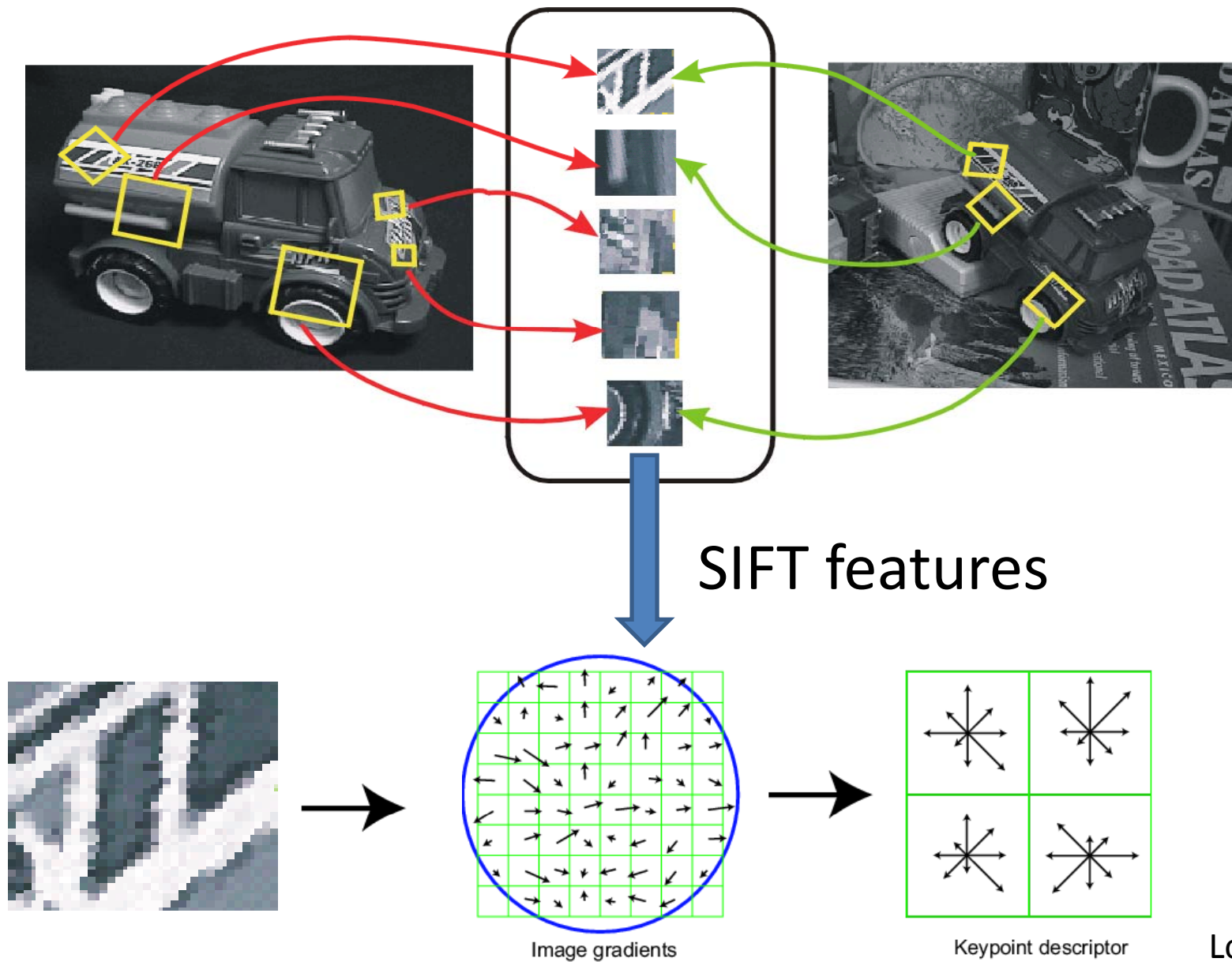
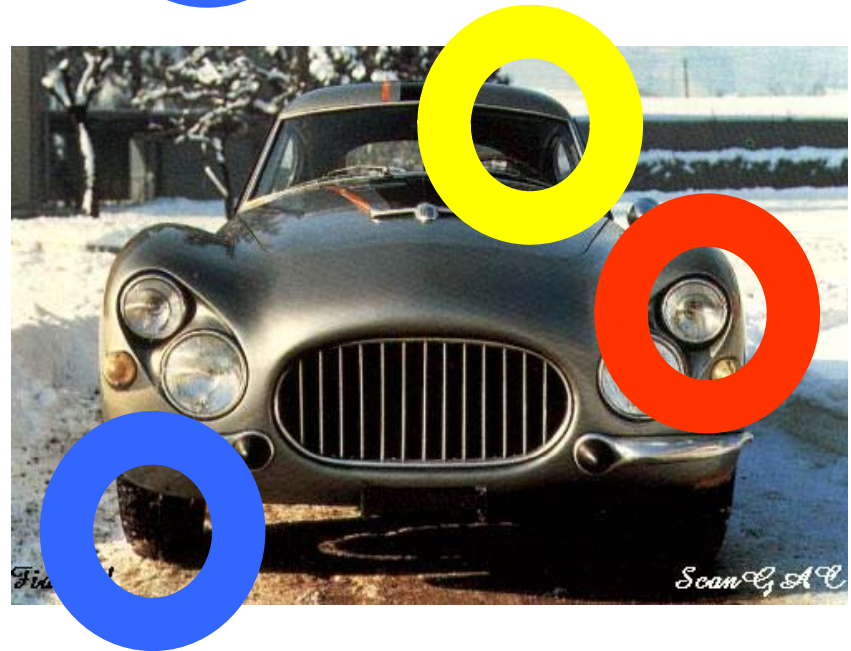


Matching single objects...



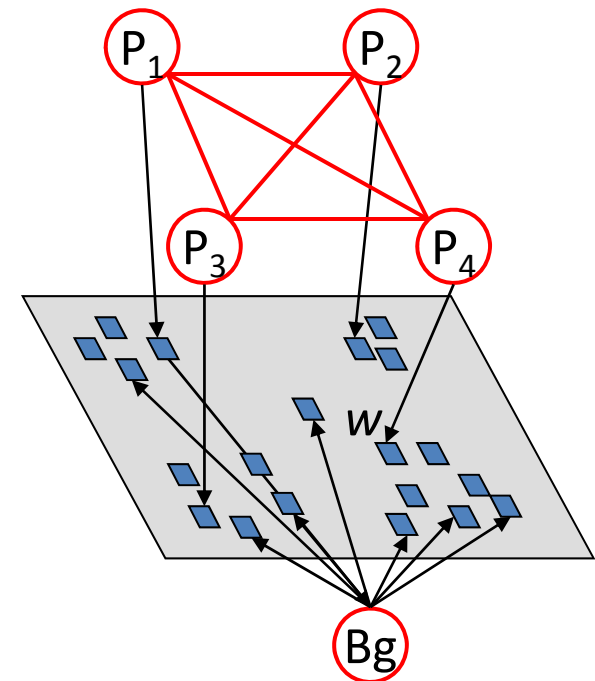
Lowe, 1999

Matching a class of objects...

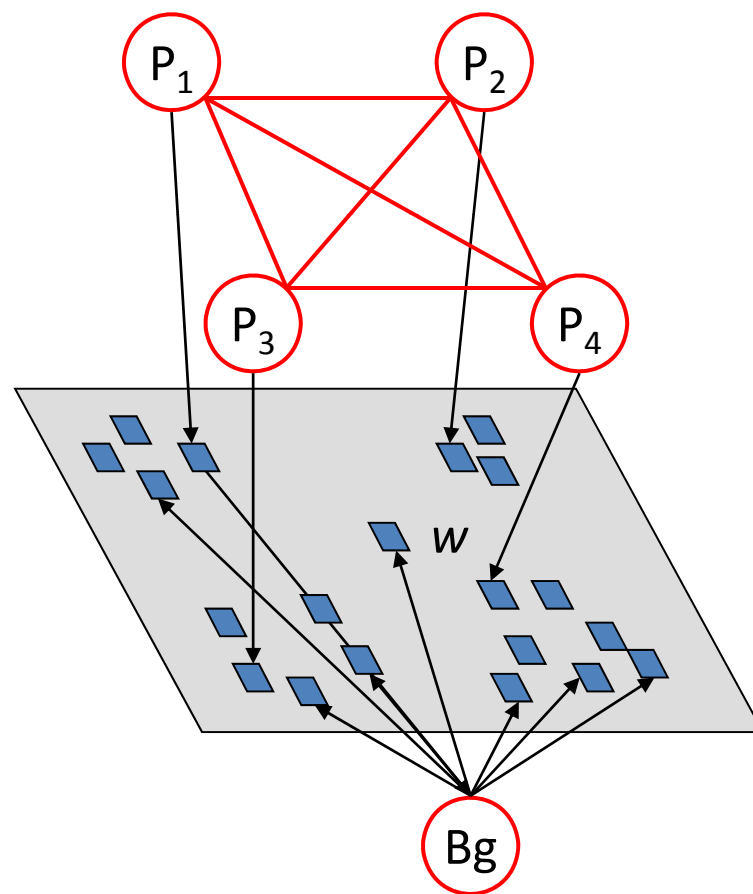


Part-based representation: constellation model

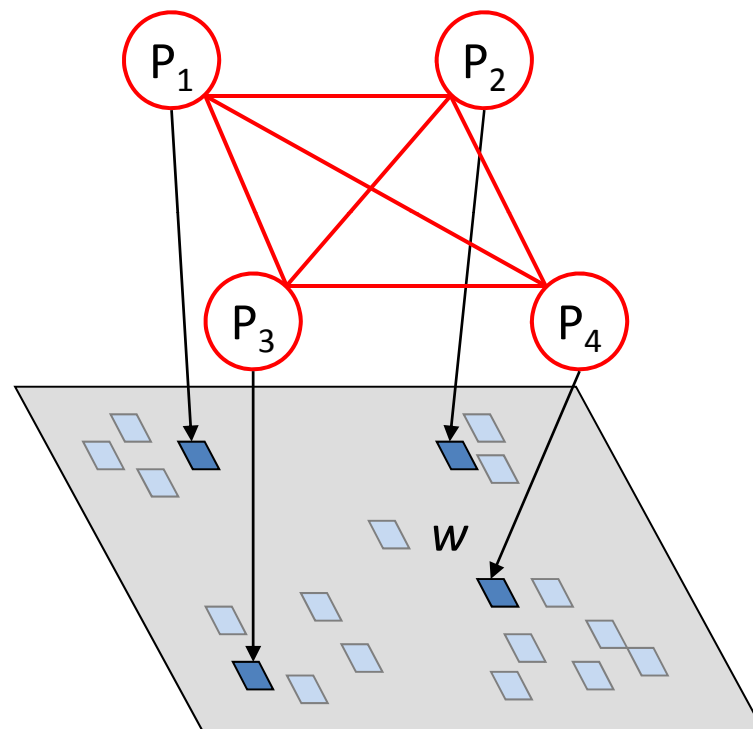
- Fischler & Elschlager 1973
- Yuille '91
- Brunelli & Poggio '93
- Lades, v.d. Malsburg et al. '93
- Cootes, Lanitis, Taylor et al. '95
- Amit & Geman '95, '99
- Perona et al. '95, '96, '98, '00, '03, '04, '05
- Felzenszwalb & Huttenlocher '00, '04
- Crandall & Huttenlocher '05, '06
- Leibe & Schiele '03, '04
- Many papers since '00



