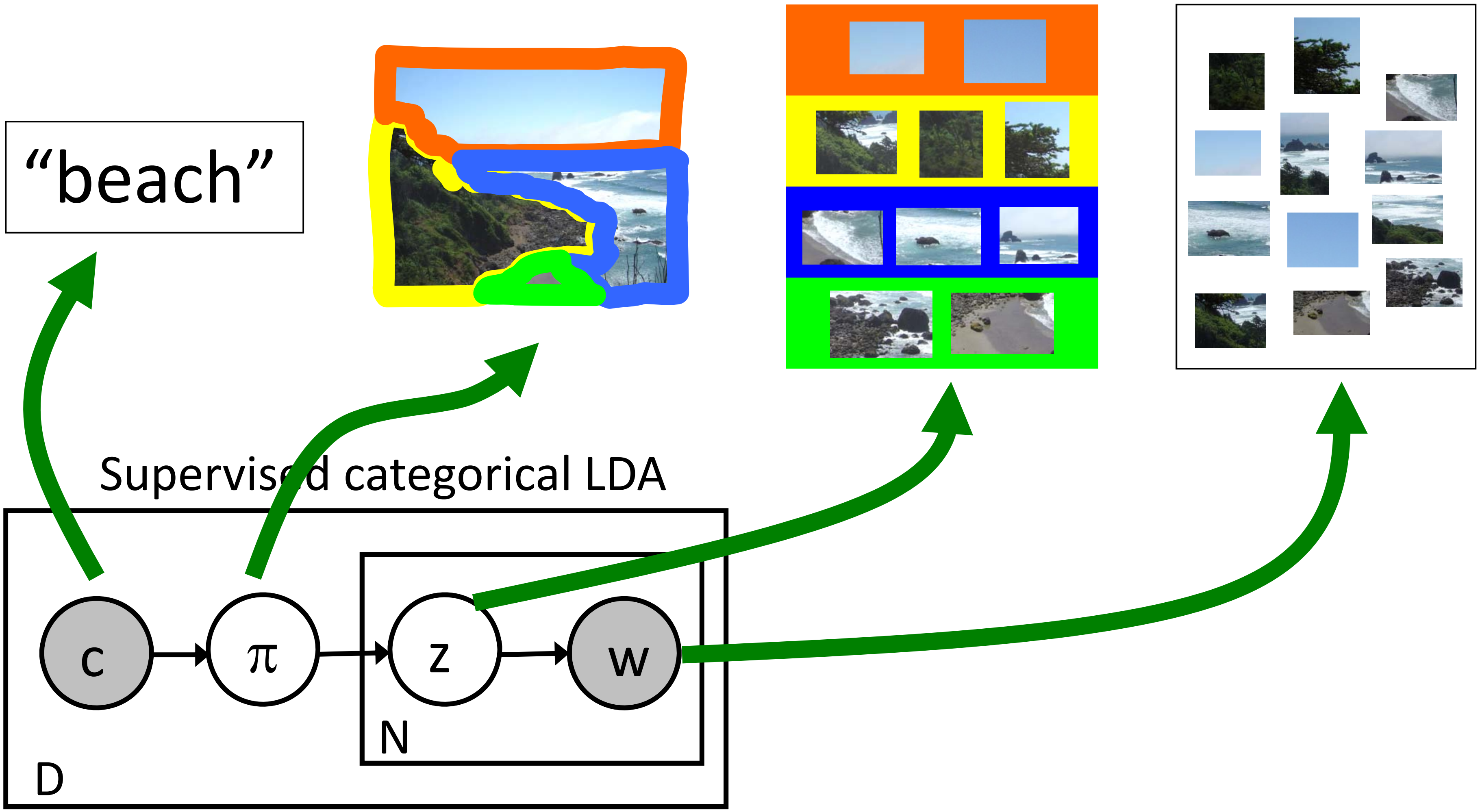


Csurka et al. 2004; earlier work in texture: Leung & Malik, 1999

Analogy to textual documents

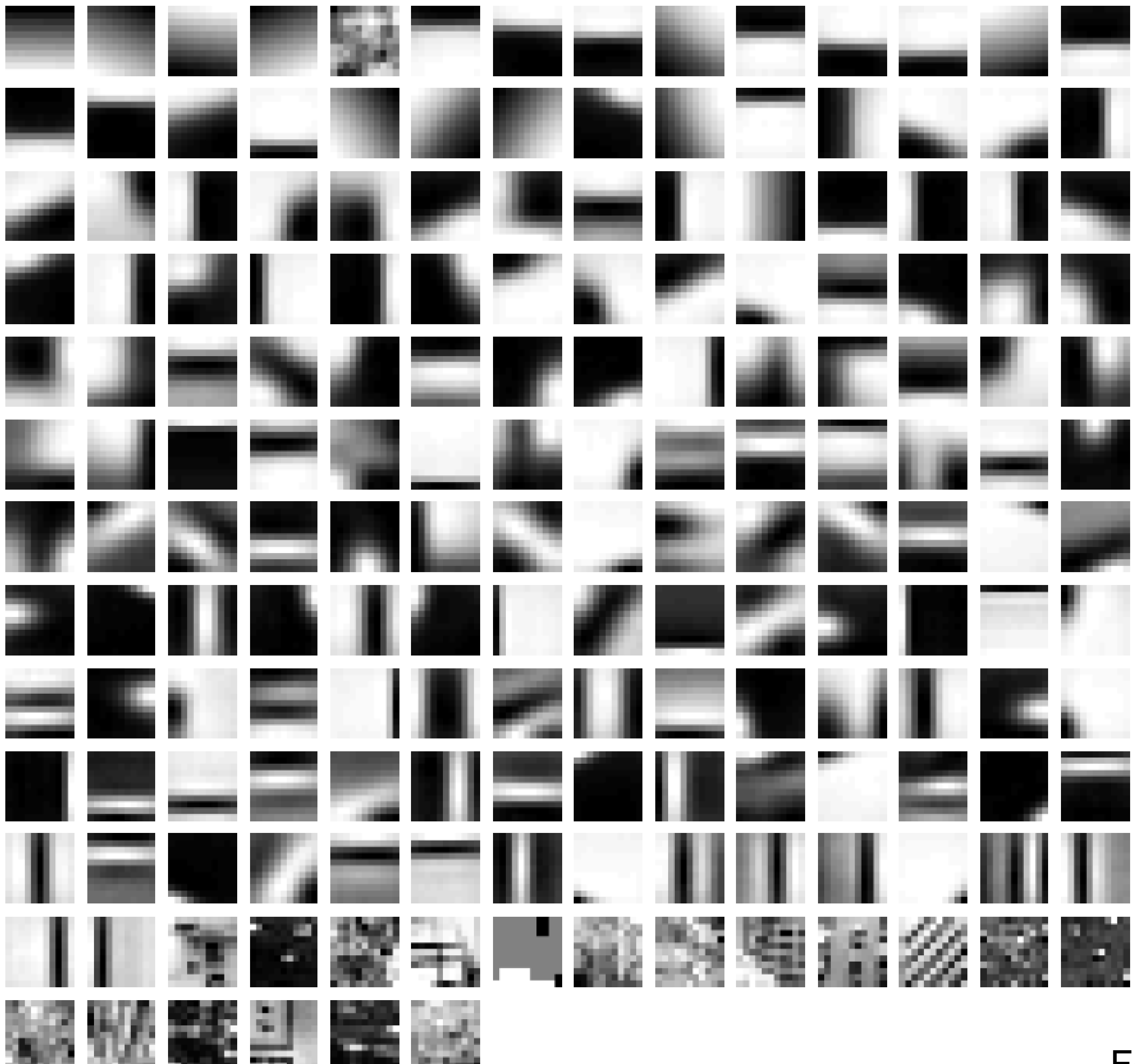


Natural scene categorization



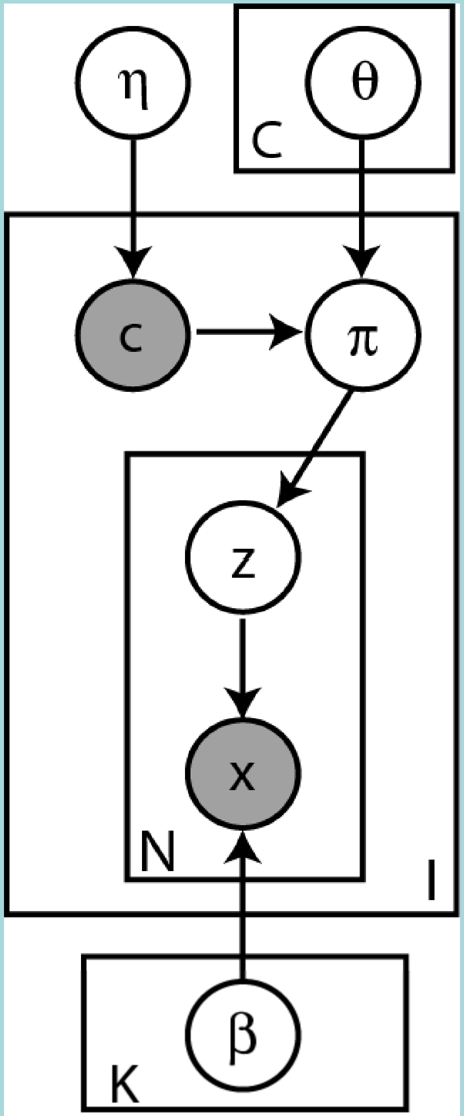
Fei-Fei et al. CVPR 2005

codebook

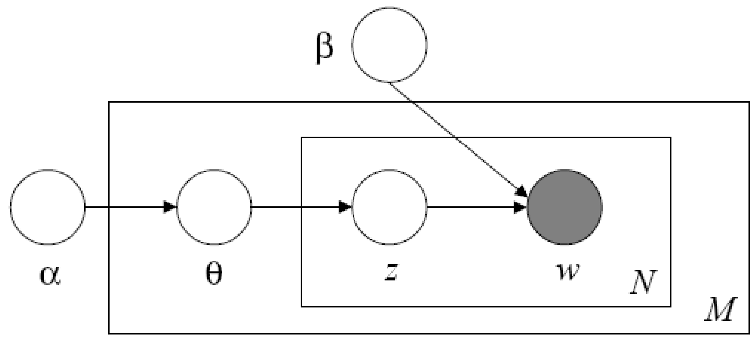


Fei-Fei et al. 2005

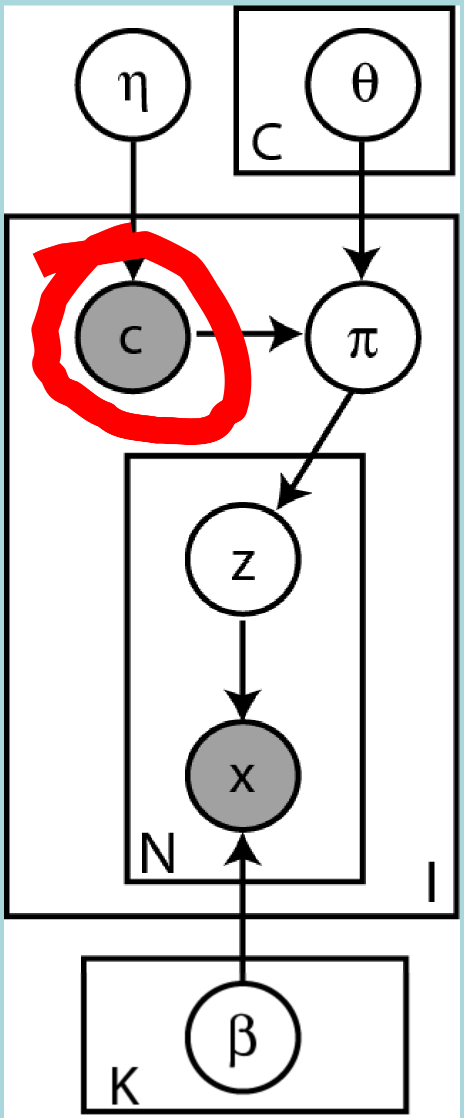
A Generative Model



LDA: Blei, Ng, & Jordan. 2003

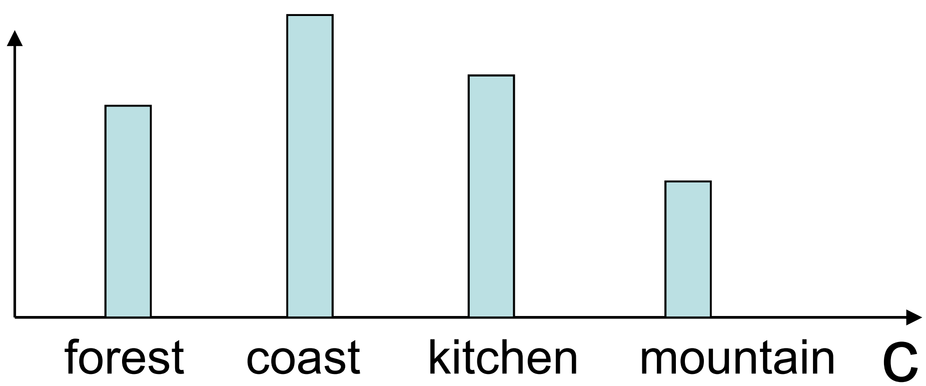


A Generative Model



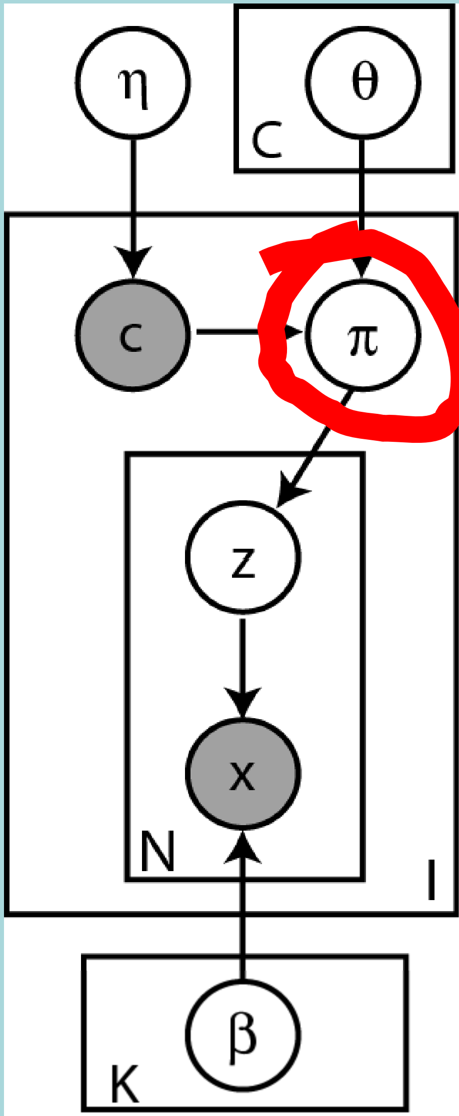
scene category

discrete variable: $c \sim p(c|\eta)$



Fei-Fei & Perona (CVPR 2005)