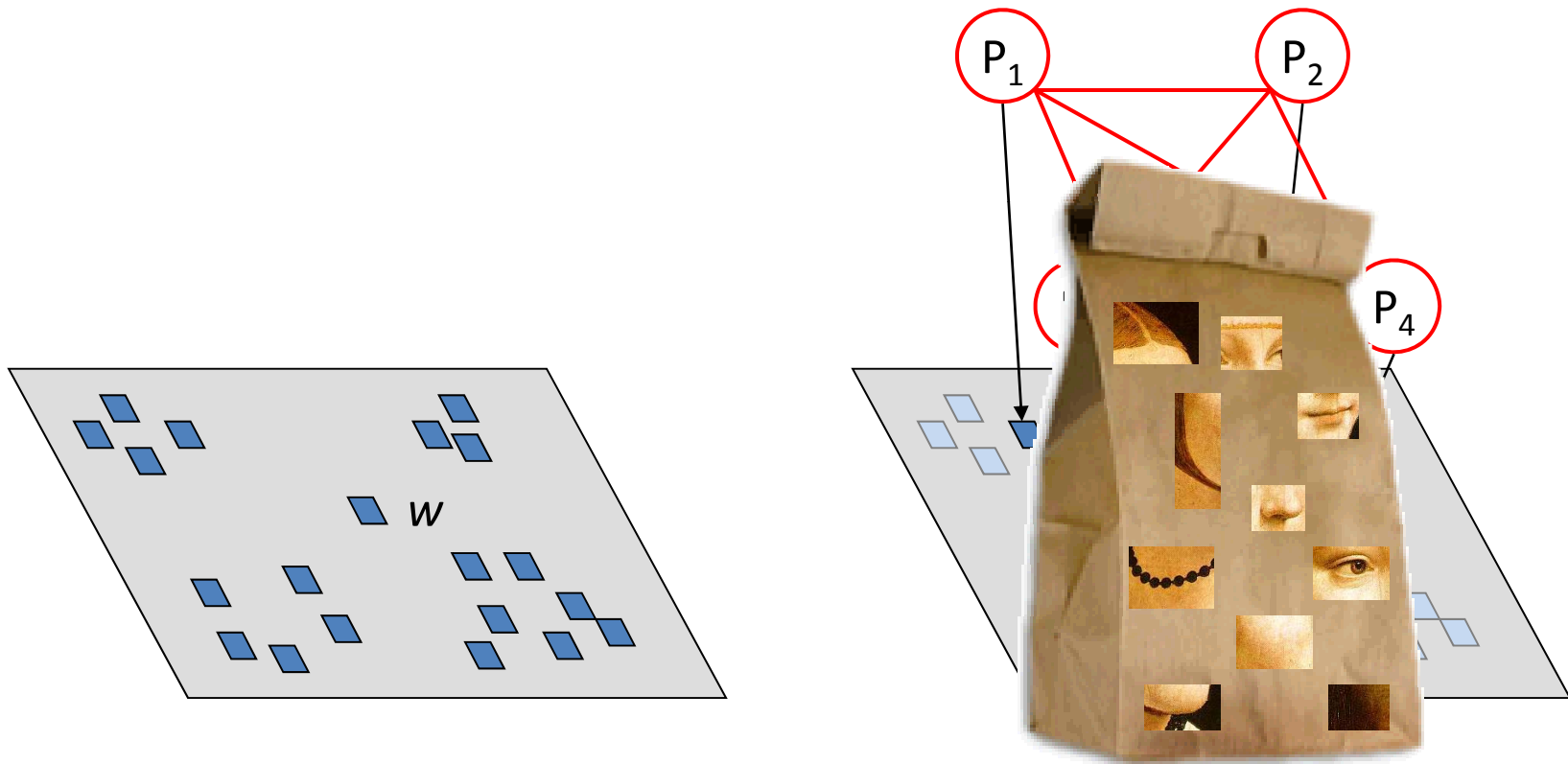
Part-based representation: constellation model



Part-based representation: constellation model

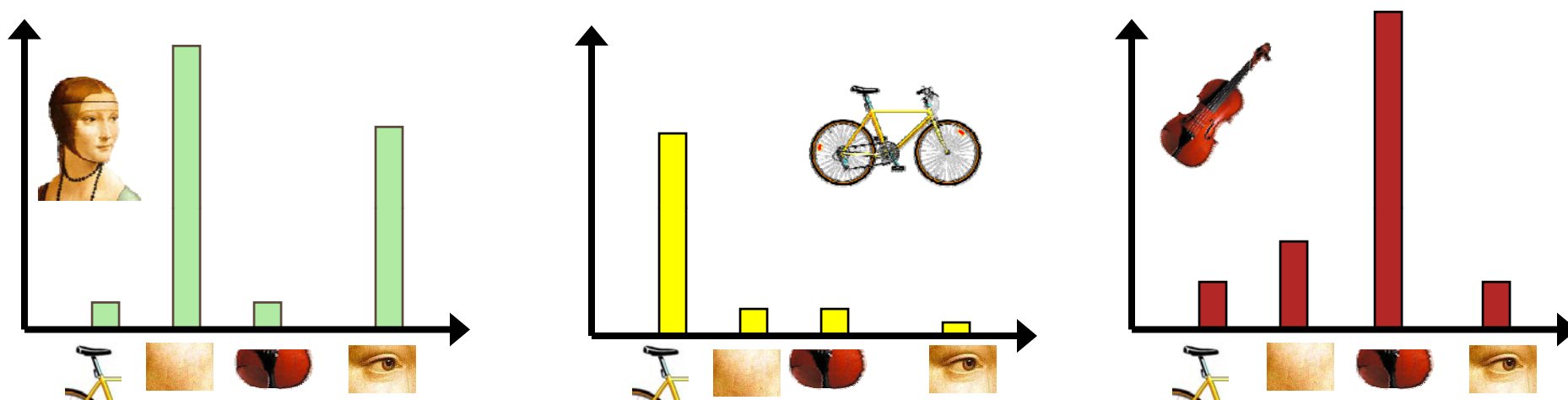


“Bag of words” representation

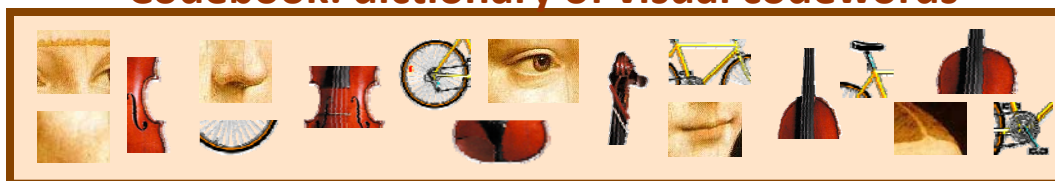


Csurka et al. 2004; earlier work in texture: Leung & Malik, 1999

“Bag of words” representation



Codebook: dictionary of visual codewords



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

Analogy to textual documents

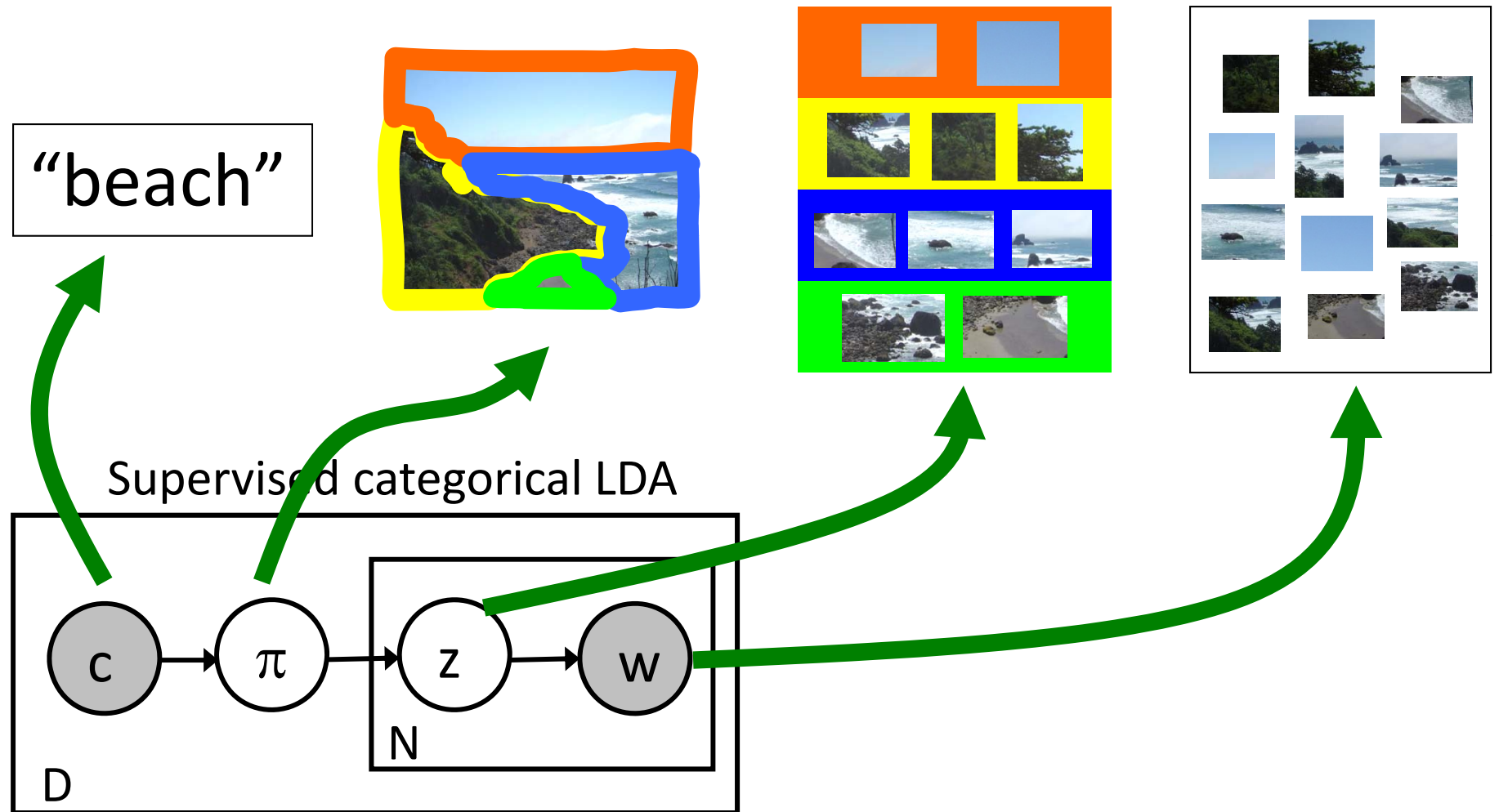
Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially on the messages that reach our eyes. For a long time, the retinal image was considered as a movie screen. It is now discovered that the visual centers in the brain are like a more complex system following the path of the message to the various cells of the cortex. Hubel and Wiesel have demonstrated that the *message about the image falling on the retina undergoes a point-by-point analysis in a system of nerve cells stored in columns. In this system each cell has its specific function and is responsible for a specific detail in the pattern of the retinal image.*