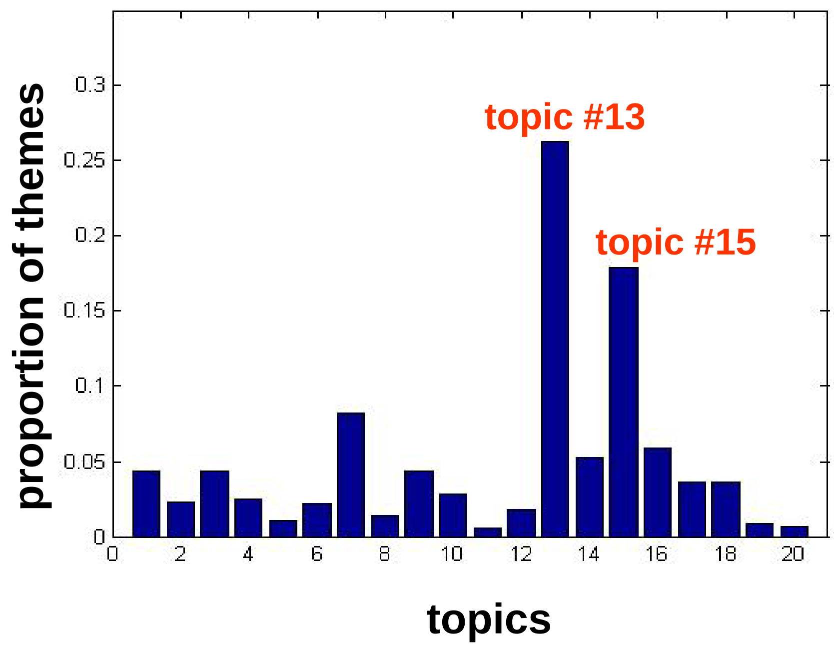
A Generative Model



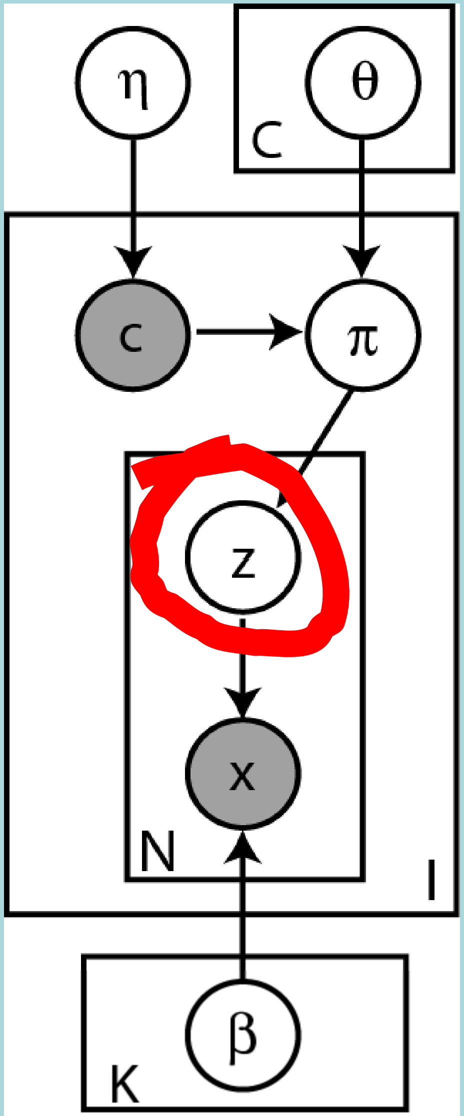
mixing parameter for the latent topics

$$\pi \sim p(\pi | c, \theta)$$
$$\sim Dir(\pi | c, \theta)$$

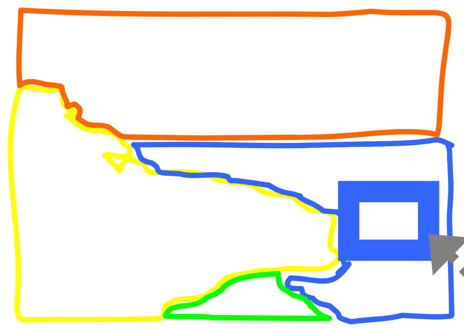
where $\sum_{k=1}^{K} \pi_k = 1$



A Generative Model



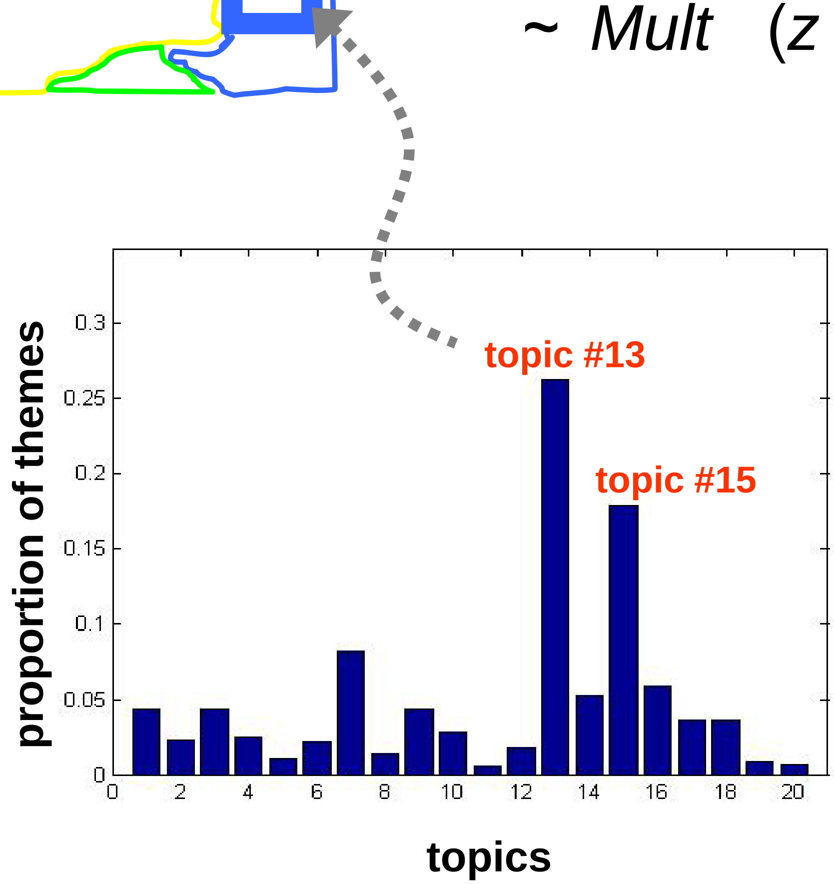
topic label



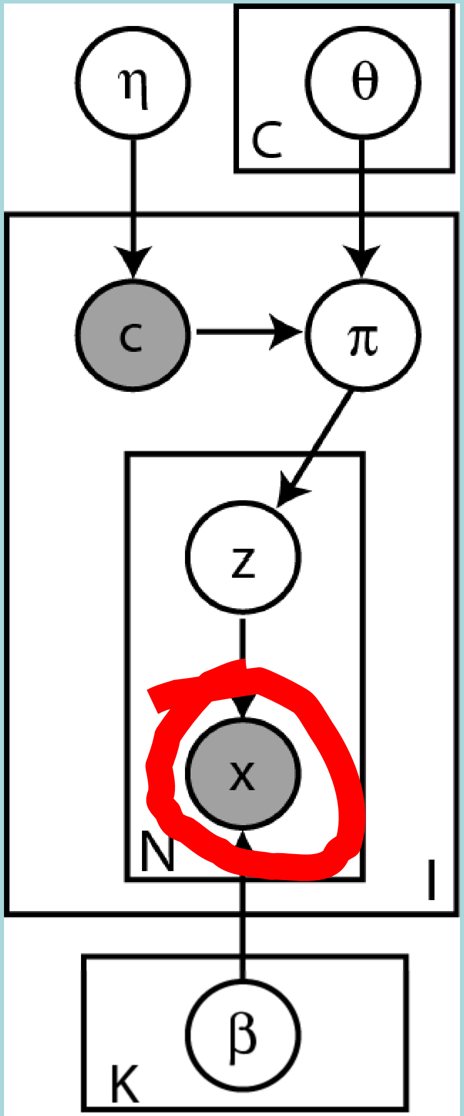
discrete variable:

$$z \sim p(z \mid \pi)$$

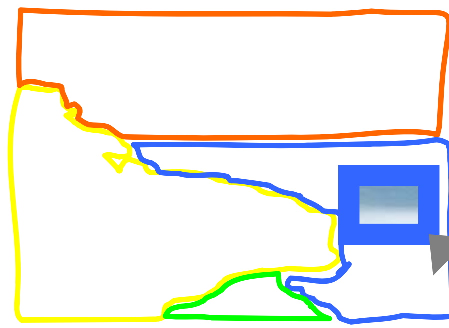
$$\sim Mult(z \mid \pi)$$



A Generative Model



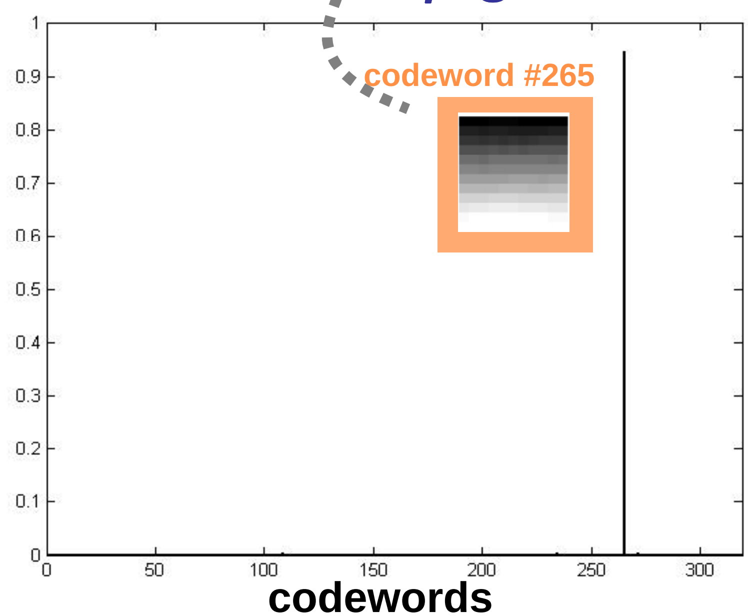
patch label



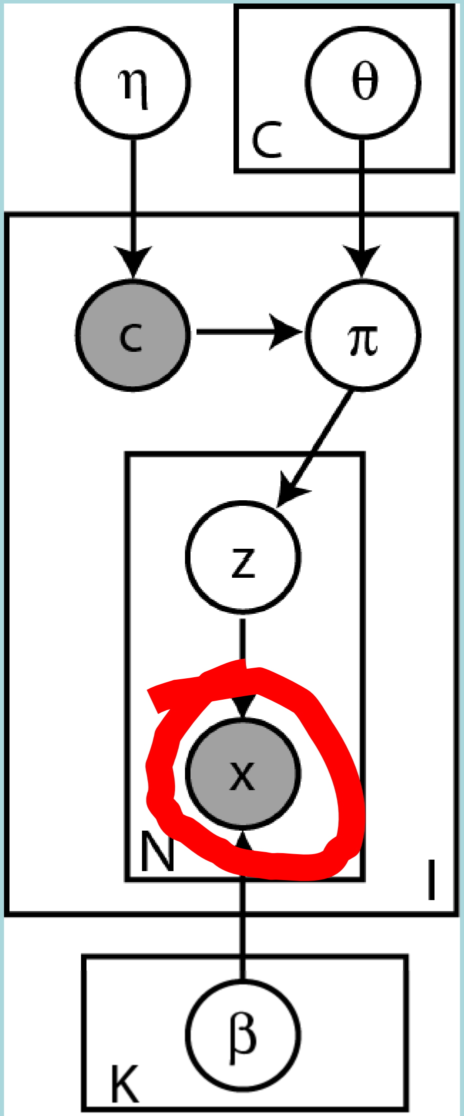
discrete variable:

$$x \sim p(x \mid z, \beta)$$
$$\sim Mult \; (x \mid z, \beta)$$

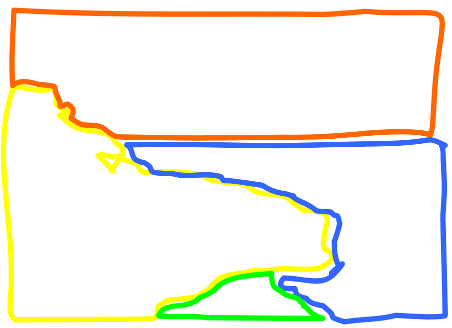
expected value of β given 'z=13'