**sensory, brain,
visual, perception,
retinal, cerebral cortex,
eye, cell, optical
nerve, image
Hubel, Wiesel**

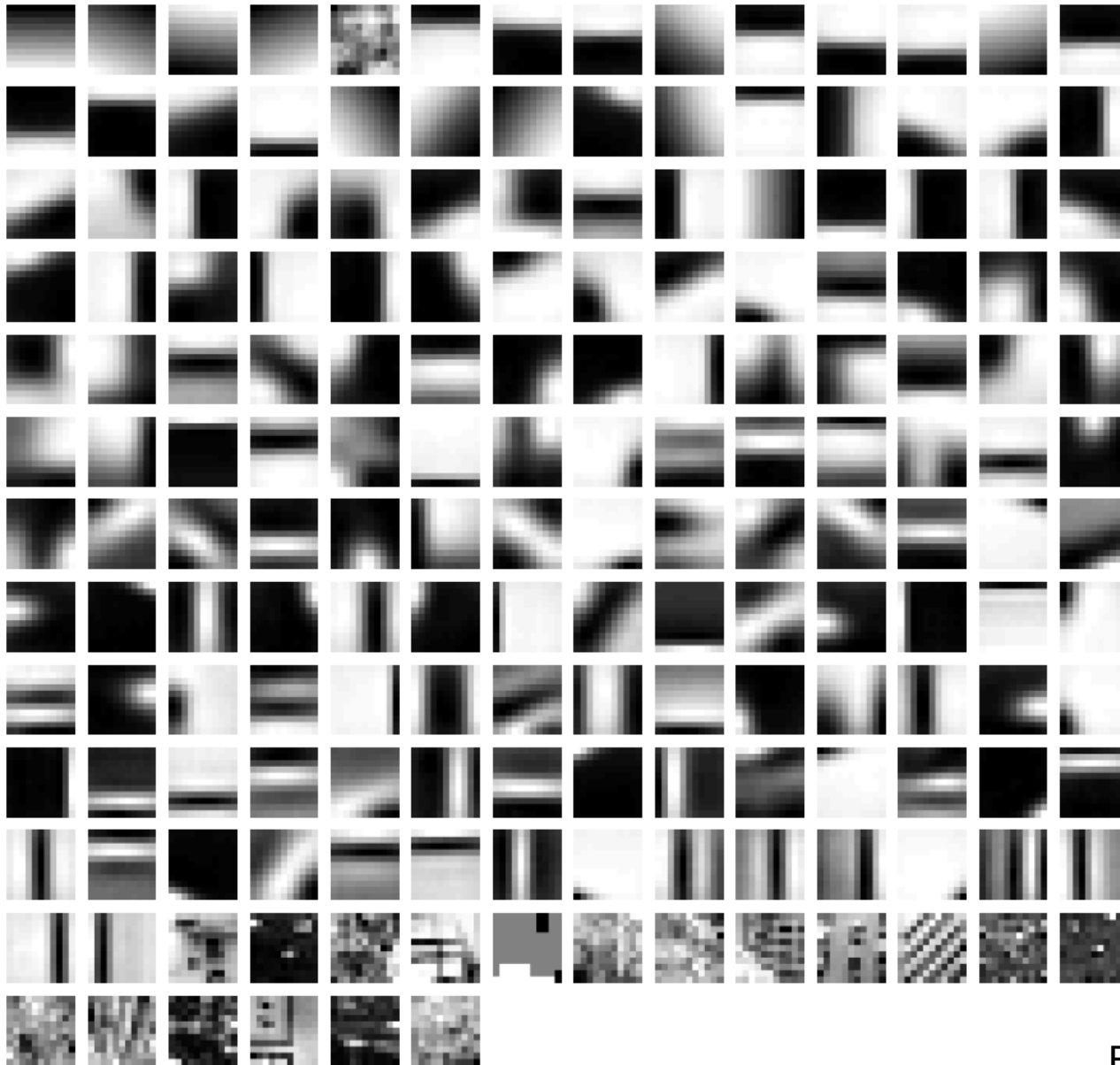
China is forecasting a trade surplus of \$90bn (£51bn) to \$100bn this year, a threefold increase on 2004's \$32bn. The Commerce Ministry said the surplus would be created by a predicted 30% increase in exports to \$750bn, compared with \$560bn in 2004. The increase will annoy the US because it will reduce China's trade deficit. China's government has agreed to let the yuan rise against the dollar, but the government also needs to keep the yuan's value low to meet the demand for exports. China has said it will let the yuan rise against the dollar, but it has not permitted it to trade within a narrow band. The US wants the yuan to be allowed to rise freely. However, Beijing has made it clear that it will take its time and tread carefully before allowing the yuan to rise further in value.

**China, trade,
surplus, commerce,
exports, imports, US,
yuan, bank, domestic,
foreign, increase,
trade, value**

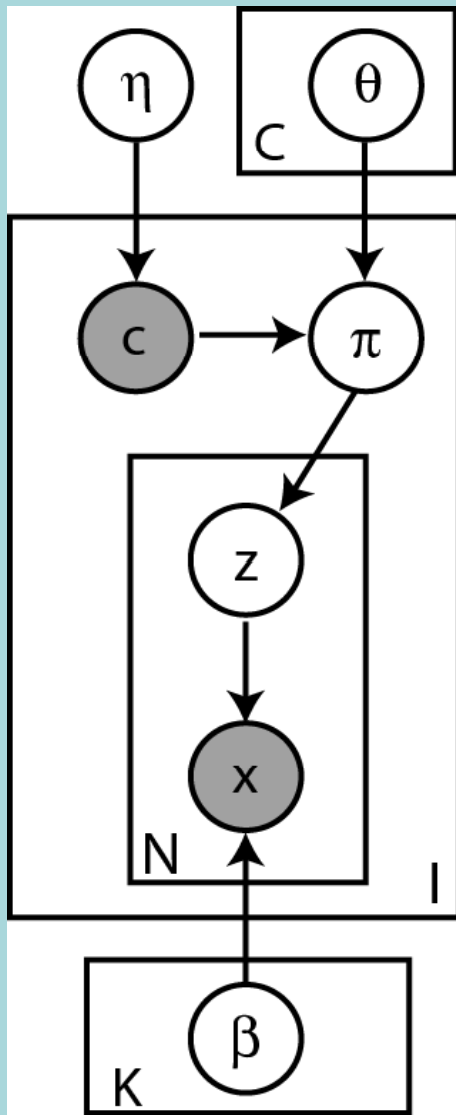
Natural scene categorization



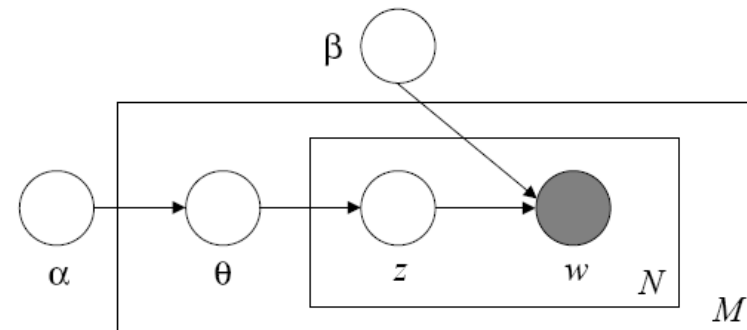
codebook



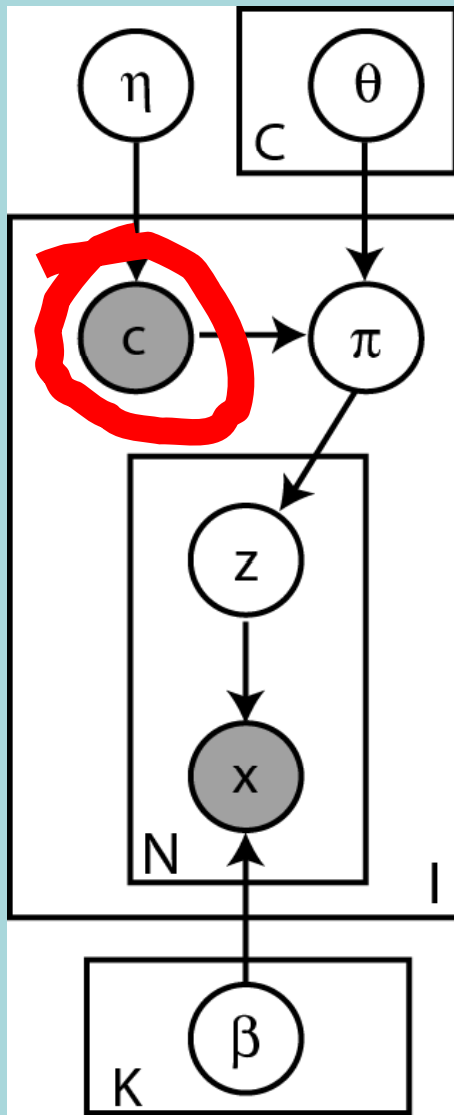
A Generative Model



LDA: Blei, Ng, & Jordan. 2003



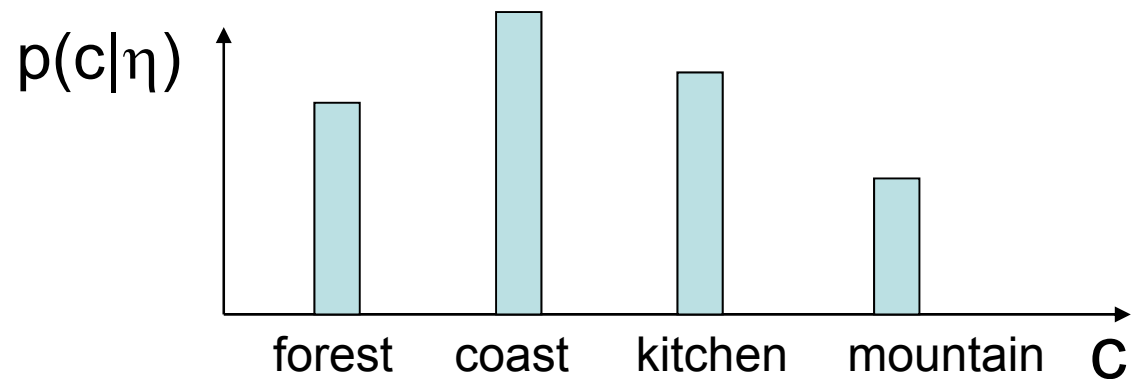
A Generative Model



scene category

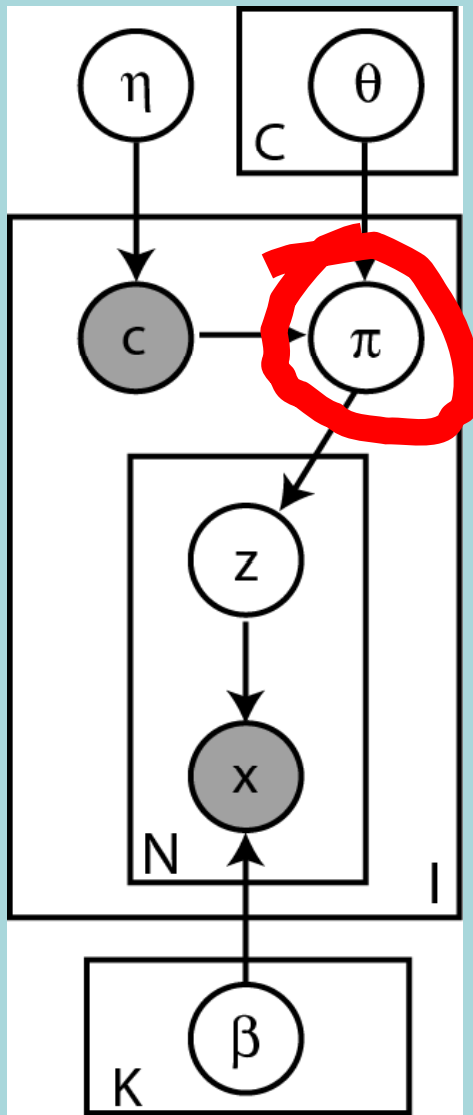


discrete variable: $c \sim p(c|\eta)$



Fei-Fei & Perona (CVPR 2005)