A Generative Model



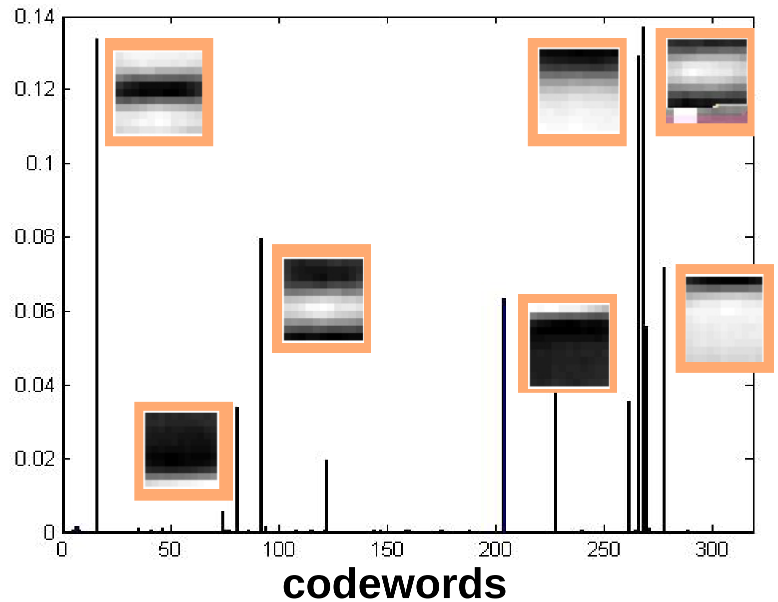
patch label

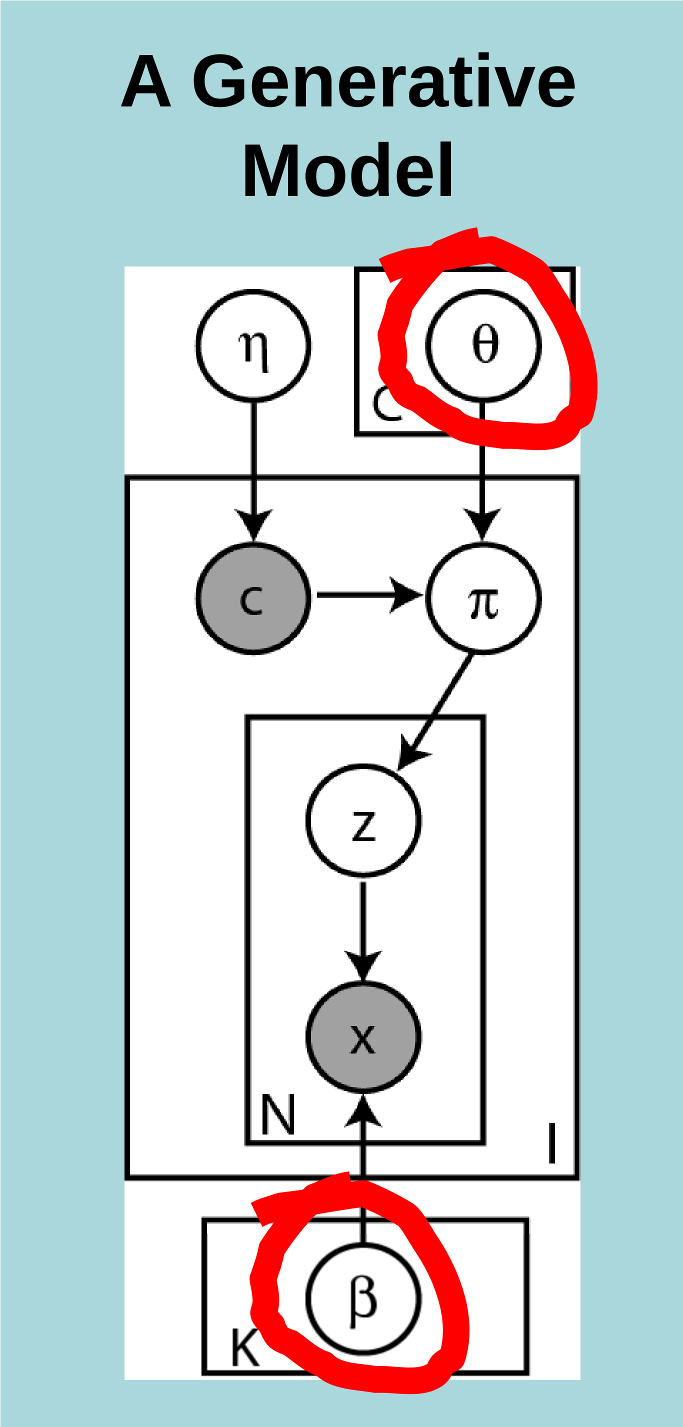


discrete variable:

$$x \sim p(x|z,\beta)$$
$$\sim Mult(x|z,\beta)$$

expected value of β given 'z=15'





learning

Find the 'best' θ and β

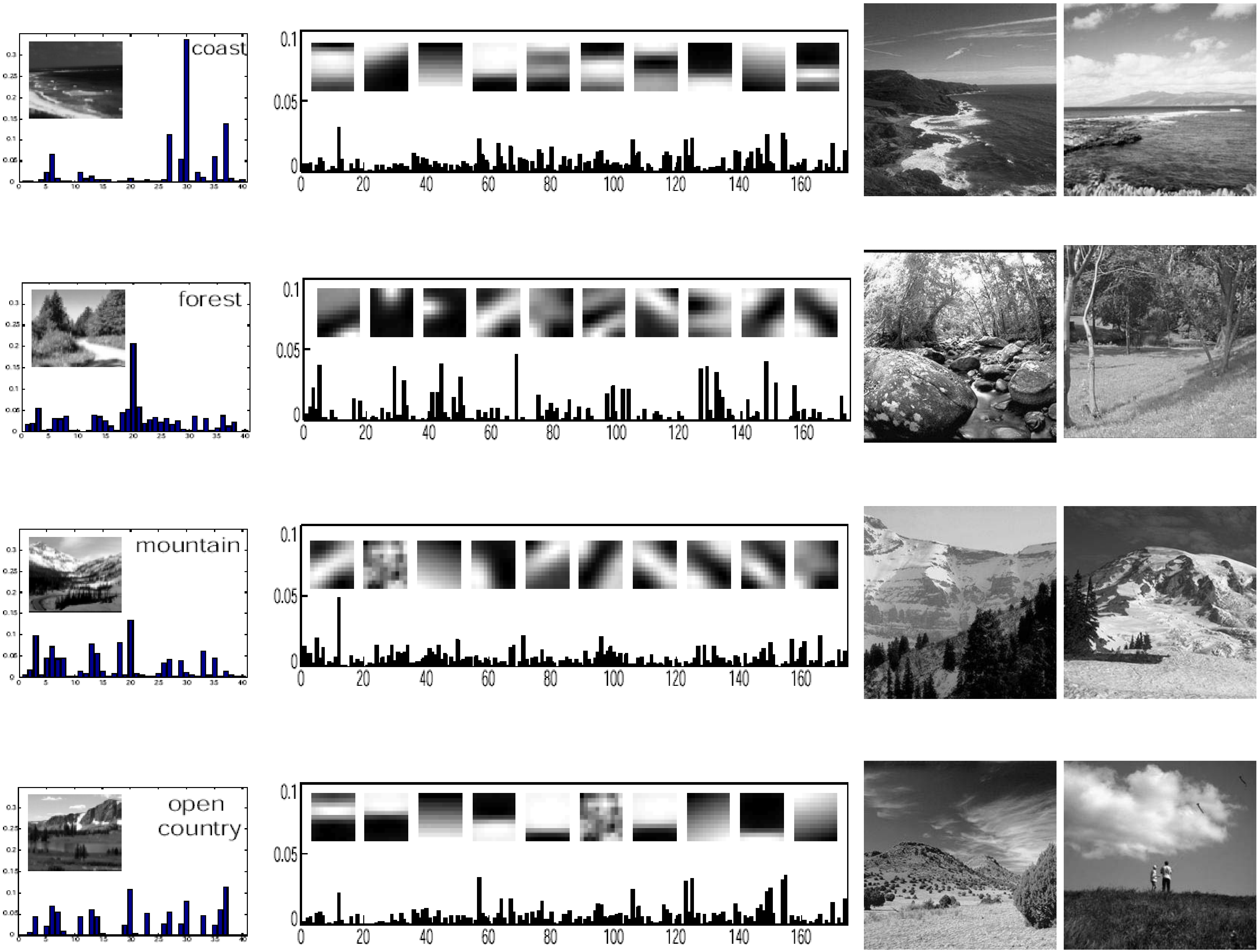
joint probability

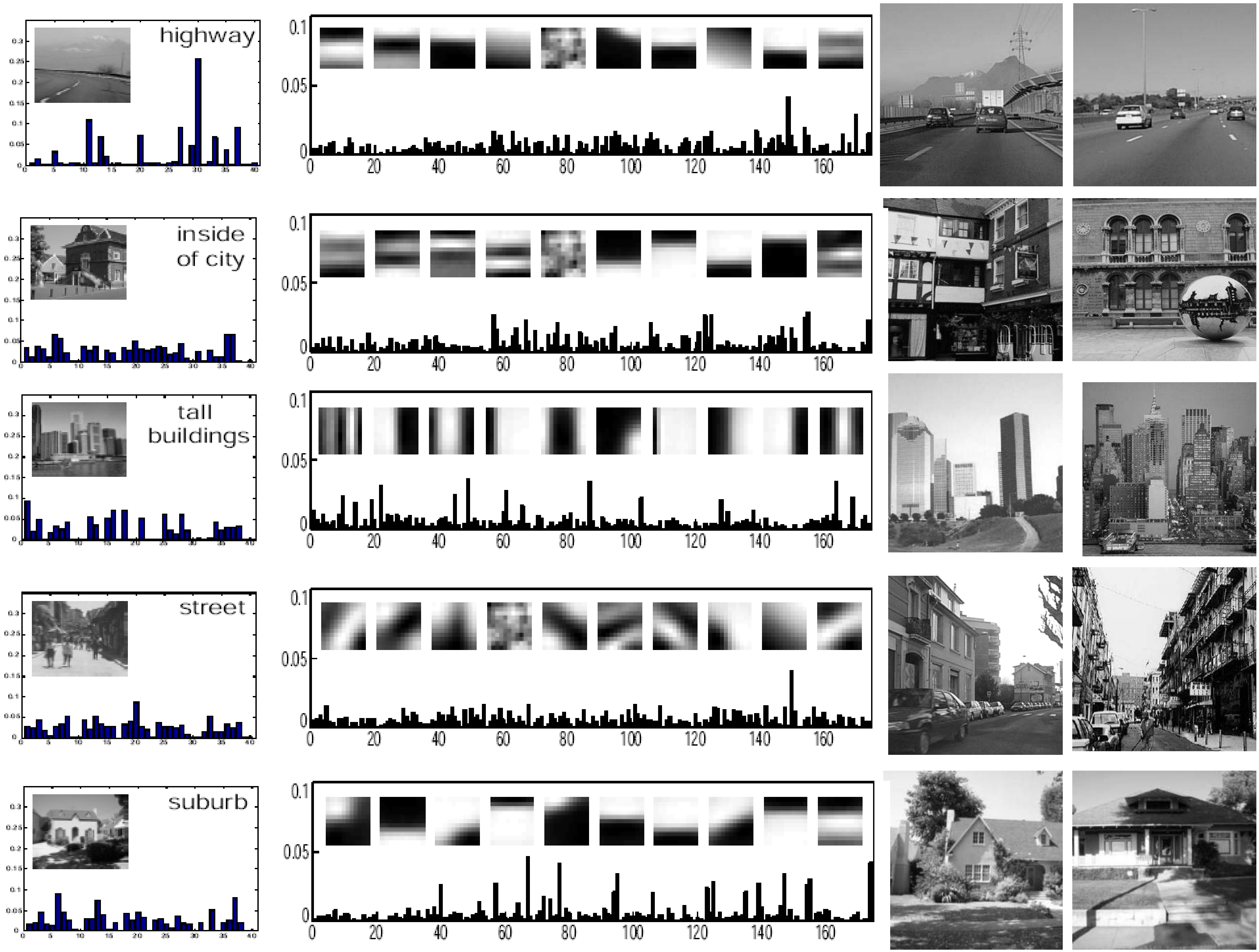
$$p(x, z, \pi | \theta, \beta, c) = p(\pi | c, \theta) \prod_{n}^{N} p(z_n | \pi) p(x_n | z_n, \beta)$$

$$p(x | \theta, \beta, c) = \int p(\pi | c, \theta) \left(\prod_{n}^{N} \sum_{z_n} p(z_n | \pi) p(x_n | z_n, \beta) \right) d\pi$$

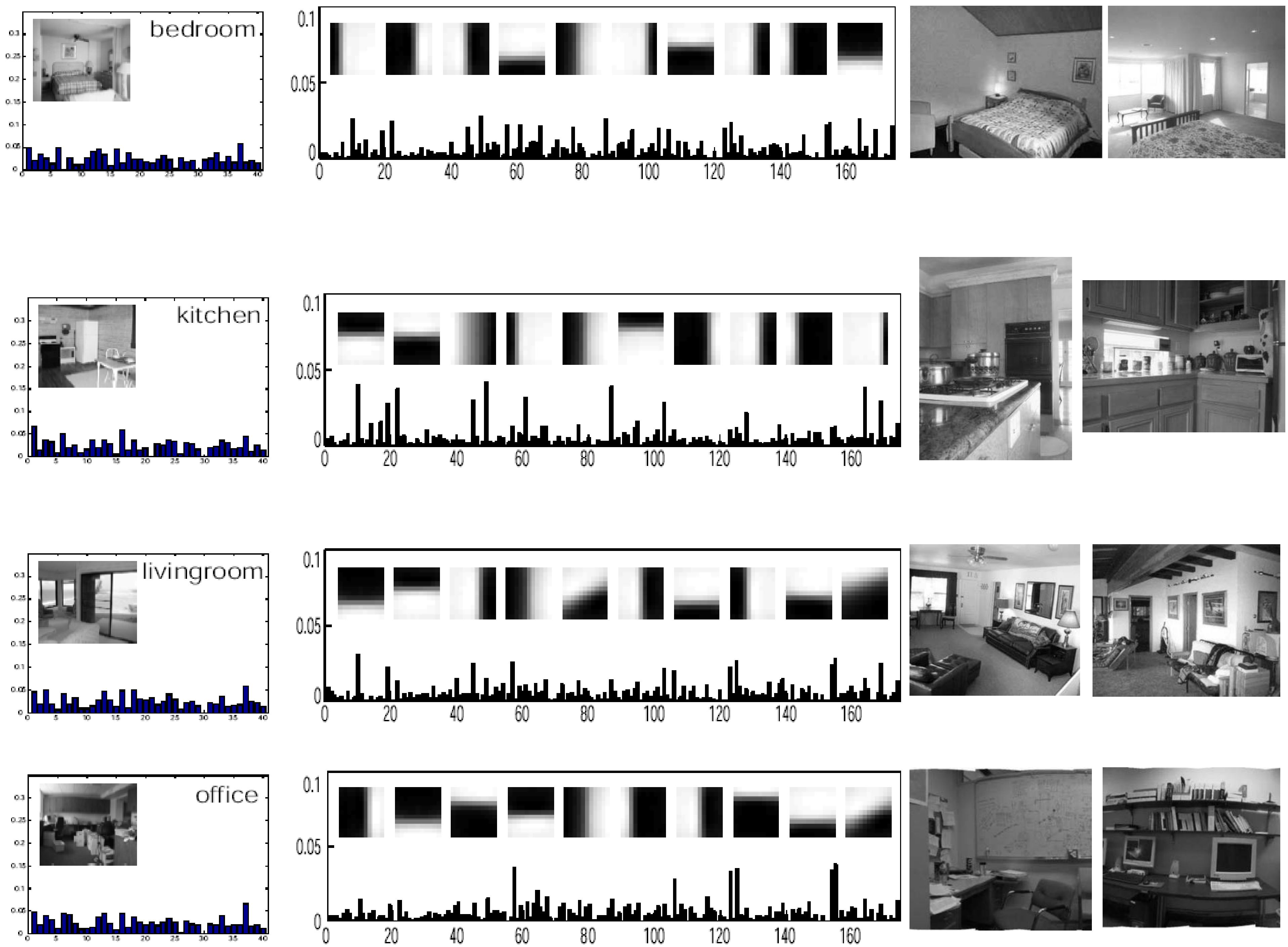
- exact inference is intractable
- use Variational Inference

Fei-Fei & Perona (CVPR 2005)