A Generative Model

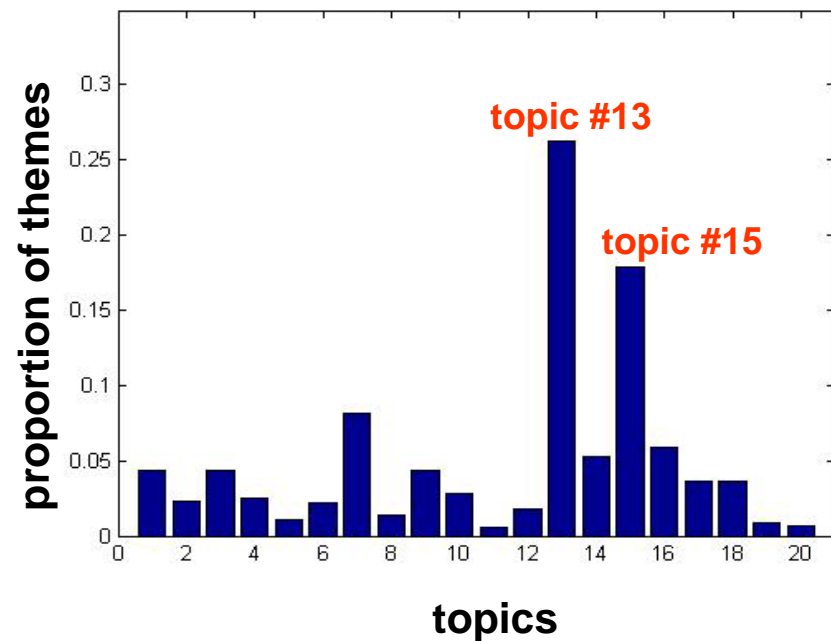


mixing parameter for the latent topics

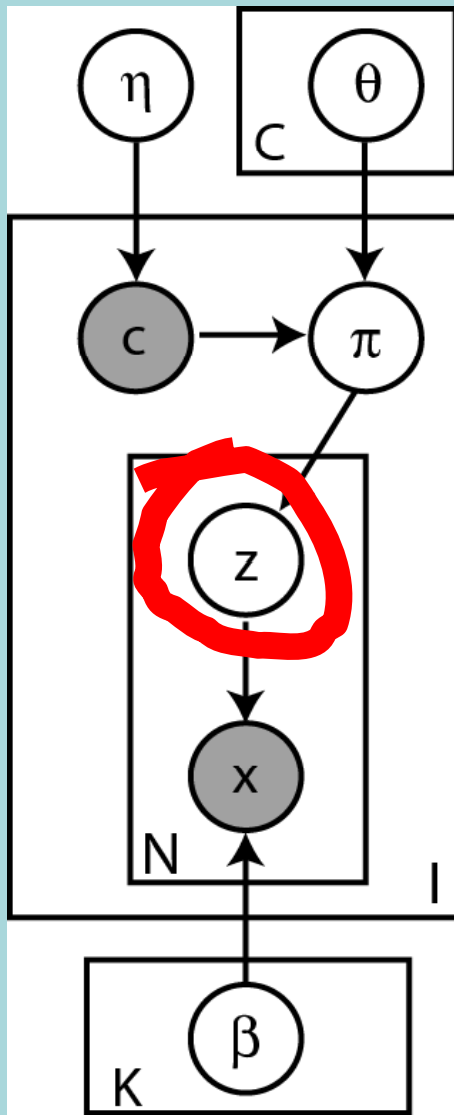


$$\pi \sim p(\pi | c, \theta) \\ \sim \text{Dir}(\pi | c, \theta)$$

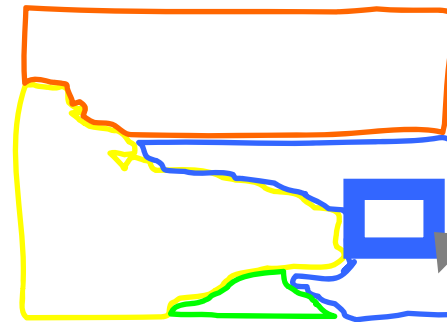
$$\text{where } \sum_{k=1}^K \pi_k = 1$$



A Generative Model



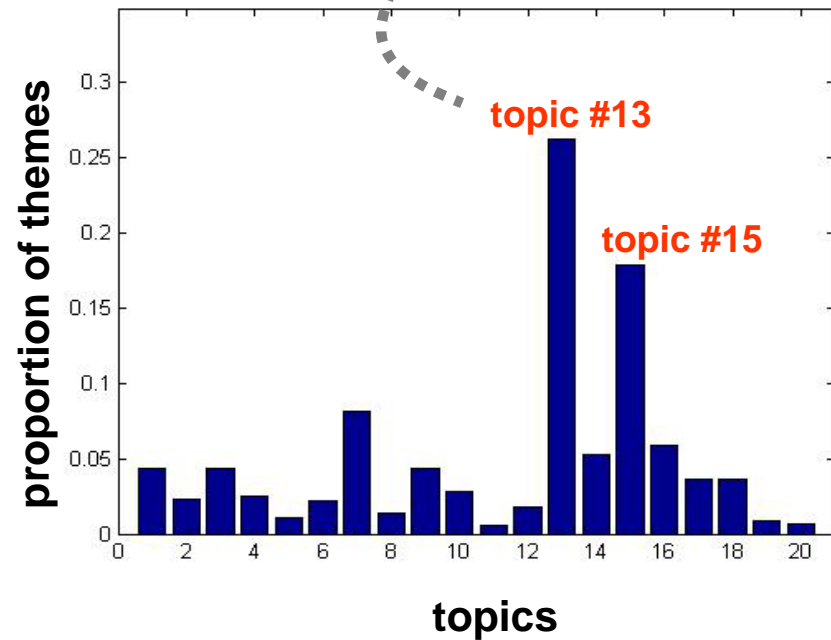
topic label



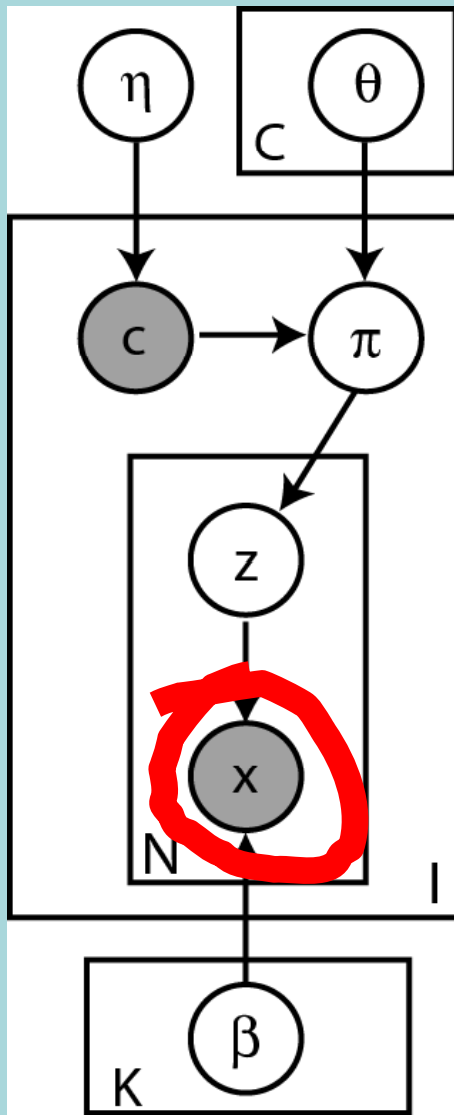
discrete variable:

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

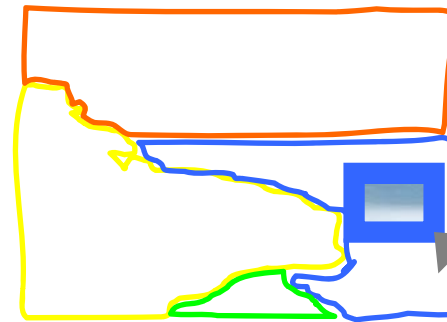
$$\sim \text{Mult}(z|\pi)$$



A Generative Model



patch label

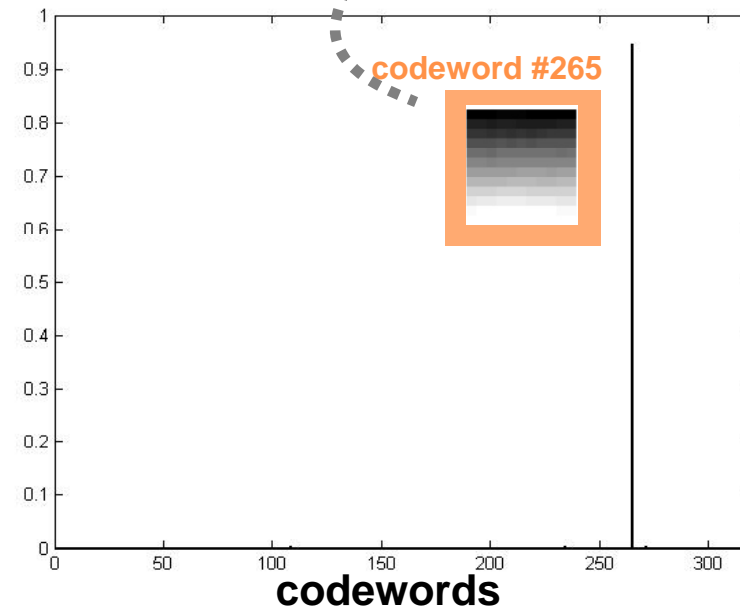


discrete variable:

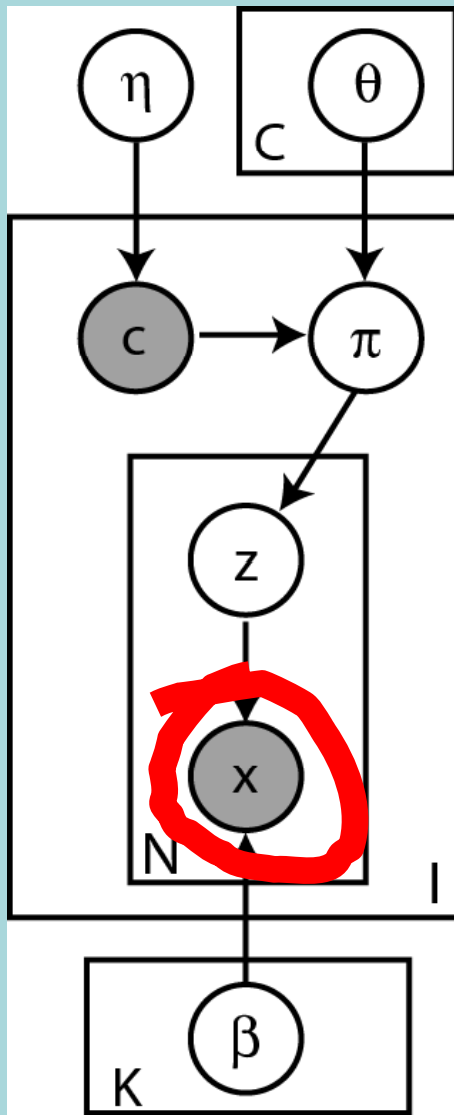
$$x \sim p(x|z, \beta)$$

$$\sim \text{Mult}(x|z, \beta)$$

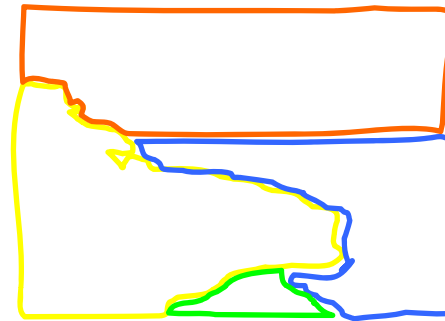
expected value of β given 'z=13'



A Generative Model



patch label

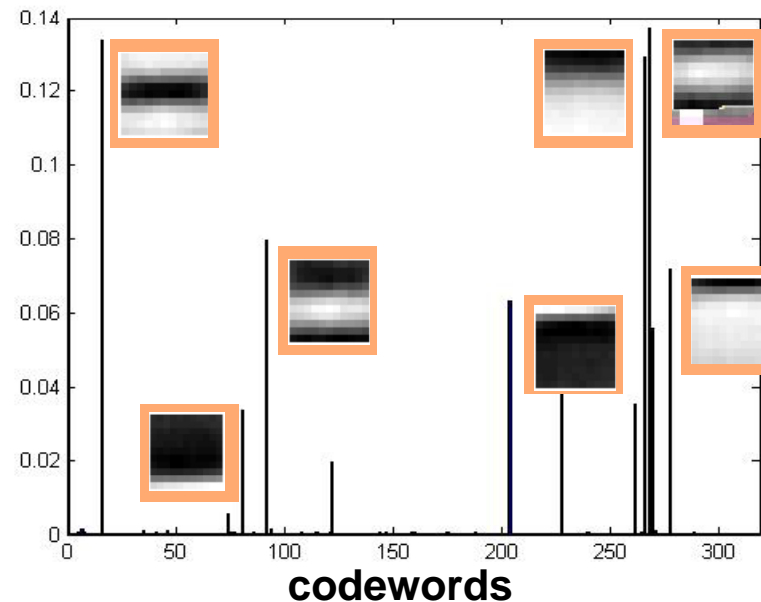


discrete variable:

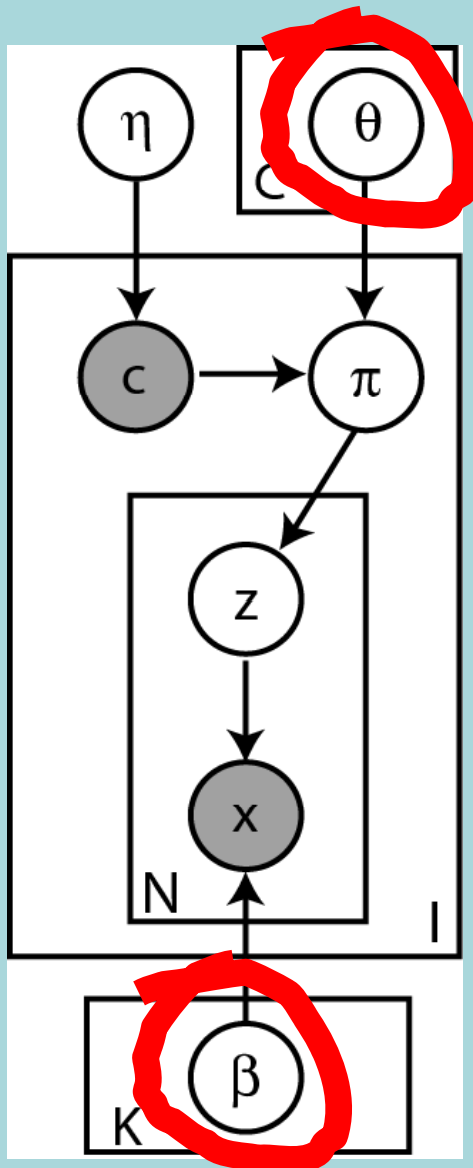
$$x \sim p(x|z, \beta)$$

$$\sim \text{Mult}(x|z, \beta)$$

expected value of β given 'z=15'



A Generative Model



learning

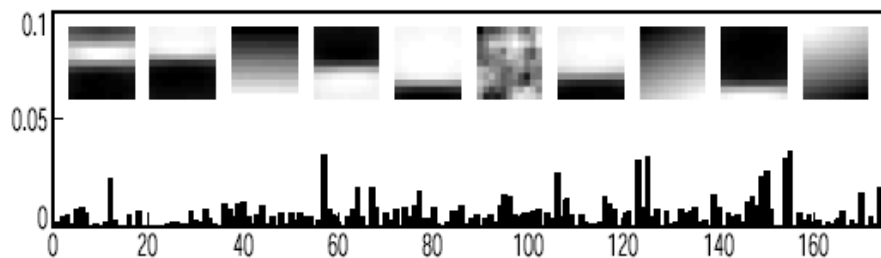
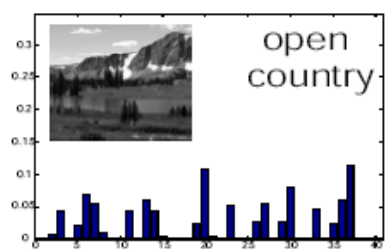
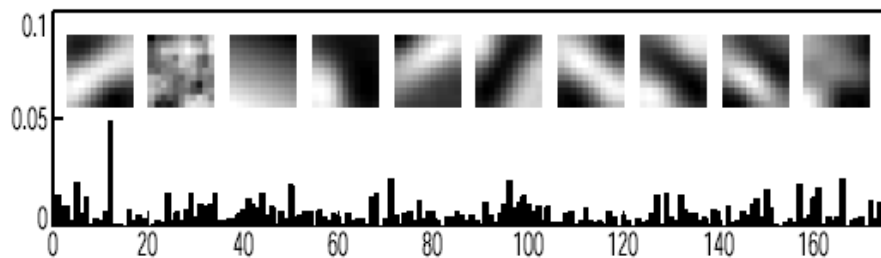
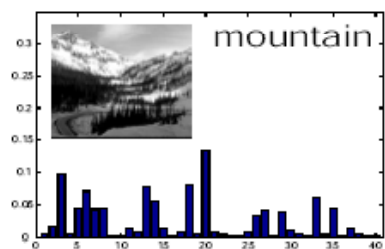
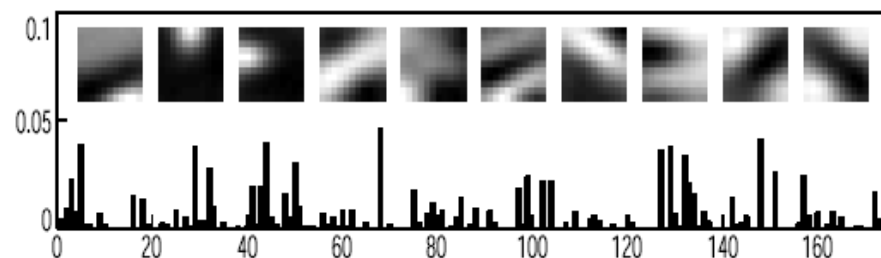
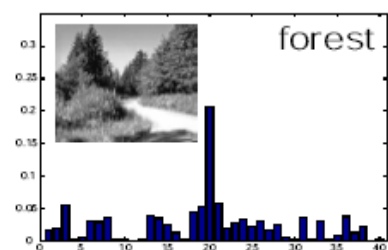
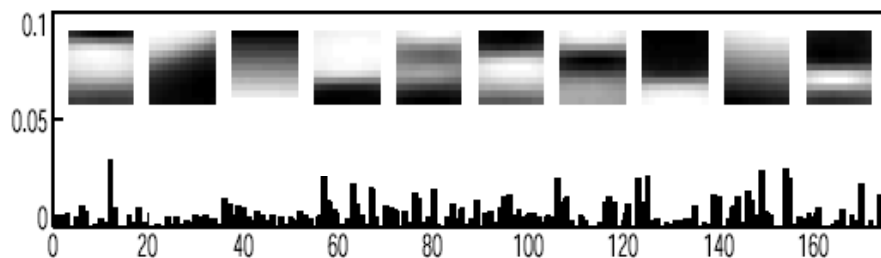
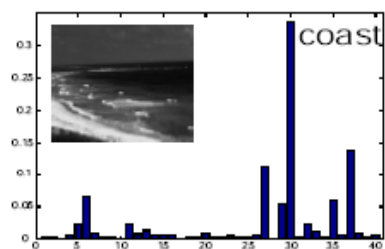
Find the 'best' θ and β

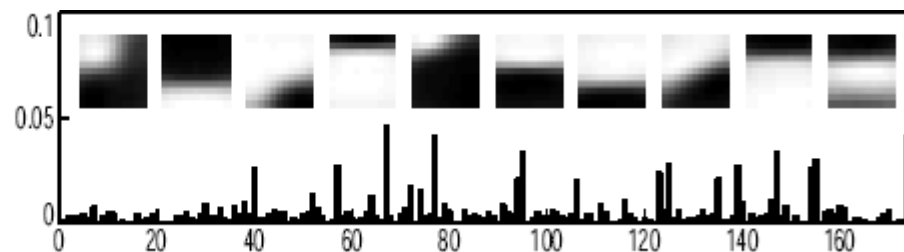
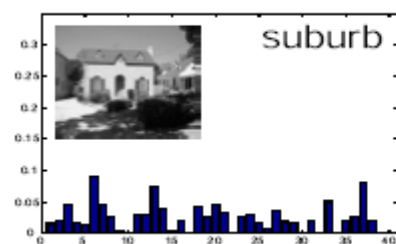
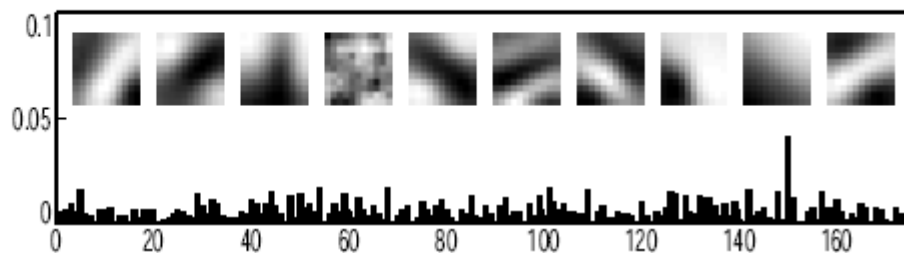
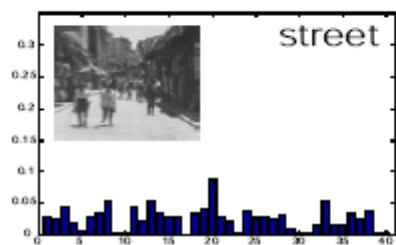
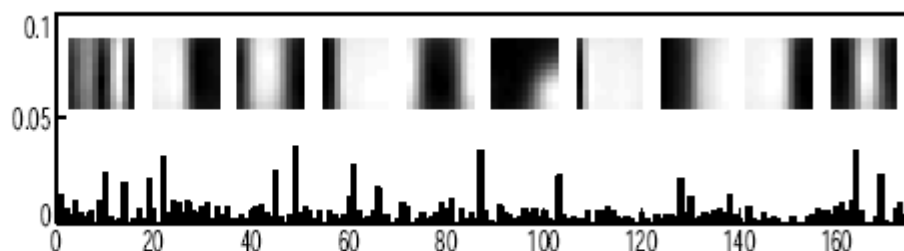
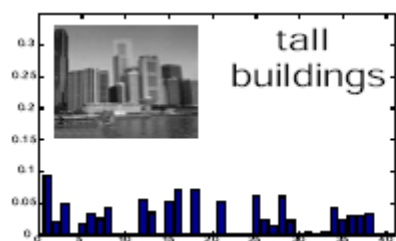
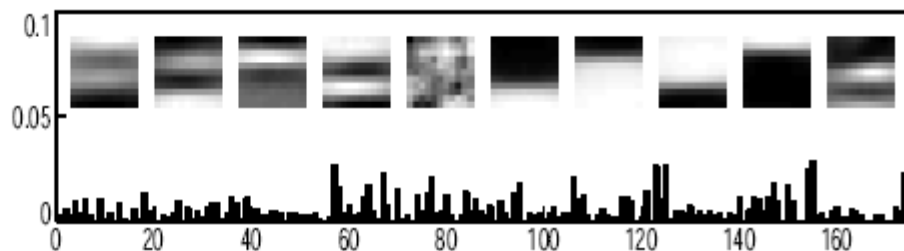
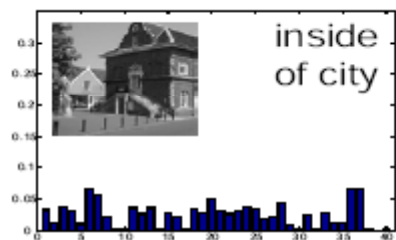
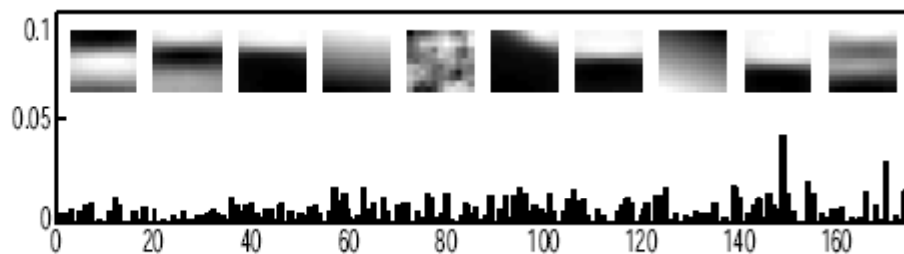
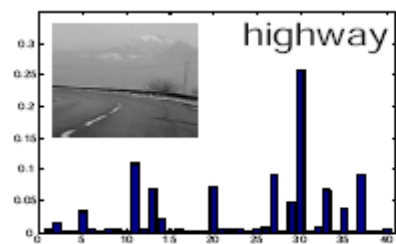
joint probability