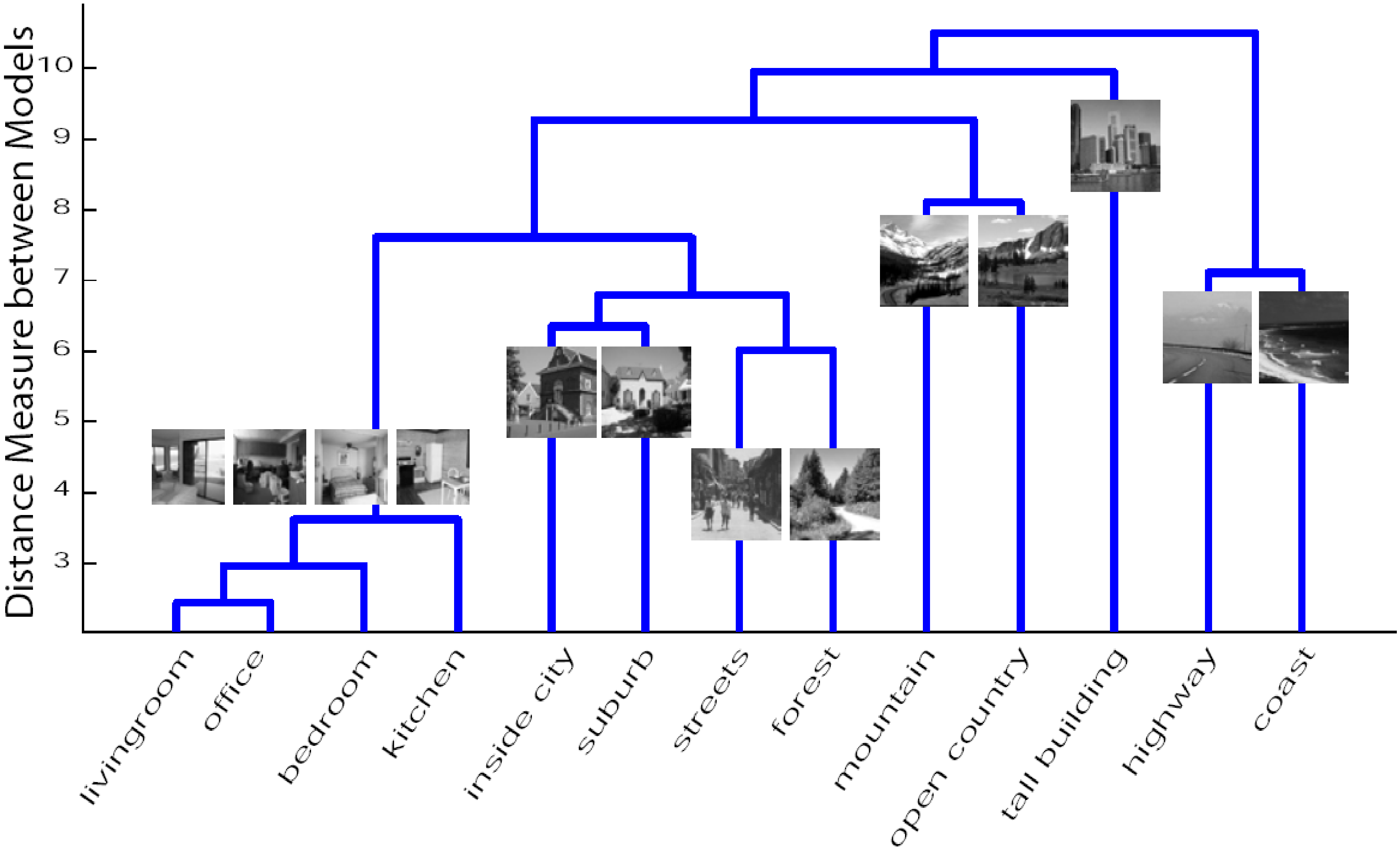

coast
forest
mountain
open
country
0.1
0.05
0
0
20
40
60
80
100
120
140
160



highway
inside of city
tall buildings
street
suburb
0.1
0.05
0
0
20
40
60
80
100
120
140
160

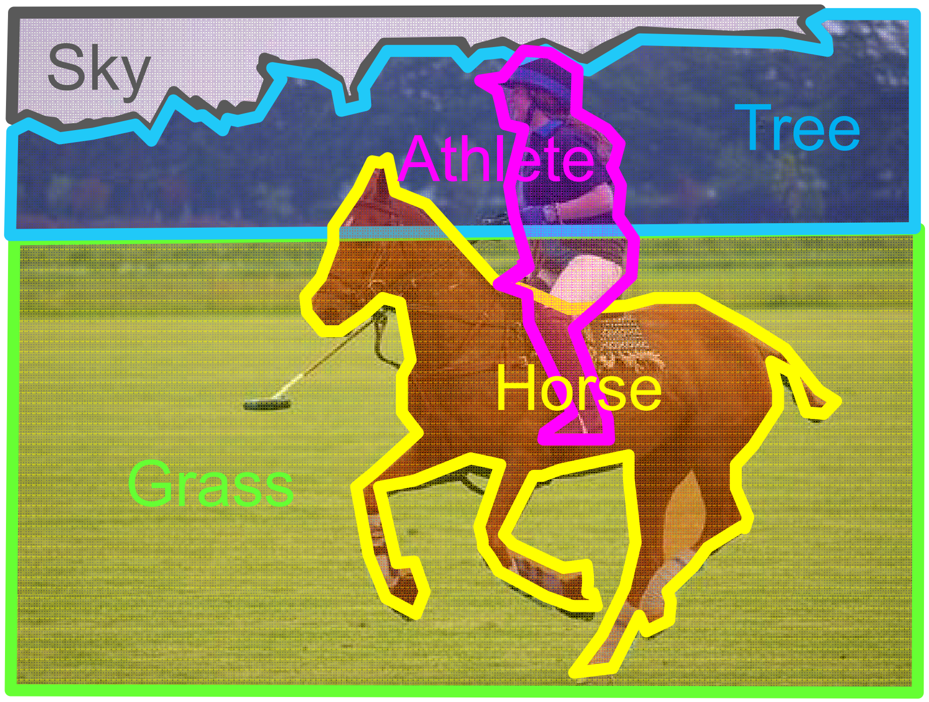

bedroom
kitchen
livingroom
office
0.1
0.05
0
0
20
40
60
80
100
120
140
160

model distance based on topic distribution



Fei-Fei & Perona (CVPR 2005)

Classification | Annotation | Segmentation

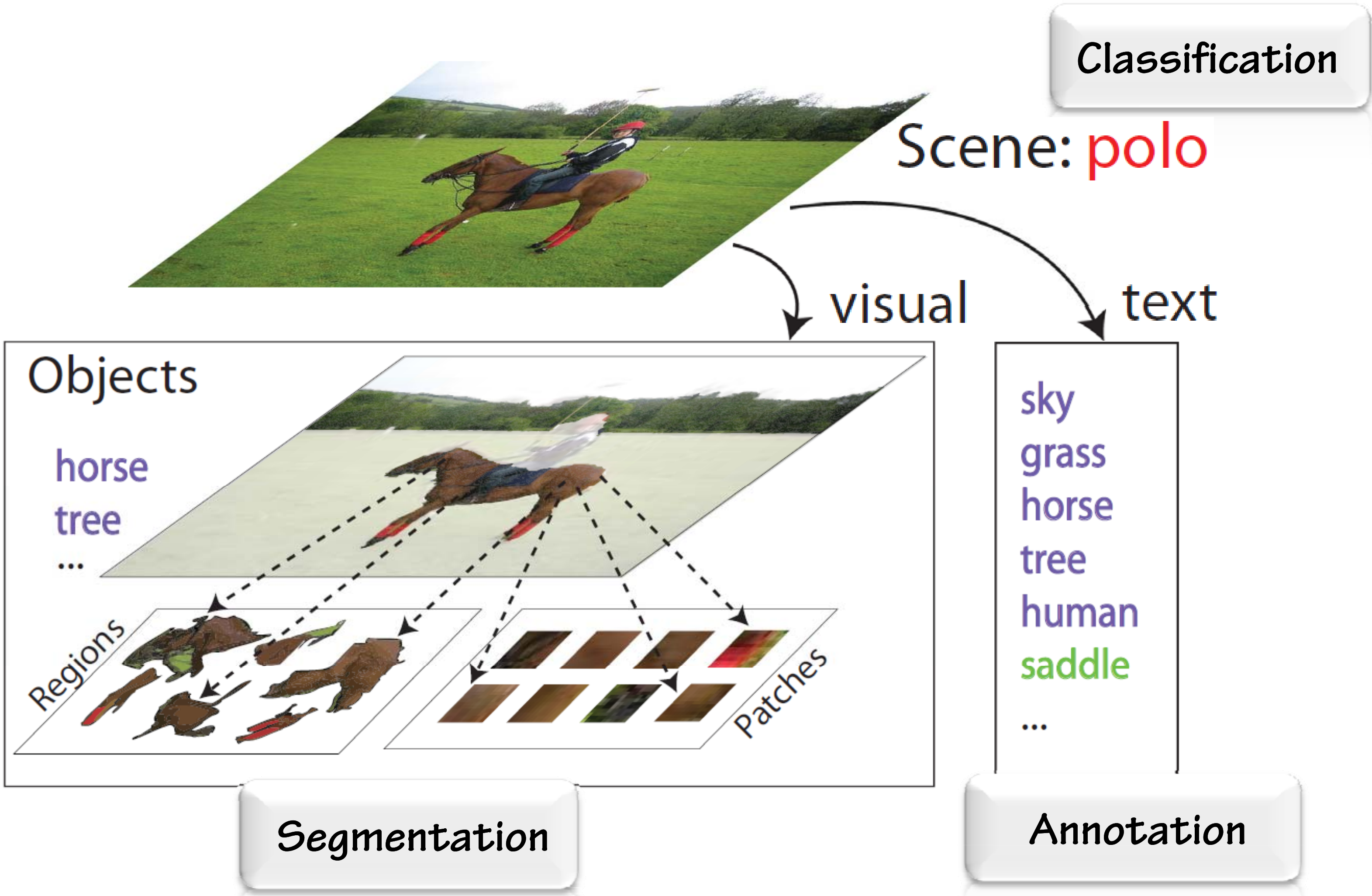


class: Polo

Athlete
Horse
Grass
Trees
Sky
Saddle

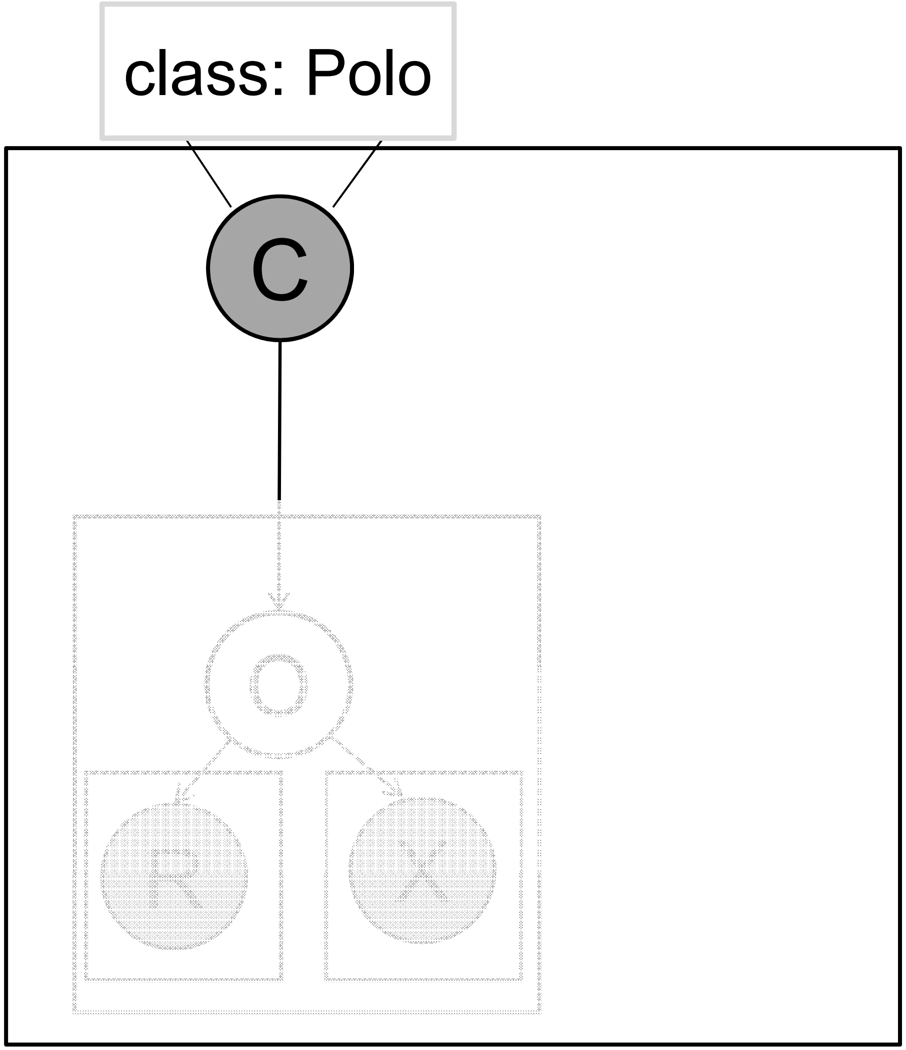
Li, Socher, & Fei-Fei, *CVPR,* 2009

Towards total scene understanding



Li, Socher, & Fei-Fei, *CVPR,* 2009

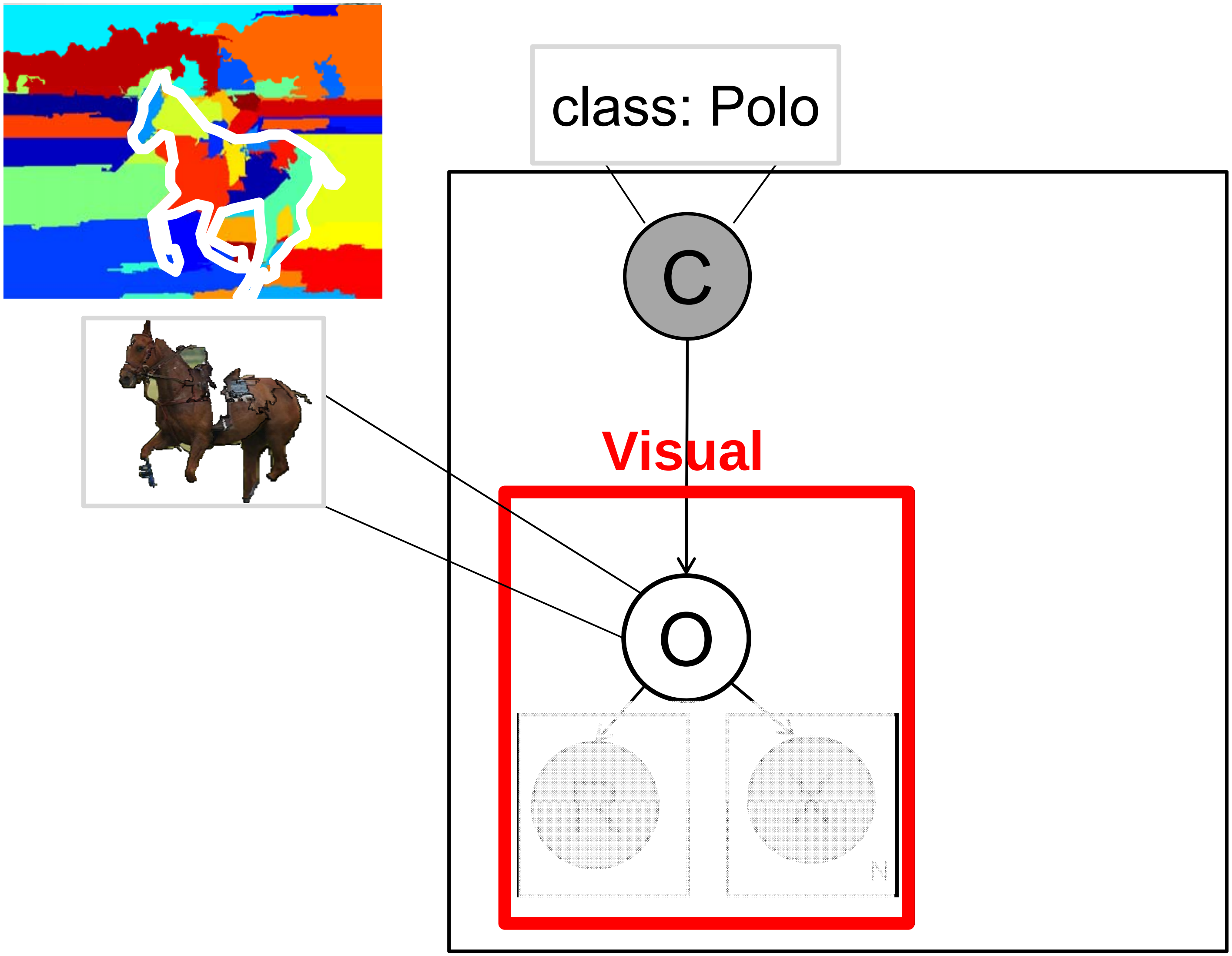
A joint model for image classification, annotation & segmentation