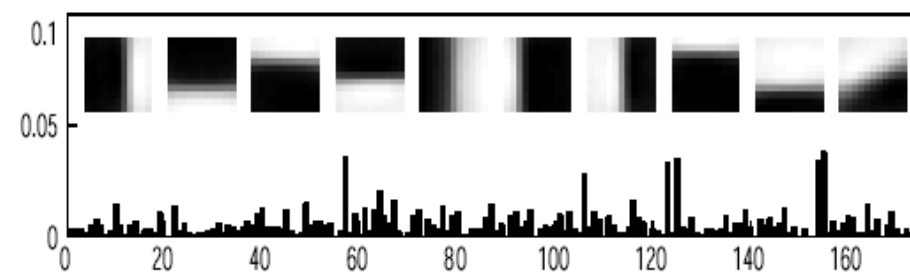
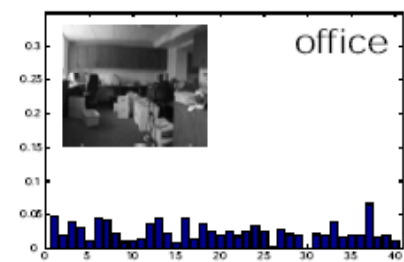
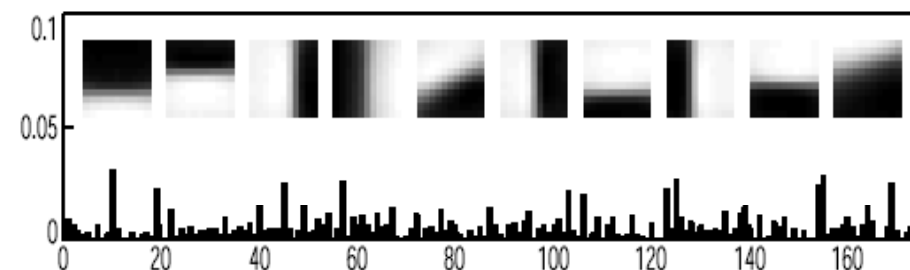
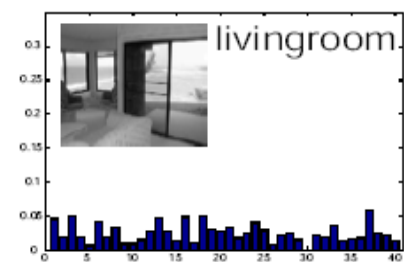
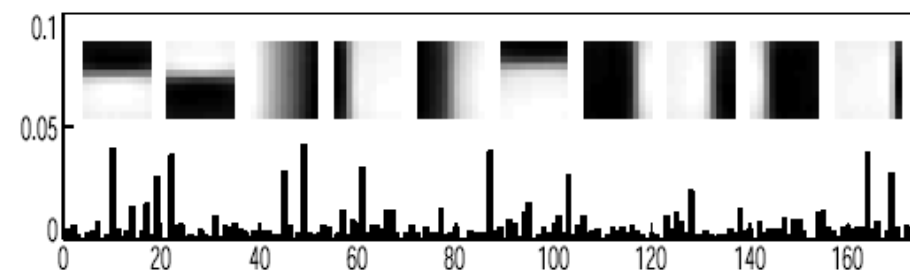
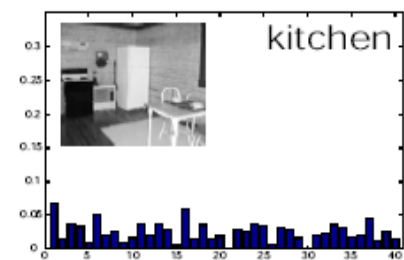
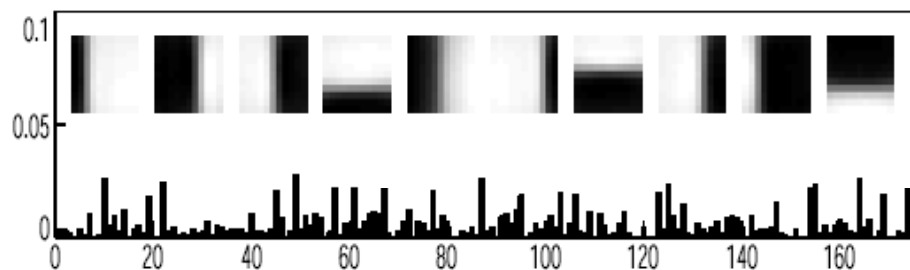
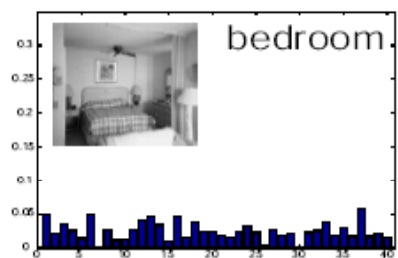
$$p(x, z, \pi | \theta, \beta, c) = p(\pi | c, \theta) \prod_n^N p(z_n | \pi) p(x_n | z_n, \beta)$$

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

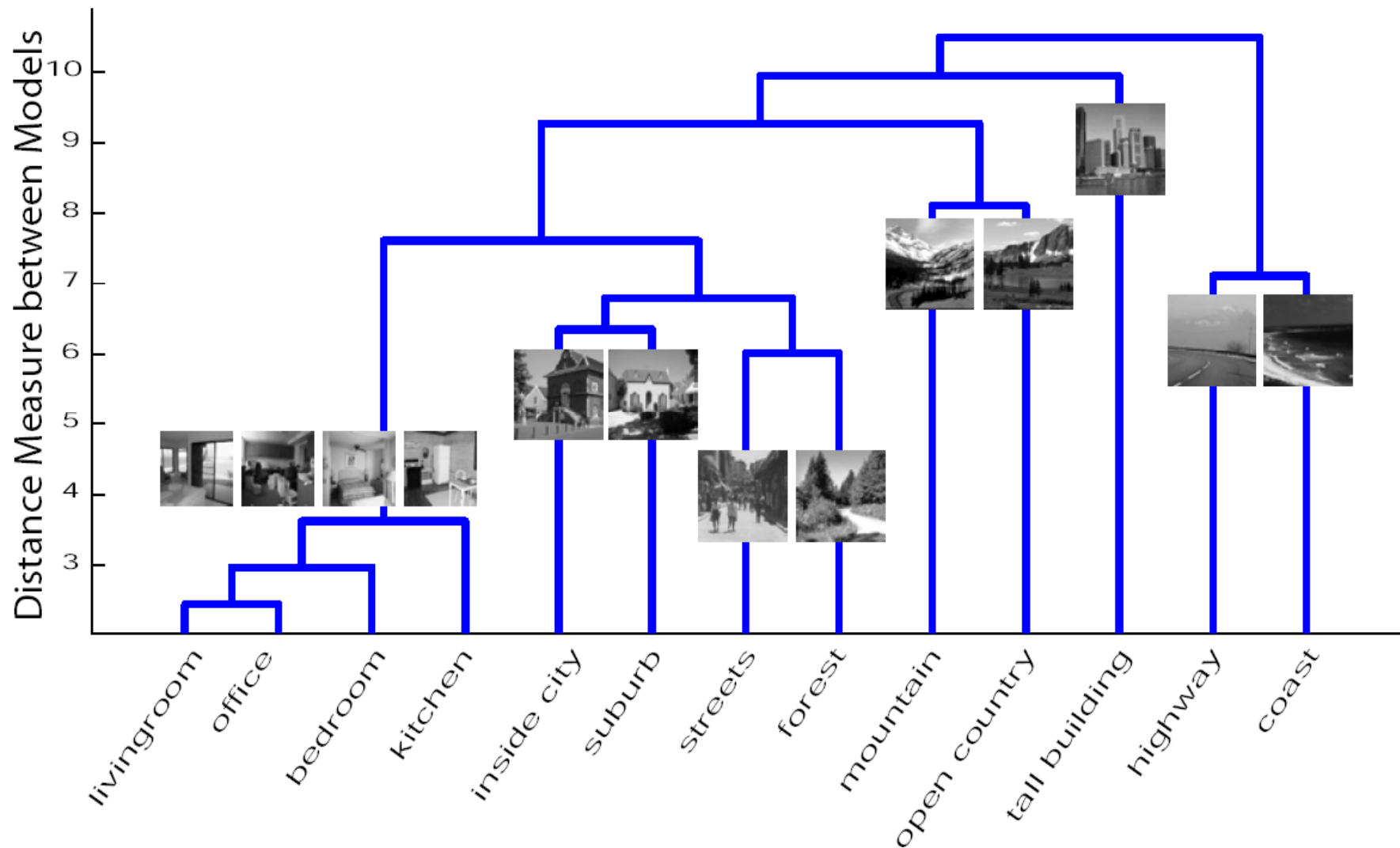
- exact inference is intractable
- use Variational Inference







model distance based on topic distribution

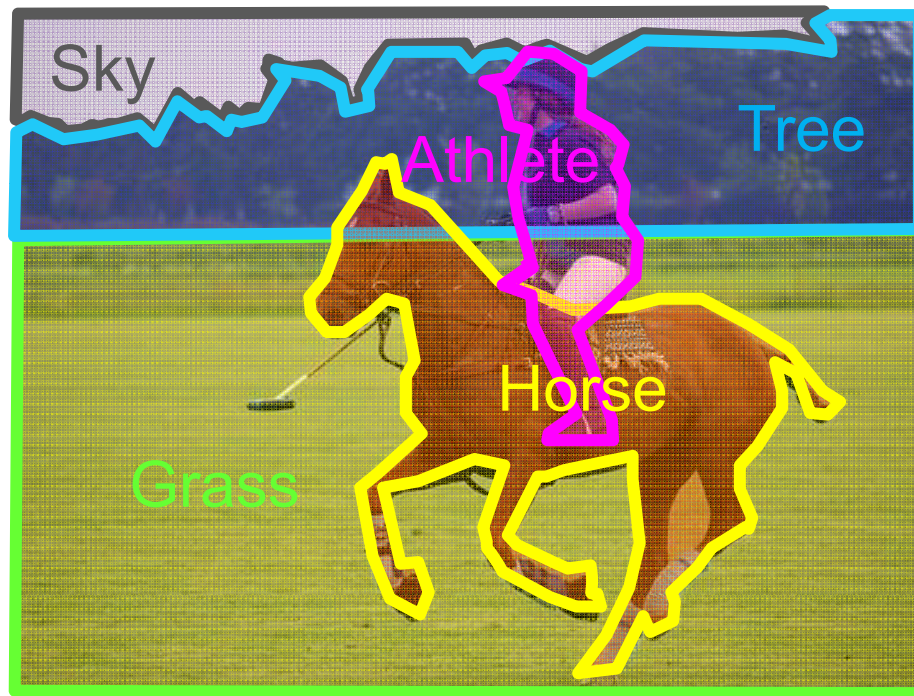


Fei-Fei & Perona (CVPR 2005)