$$p(C, \mathbf{O}, \mathbf{R}, \mathbf{X}, \mathbf{S}, \mathbf{T}, \mathbf{Z} | \eta, \alpha, \beta, \gamma, \theta, \varphi) = p(C) \cdot$$ **Visual Component**

$\cdot$ **Text Component**

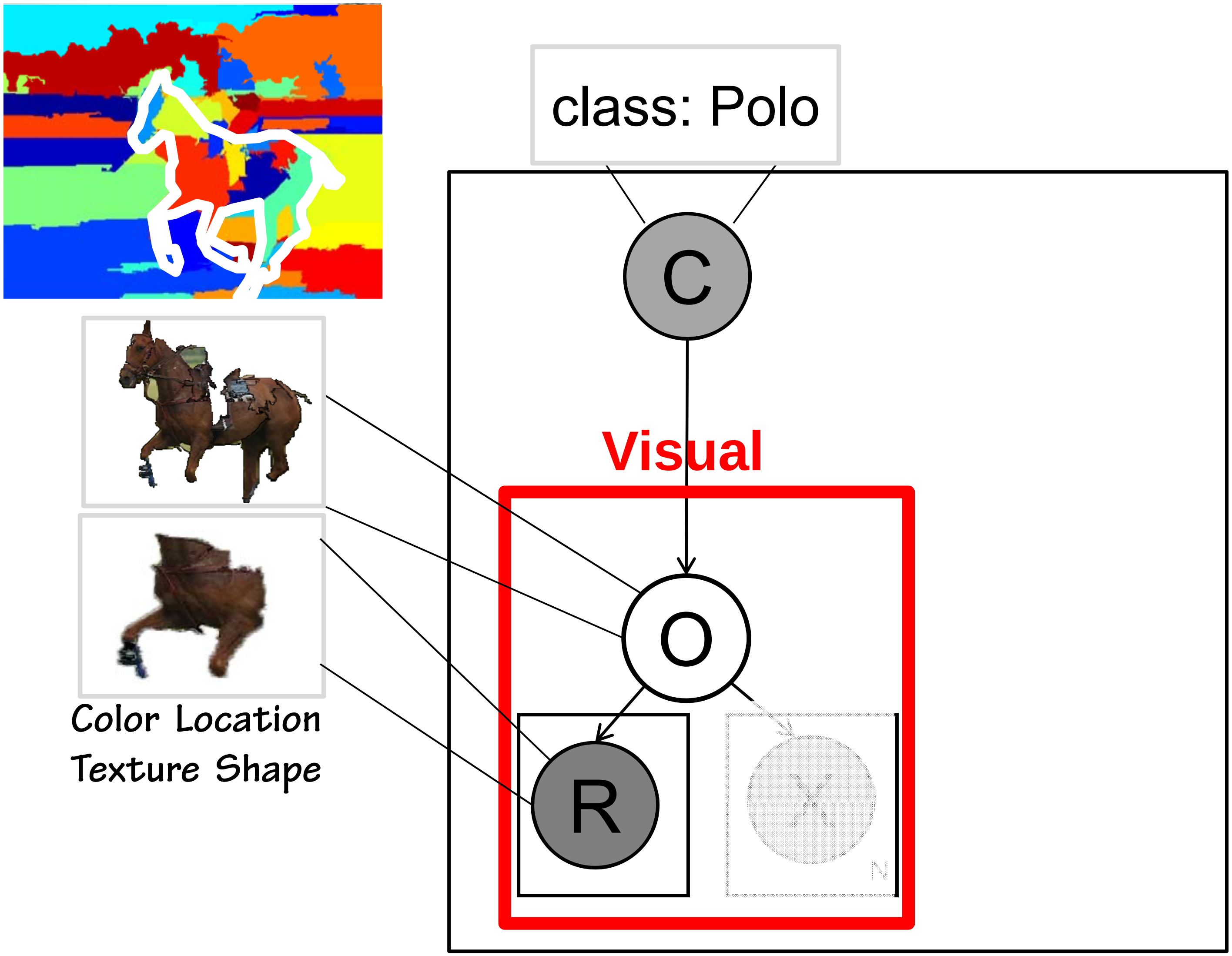
A joint model for image classification, annotation & segmentation



$$p(C, \mathbf{O}, \mathbf{R}, \mathbf{X}, \mathbf{S}, \mathbf{T}, \mathbf{Z} | \eta, \alpha, \beta, \gamma, \theta, \varphi) = p(C) \cdot \left(\prod_{n=1}^{N_r} p(O_n | \eta, C) \right)$$

$\cdot$ **Text Component**

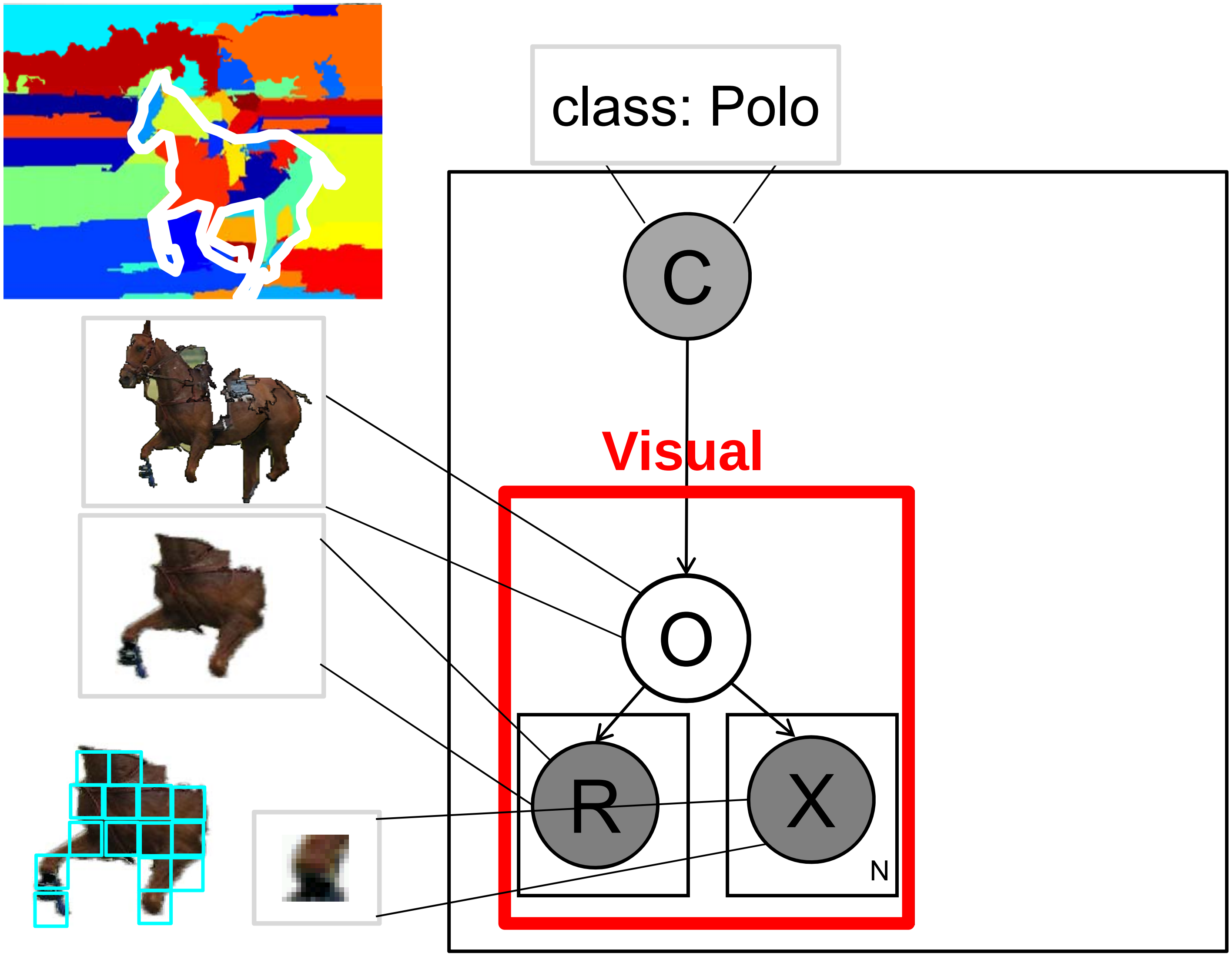
A joint model for image classification, annotation & segmentation



$$p(C, \mathbf{O}, \mathbf{R}, \mathbf{X}, \mathbf{S}, \mathbf{T}, \mathbf{Z} | \eta, \alpha, \beta, \gamma, \theta, \varphi) = p(C) \cdot \boxed{\left(\prod_{n=1}^{N_r} p(O_n | \eta, C)\right) \prod_{n=1}^{N_r} \left(\left(\prod_{i=1}^{N_F} p(R_{ni} | O_n, \alpha_i)\right)\right.}$$

$\cdot$ **Text Component**

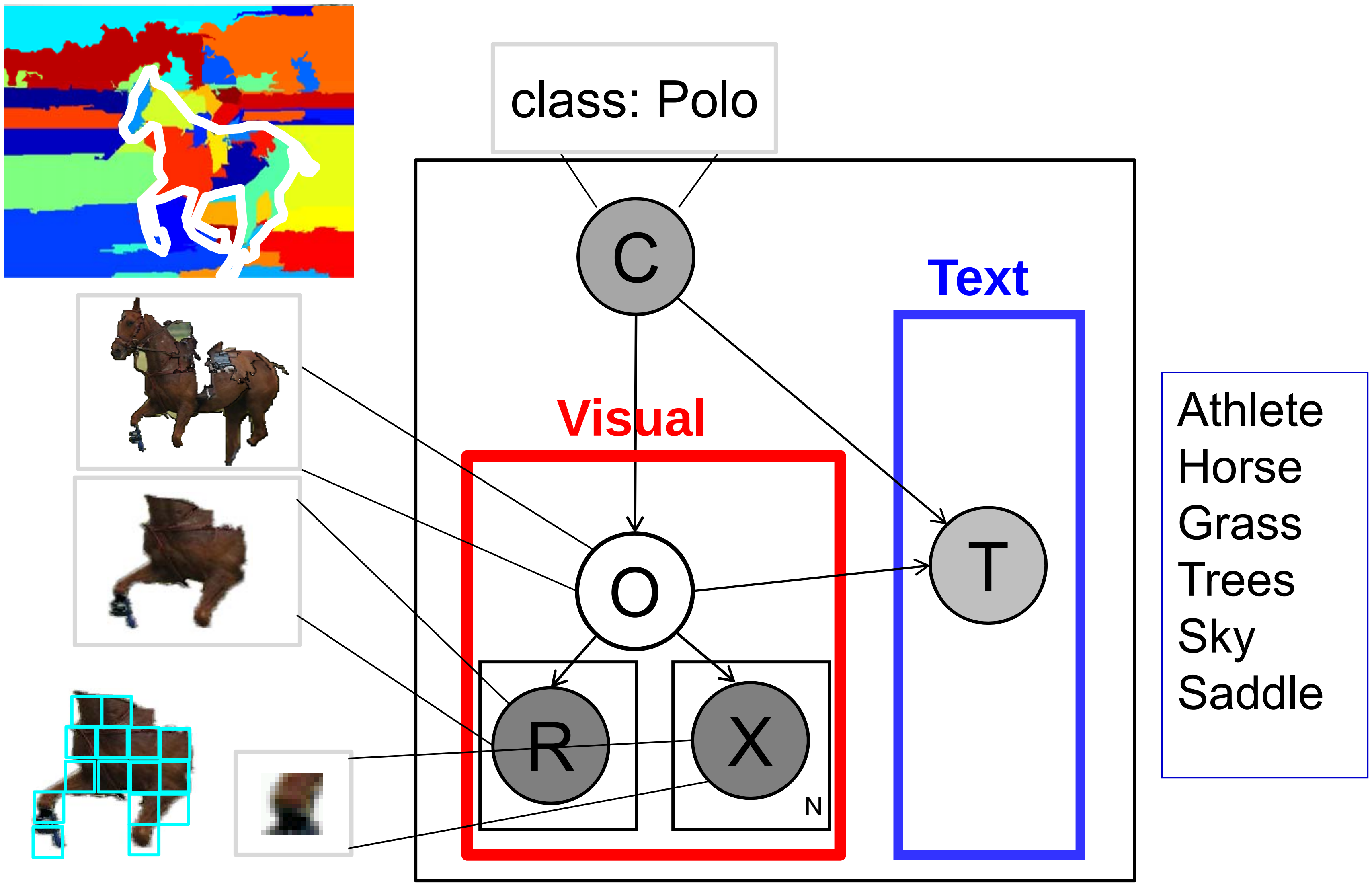
A joint model for image classification, annotation & segmentation



$$p(C,\mathbf{O},\mathbf{R},\mathbf{X},\mathbf{S},\mathbf{T},\mathbf{Z}|\eta,\alpha,\beta,\gamma,\theta,\varphi)= p(C)\cdot\left(\prod_{n=1}^{N_r} p(O_n|\eta,C)\right)\prod_{n=1}^{N_r}\left(\left(\prod_{i=1}^{N_F} p(R_{ni}|O_n,\alpha_i)\right)\cdot\prod_{r=1}^{A_r} p(X_{nr}|O_n,\beta)\right.$$