Classification

Annotation

Segmentation

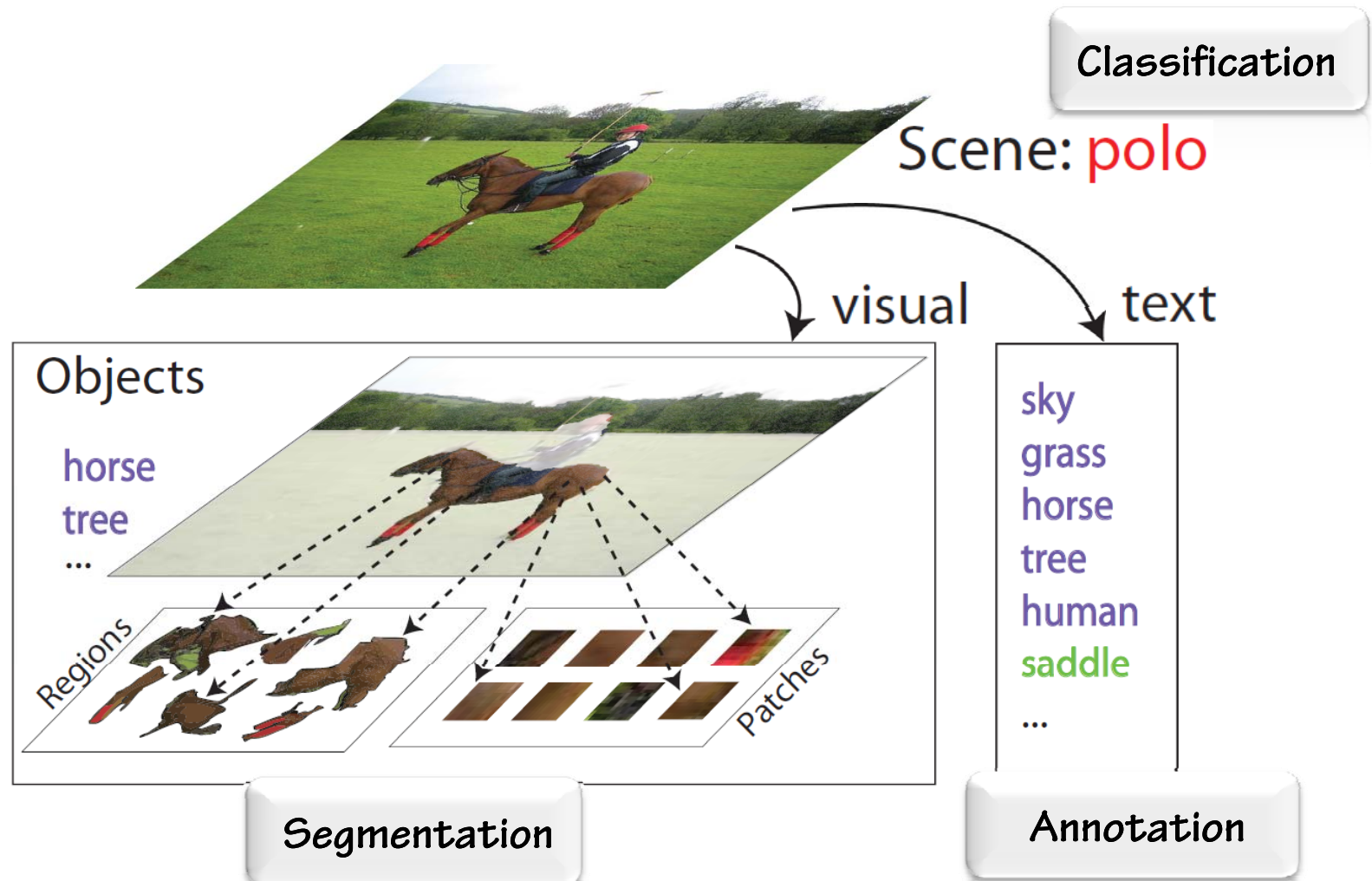


class: **Polo**

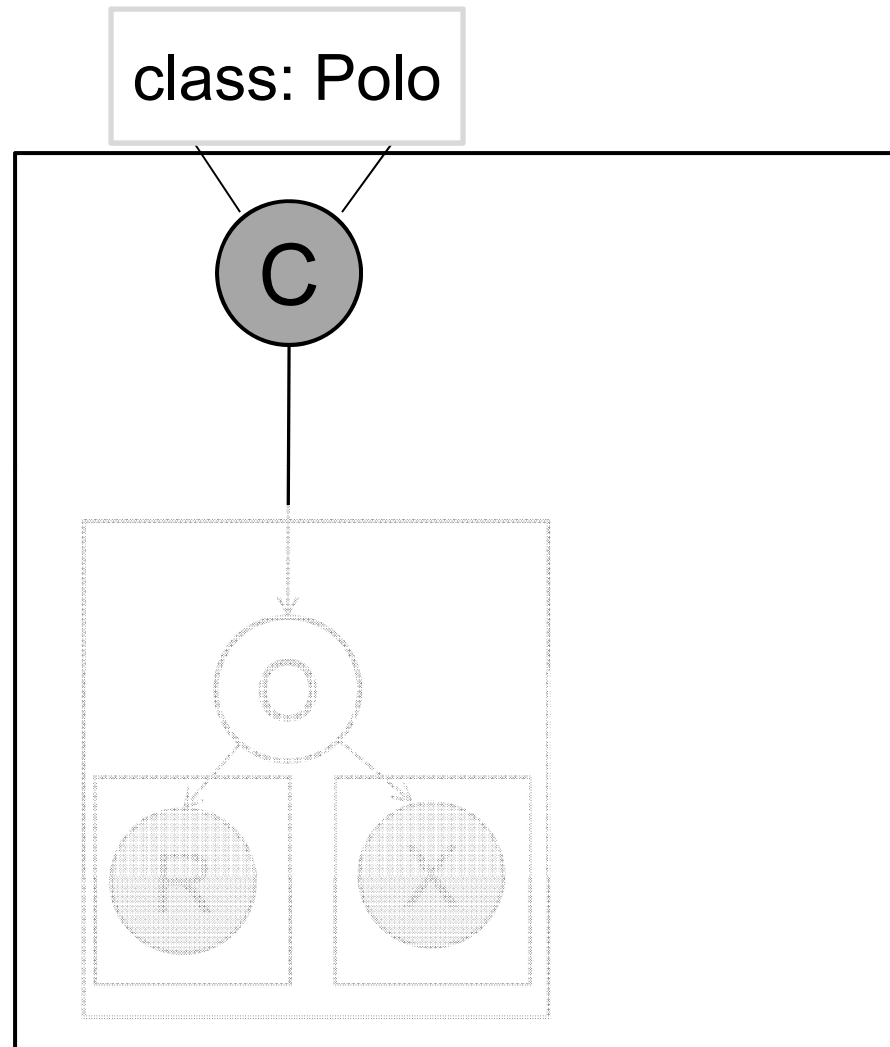
Athlete
Horse
Grass
Trees
Sky
Saddle

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

Towards total scene understanding

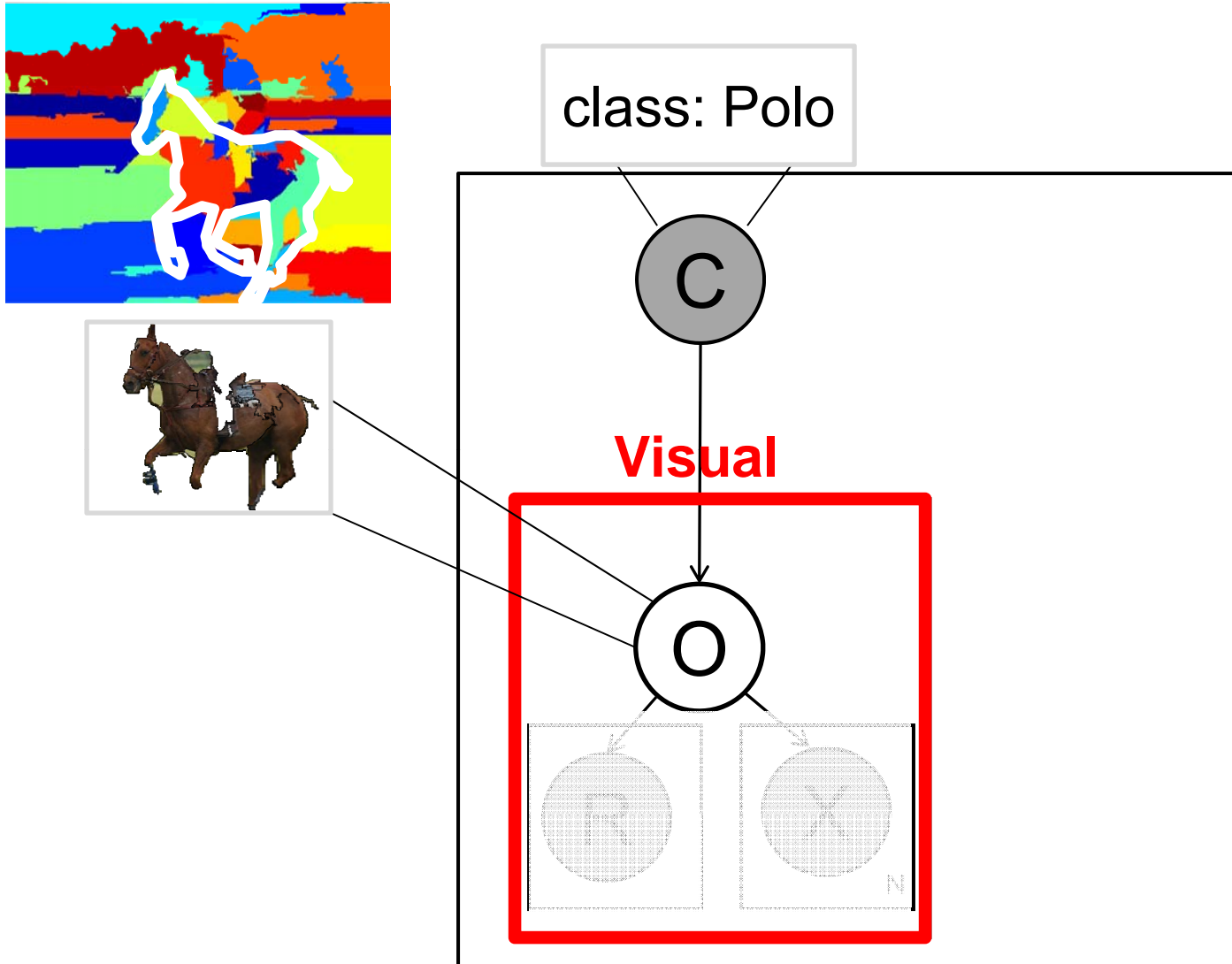


A joint model for image classification, annotation & segmentation



$$p(C, O, R, X, S, T, Z | \eta, \alpha, \beta, \gamma, \theta, \varphi) = p(C) \cdot \begin{matrix} \text{Visual Component} \\ \text{Text Component} \end{matrix}$$