$\cdot$ **Text Component**

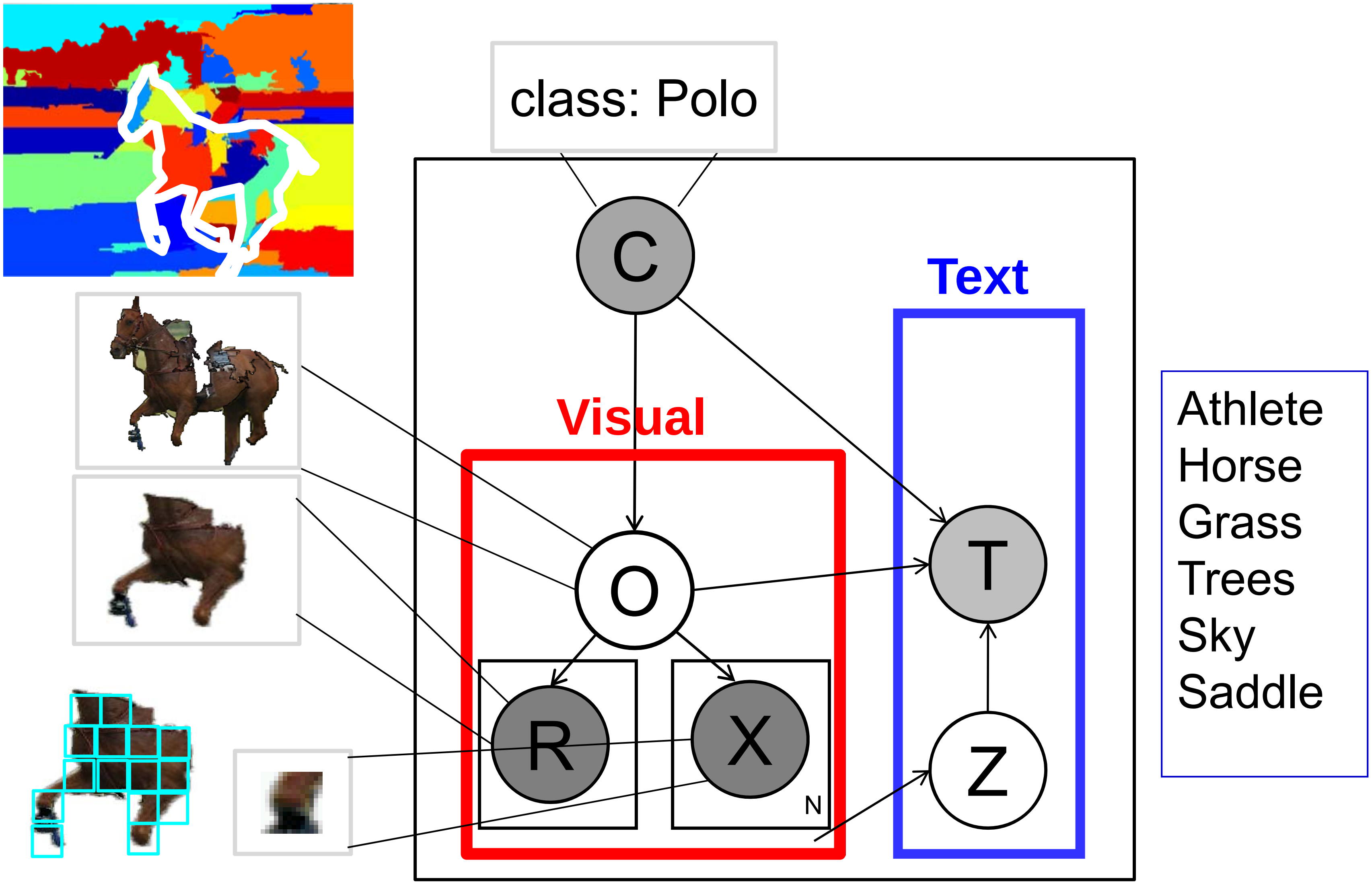
A joint model for image classification, annotation & segmentation



$$p(C, \mathbf{O}, \mathbf{R}, \mathbf{X}, \mathbf{S}, \mathbf{T}, \mathbf{Z} | \eta, \alpha, \beta, \gamma, \theta, \varphi) = p(C) \cdot \boxed{\left(\prod_{n=1}^{N_r} p(O_n|\eta, C)\right) \prod_{n=1}^{N_r} \left(\left(\prod_{i=1}^{N_F} p(R_{ni}|O_n, \alpha_i)\right) \cdot \prod_{r=1}^{A_r} p(X_{nr}|O_n, \beta)\right.}$$

$$\cdot \boxed{p(T_m|O_{Z_m}, S_m, \theta, C, \varphi)}$$

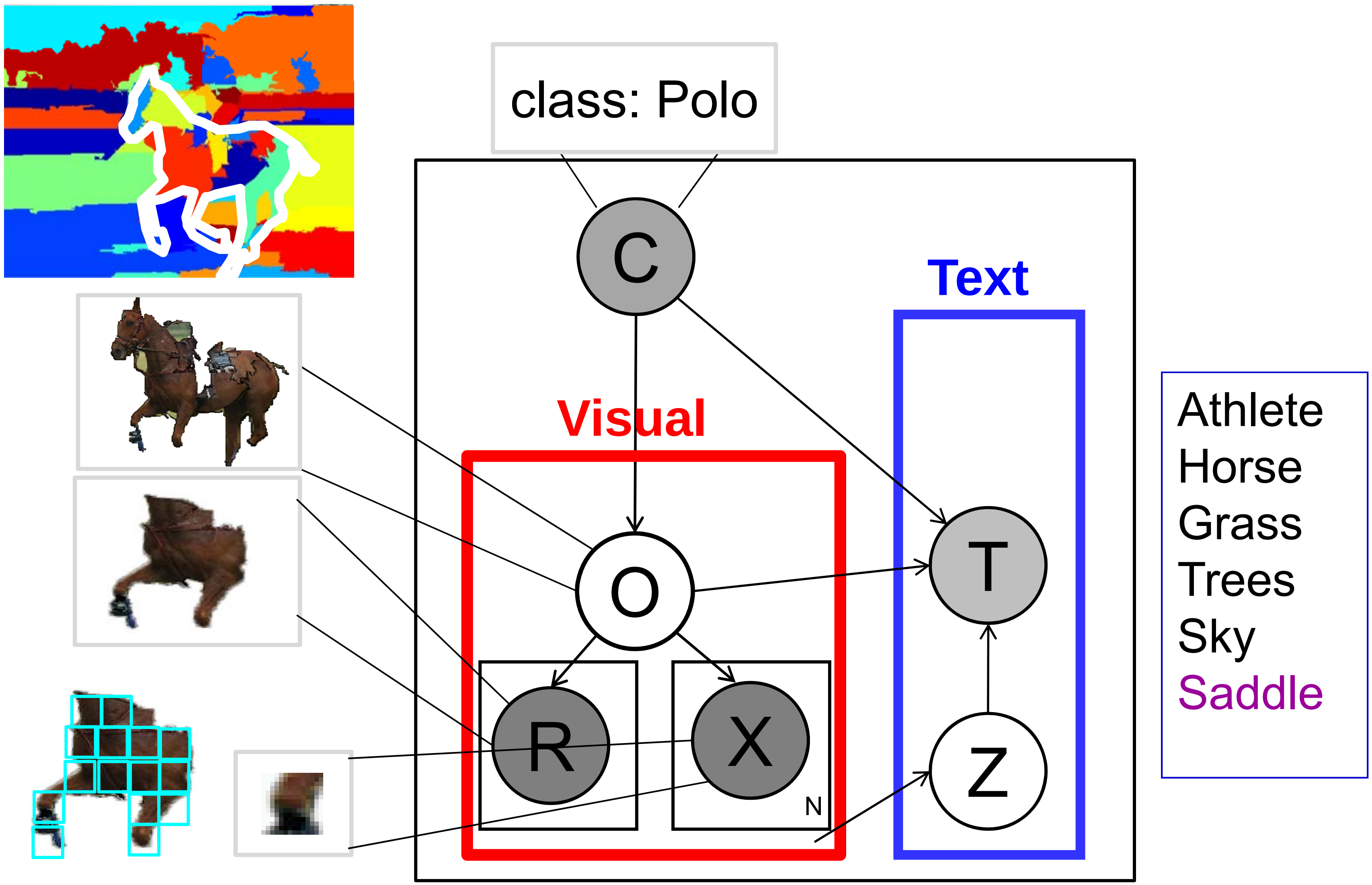
A joint model for image classification, annotation & segmentation



$$p(C, \mathbf{O}, \mathbf{R}, \mathbf{X}, \mathbf{S}, \mathbf{T}, \mathbf{Z} | \eta, \alpha, \beta, \gamma, \theta, \varphi) = p(C) \cdot \left(\prod_{n=1}^{N_r} p(O_n|\eta, C)\right) \prod_{n=1}^{N_r} \left(\left(\prod_{i=1}^{N_F} p(R_{ni}|O_n, \alpha_i)\right) \cdot \prod_{r=1}^{A_r} p(X_{nr}|O_n, \beta)\right.$$

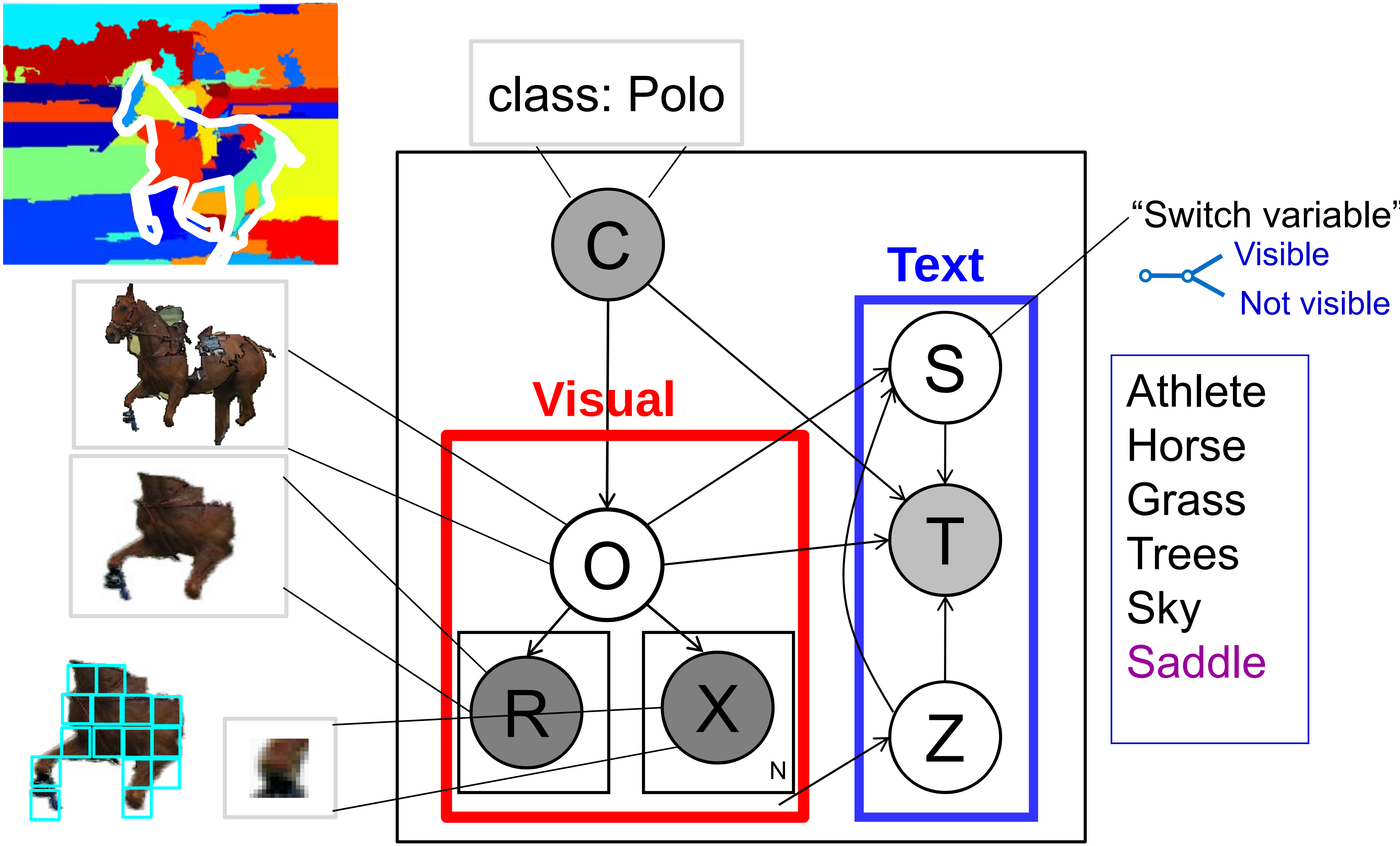
$$\cdot \prod_{m=1}^{N_t} p(Z_m|N_r) \quad p(T_m|O_{Z_m}, S_m, \theta, C, \varphi)$$

A joint model for image classification, annotation & segmentation



$$p(C, \mathbf{O}, \mathbf{R}, \mathbf{X}, \mathbf{S}, \mathbf{T}, \mathbf{Z} | \eta, \alpha, \beta, \gamma, \theta, \varphi) = p(C) \cdot \left(\prod_{n=1}^{N_r} p(O_n|\eta, C)\right) \prod_{n=1}^{N_r} \left(\left(\prod_{i=1}^{N_F} p(R_{ni}|O_n, \alpha_i)\right) \cdot \prod_{r=1}^{A_r} p(X_{nr}|O_n, \beta)\right) \cdot \prod_{m=1}^{N_t} p(Z_m|N_r) \; p(T_m|O_{Z_m}, S_m, \theta, C, \varphi)$$

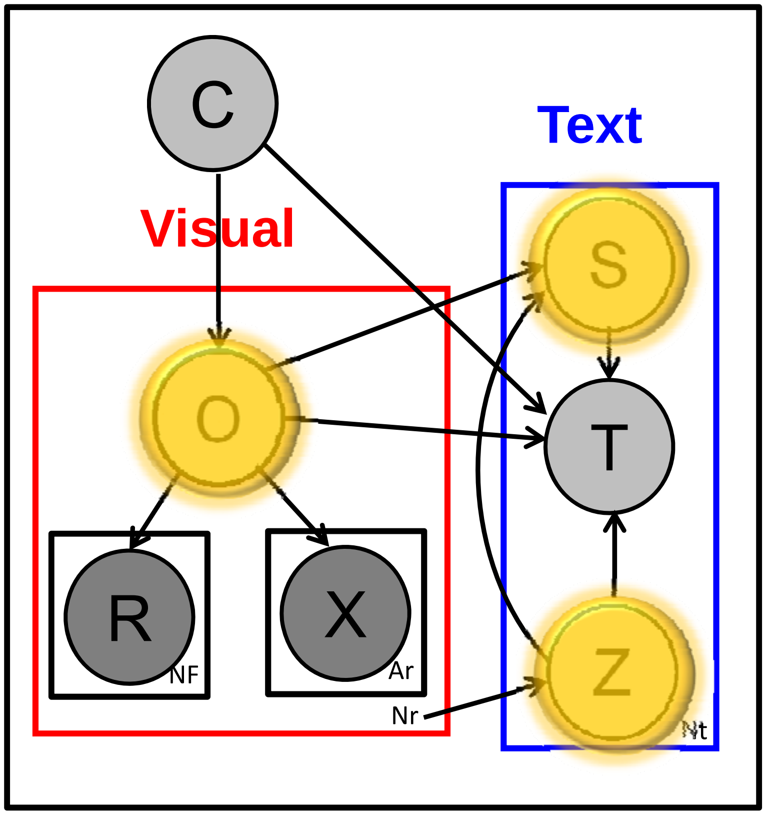
A joint model for image classification, annotation & segmentation



$$p(C, \mathbf{O}, \mathbf{R}, \mathbf{X}, \mathbf{S}, \mathbf{T}, \mathbf{Z} | \eta, \alpha, \beta, \gamma, \theta, \varphi) = p(C) \cdot \left(\prod_{n=1}^{N_r} p(O_n|\eta, C)\right) \prod_{n=1}^{N_r} \left(\left(\prod_{i=1}^{N_F} p(R_{ni}|O_n, \alpha_i)\right) \cdot \prod_{r=1}^{A_r} p(X_{nr}|O_n, \beta)\right.$$

$$\cdot \prod_{m=1}^{N_t} p(Z_m|N_r) \quad p(S_m|O_{Z_m}, \gamma) \quad p(T_m|O_{Z_m}, S_m, \theta, C, \varphi)$$

Learning



Exact Inference is Intractable!

Relationship of the random variables

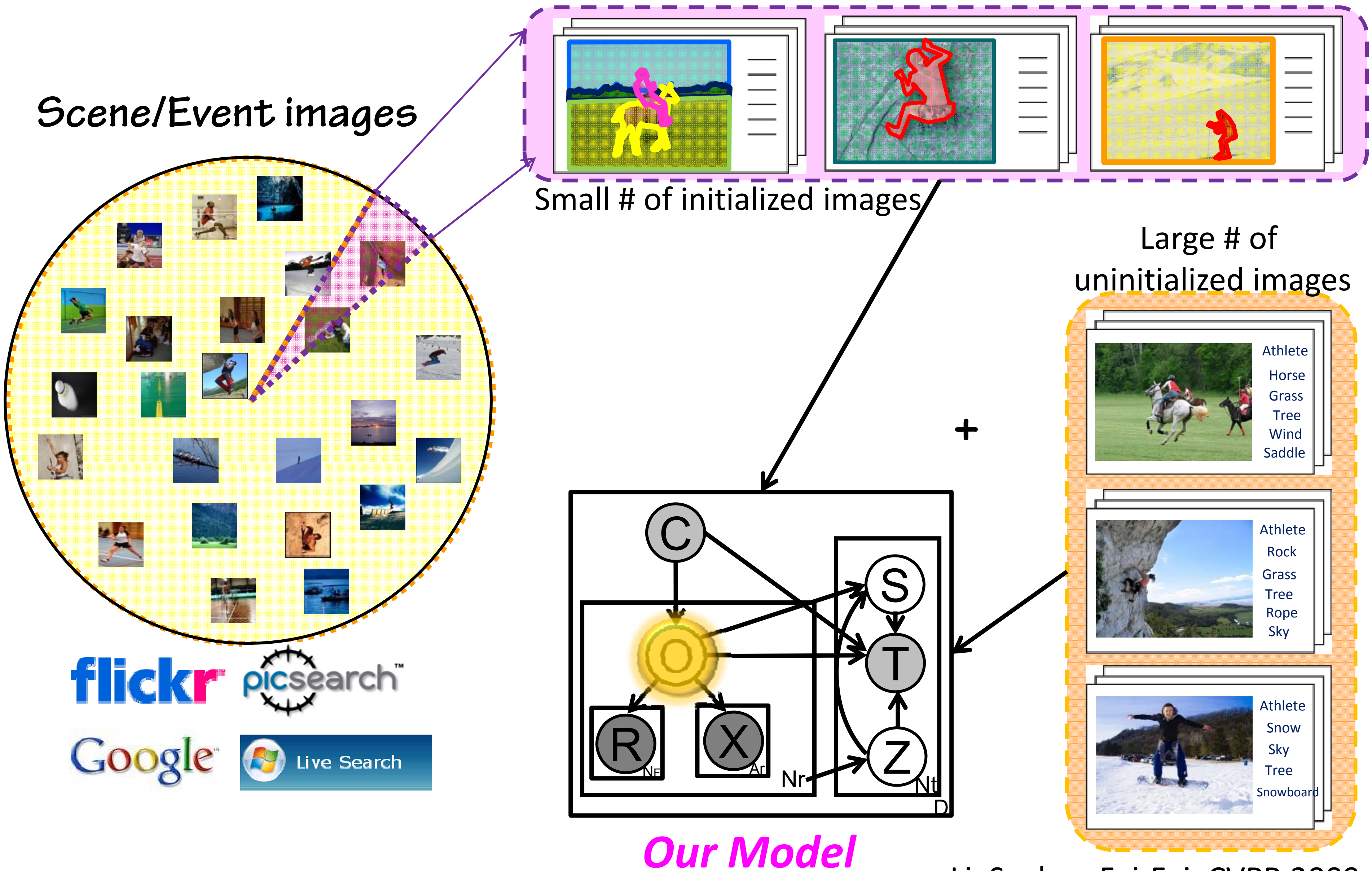
$$p(C,\mathbf{O},\mathbf{R},\mathbf{X},\mathbf{S},\mathbf{T},\mathbf{Z}|\eta,\alpha,\beta,\gamma,\theta,\varphi) = p(C)\cdot\left(\prod_{n=1}^{N_r} p(O_n|\eta,C)\right)\prod_{n=1}^{N_r}\left(\left(\prod_{i=1}^{N_F} p(R_{ni}|O_n,\alpha_i)\right)\cdot\prod_{r=1}^{A_r} p(X_{nr}|O_n,\beta)\right.$$

$$\prod_{m=1}^{N_t} p(Z_m|N_r)\,p(S_m|O_{Z_m},\gamma)\,p(T_m|O_{Z_m},S_m,\theta,C,\varphi)$$

Li, Socher, Fei-Fei, CVPR 2009

Auto-semi-supervised learning:

Small # of initialized images + Large # of uninitialized images



Li, Socher, Fei-Fei, CVPR 2009

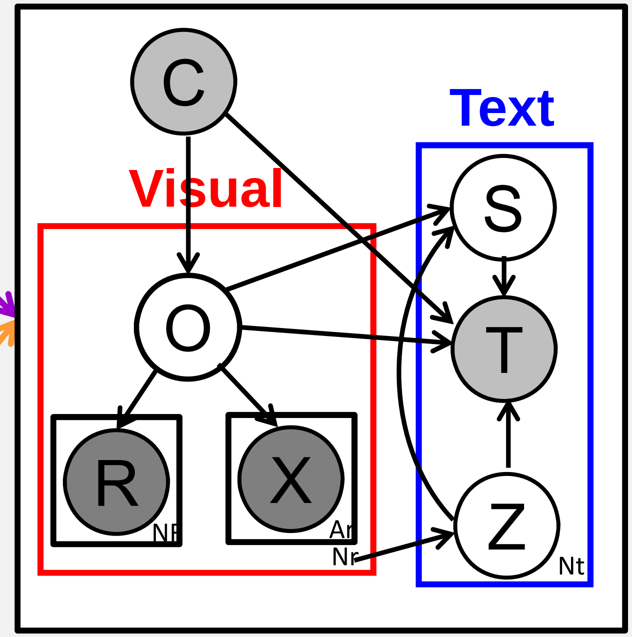
Outline

Learning

flickr Small # of automatically initialized images

Large # of uninitialized images

Model



Recognition & Experiment

- Dataset
- Learned Model
- Results

Li, Socher, Fei-Fei, CVPR 2009

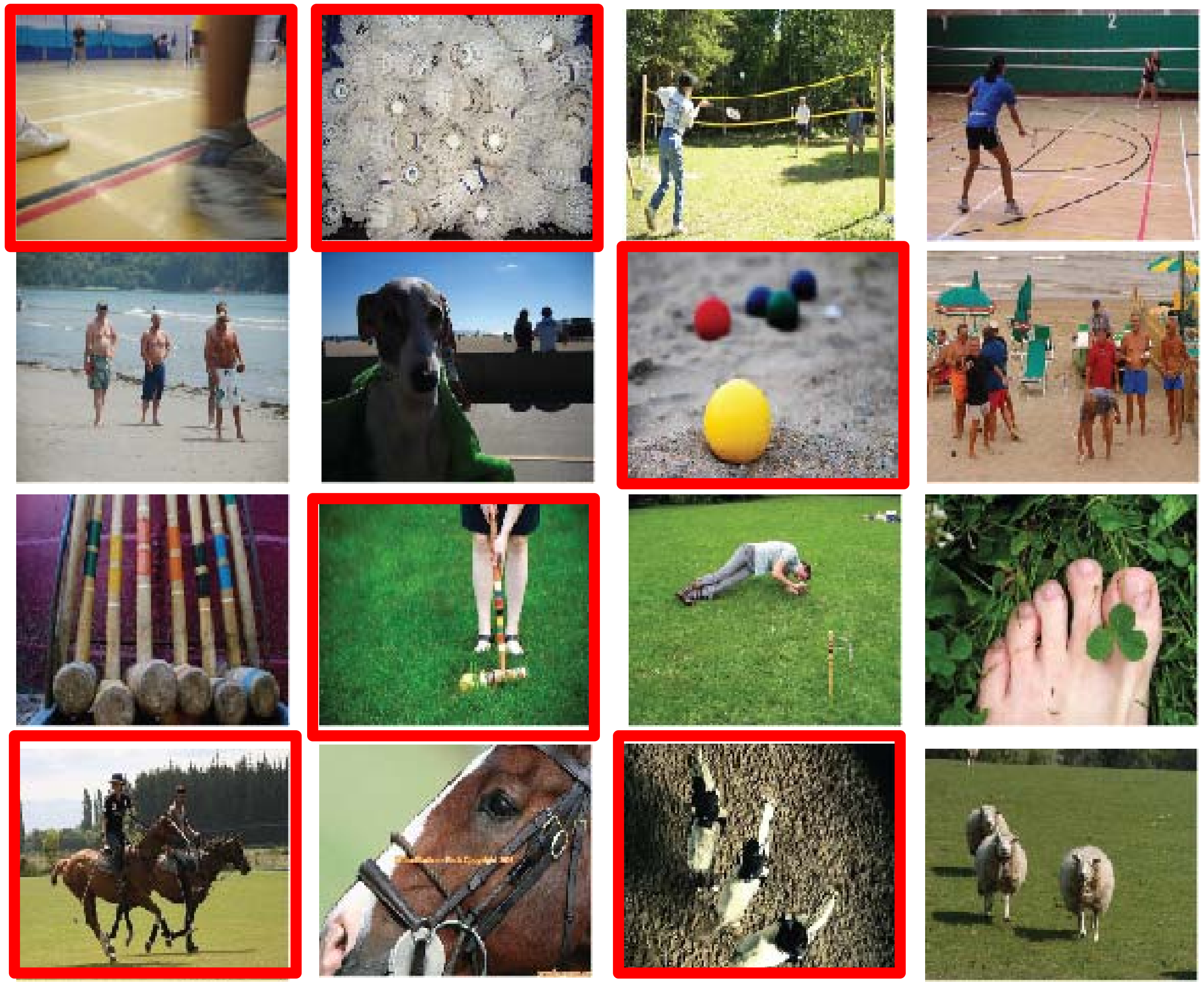
flickr 8 Event/Scene Classes

Badminton

Bocce

Croquet

Polo



Remark: Tags are not used during testing

Li, Socher, Fei-Fei, CVPR 2009

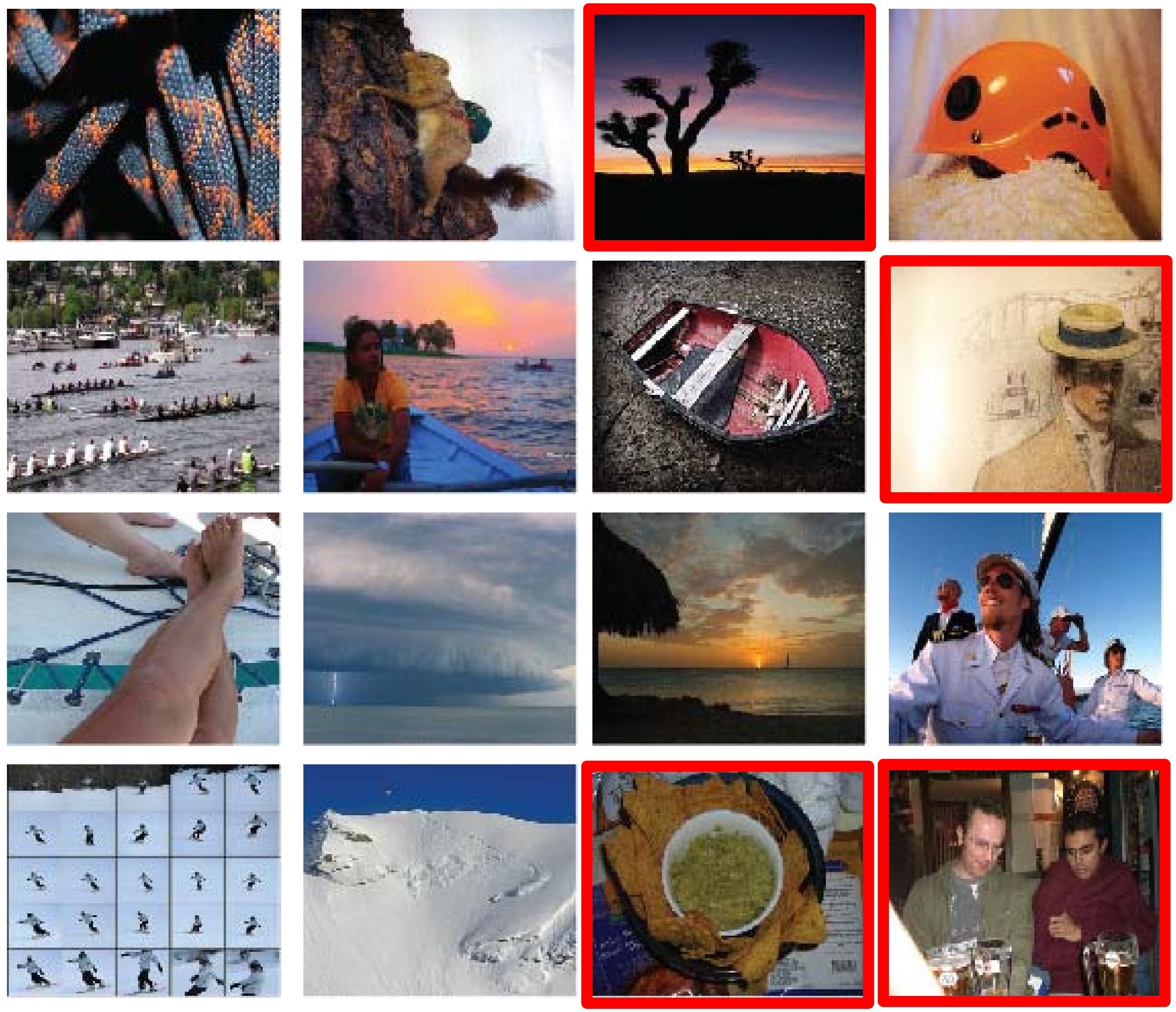
flickr 8 Event/Scene Classes

Rock climbing

Rowing

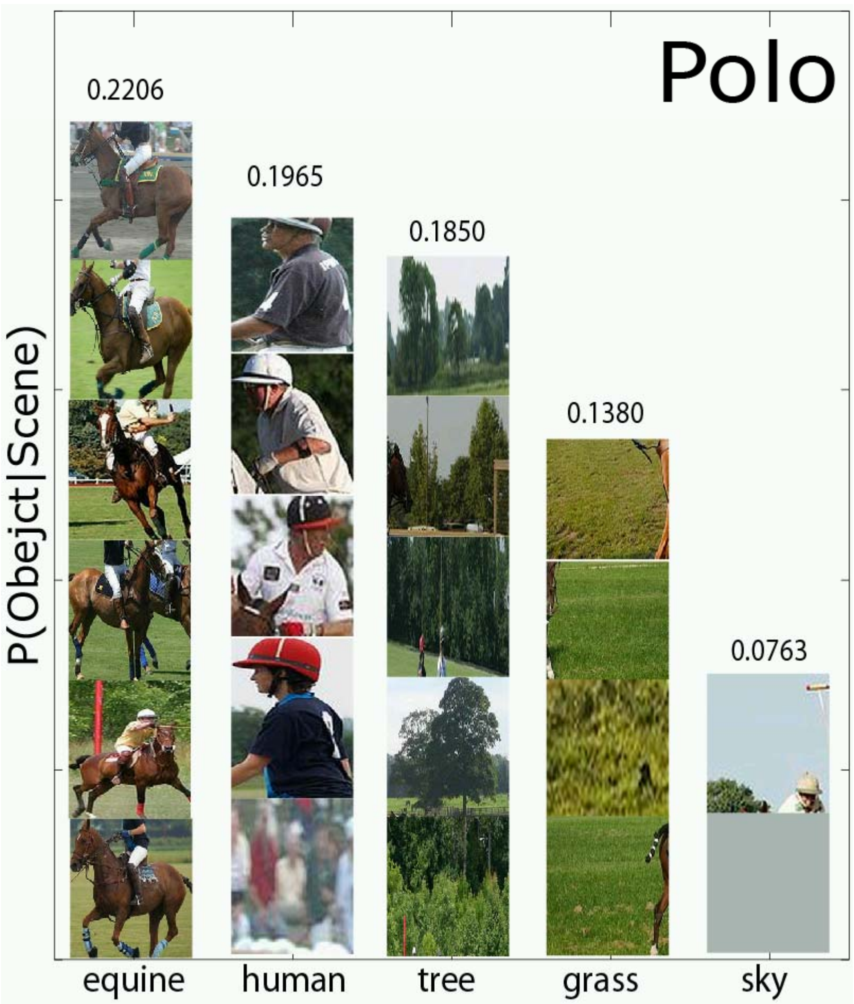
Sailing

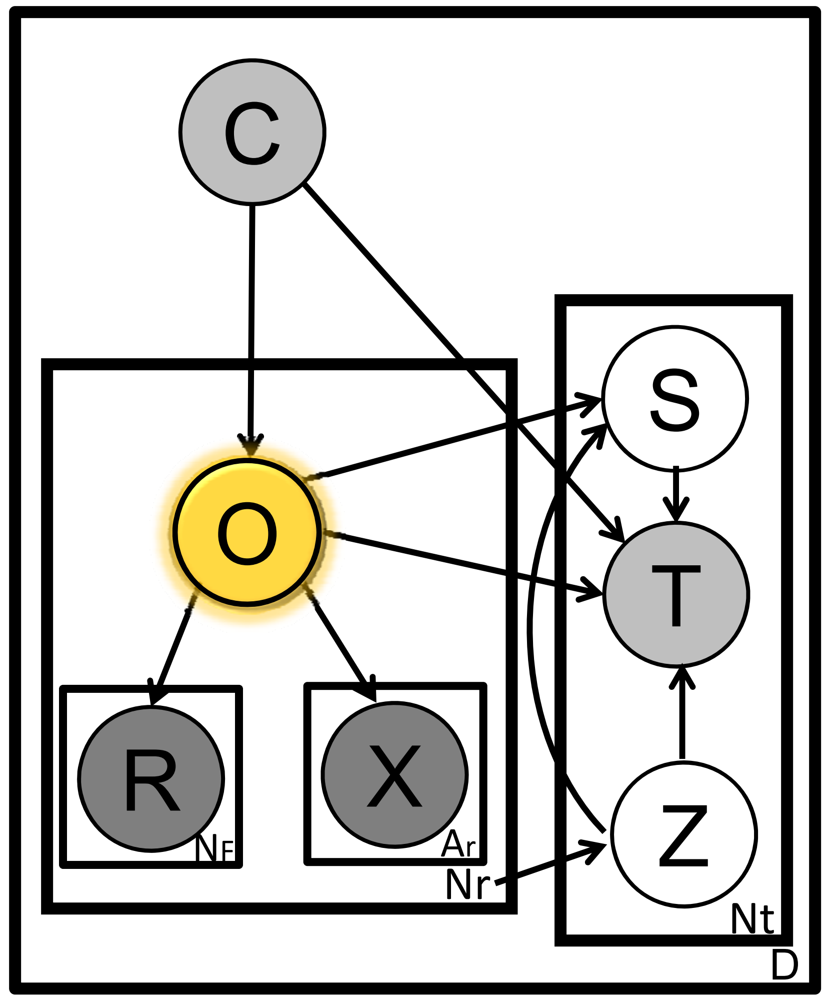
Snow boarding



Li, Socher, Fei-Fei, CVPR 2009

Learned model: O





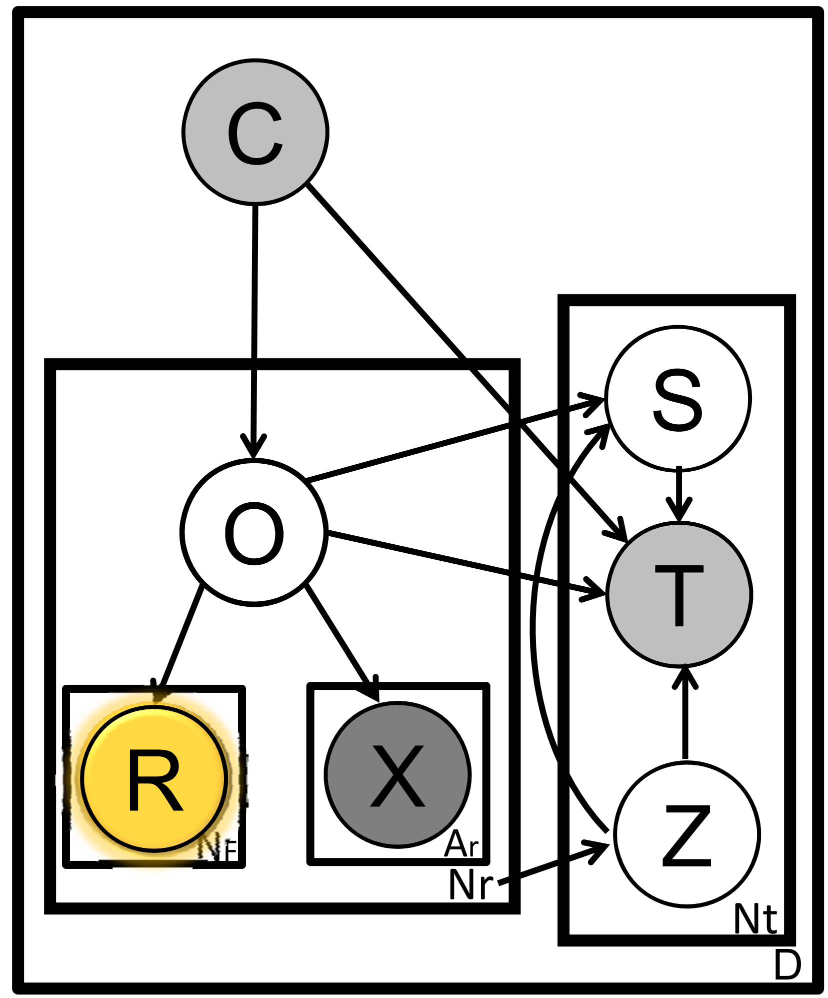
Li, Socher, Fei-Fei, CVPR 2009

Learned model: R

Athlete

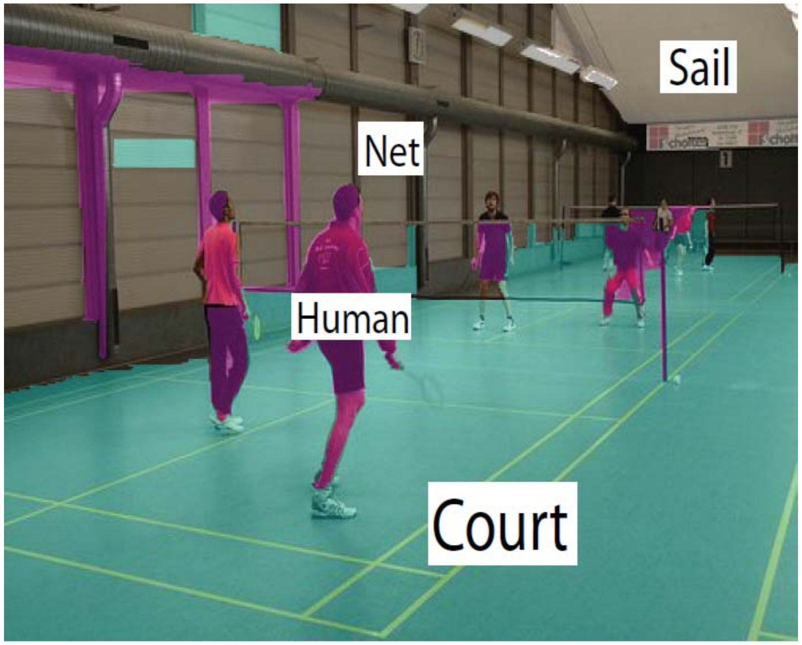
Grass

Horse

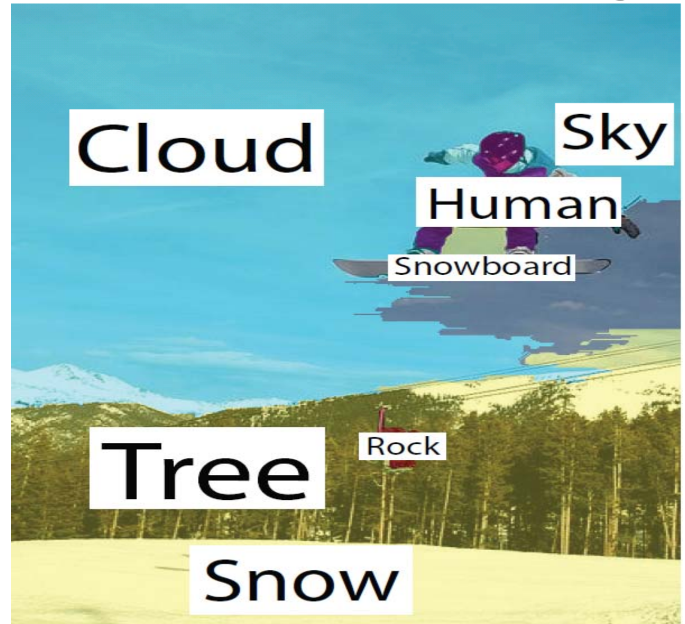


Li, Socher, Fei-Fei, CVPR 2009

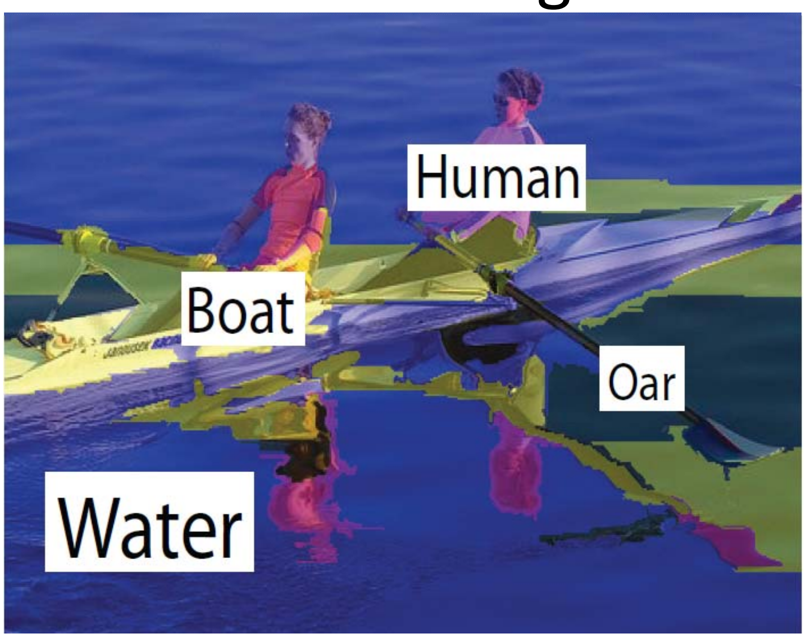
Class: Badminton



Class: Snowboarding



Class: Rowing



Class: Rock Climbing