A joint model for image classification, annotation & segmentation



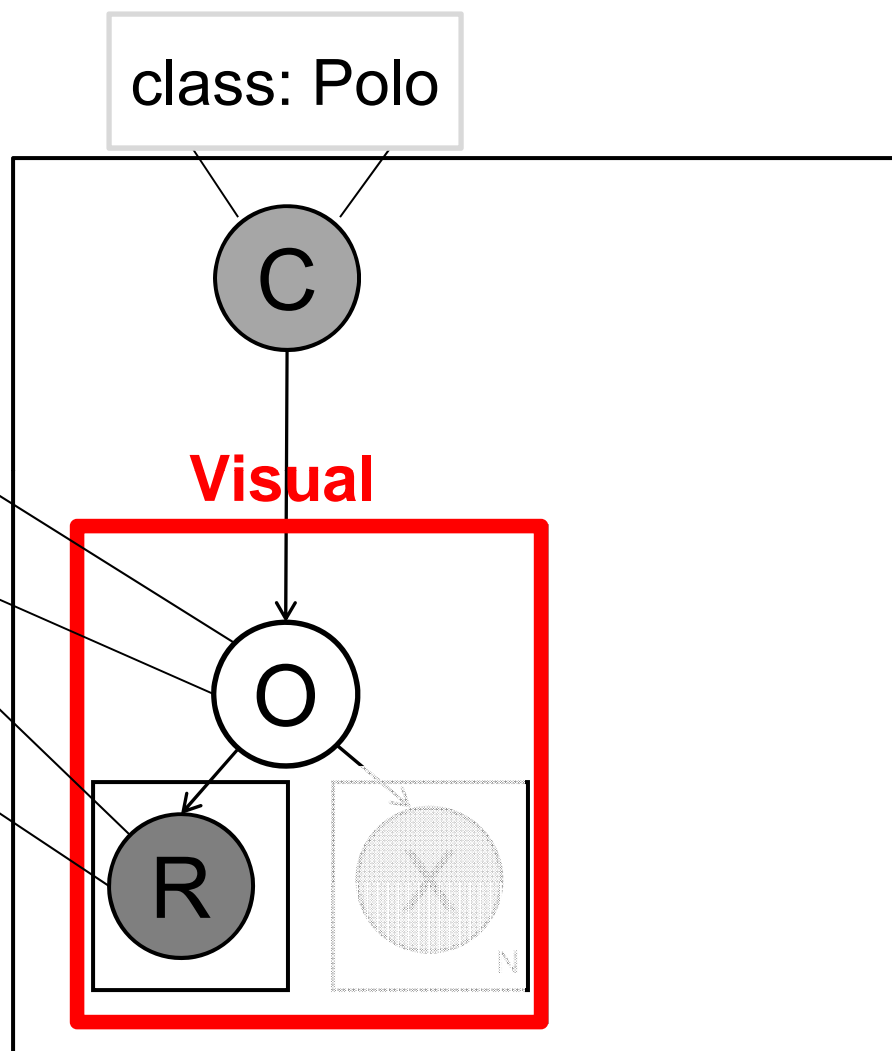
$$p(C, \mathbf{O}, \mathbf{R}, \mathbf{X}, \mathbf{S}, \mathbf{T}, \mathbf{Z} | \eta, \alpha, \beta, \gamma, \theta, \varphi) = p(C) \cdot \left(\prod_{n=1}^{N_r} p(O_n | \eta, C) \right)$$

Text Component

A joint model for image classification, annotation & segmentation



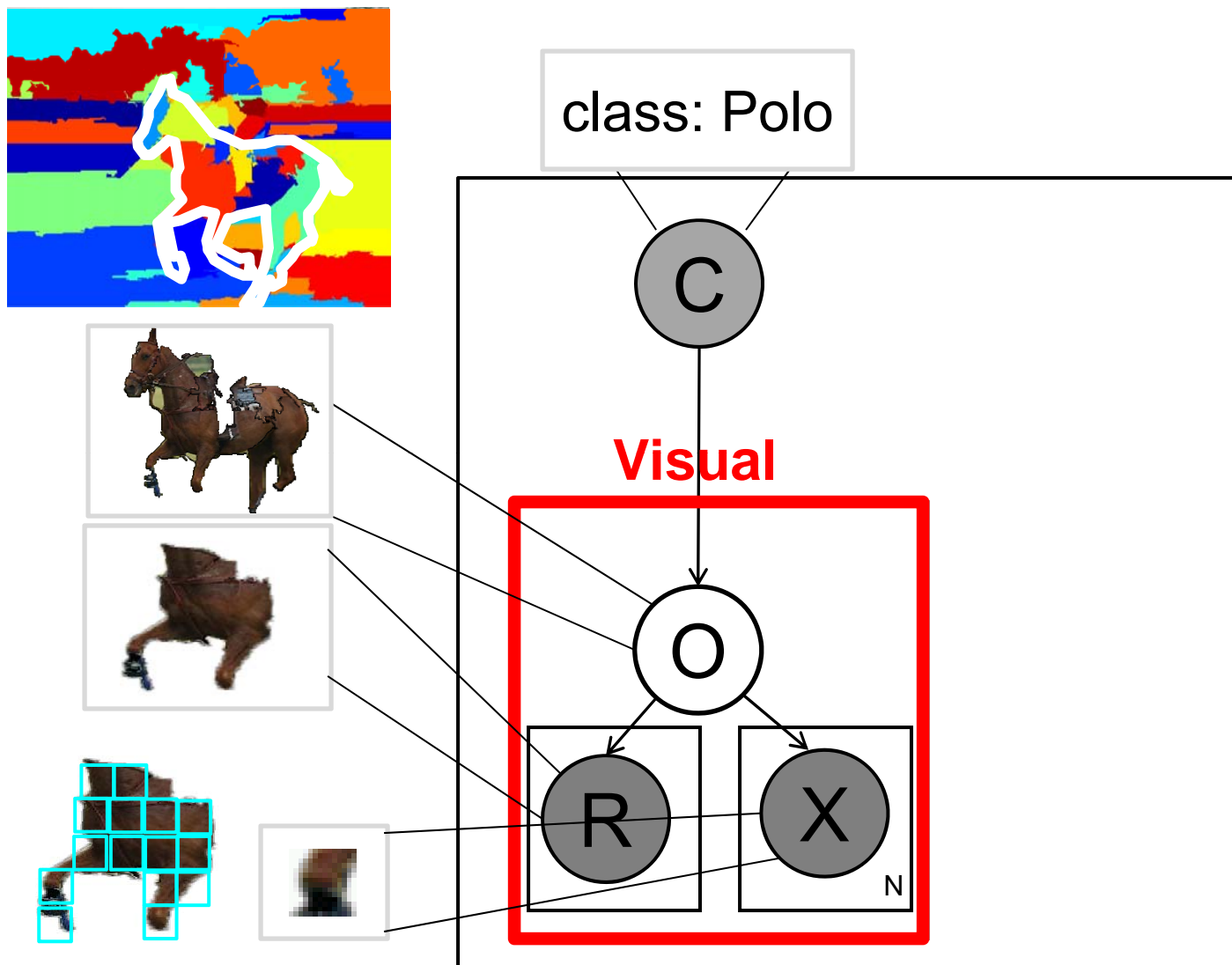
Color Location
Texture Shape



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

Text Component

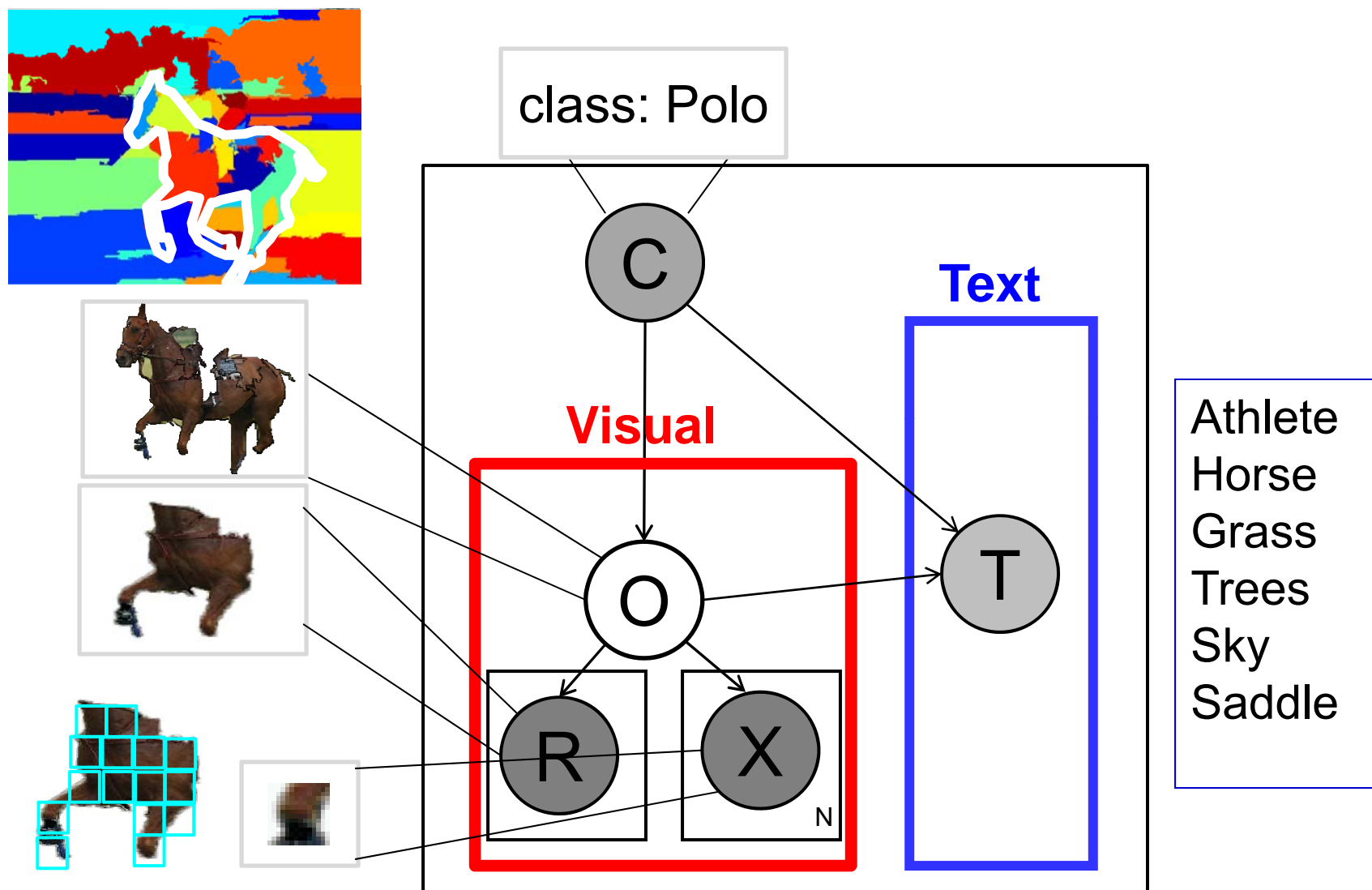
A joint model for image classification, annotation & segmentation



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

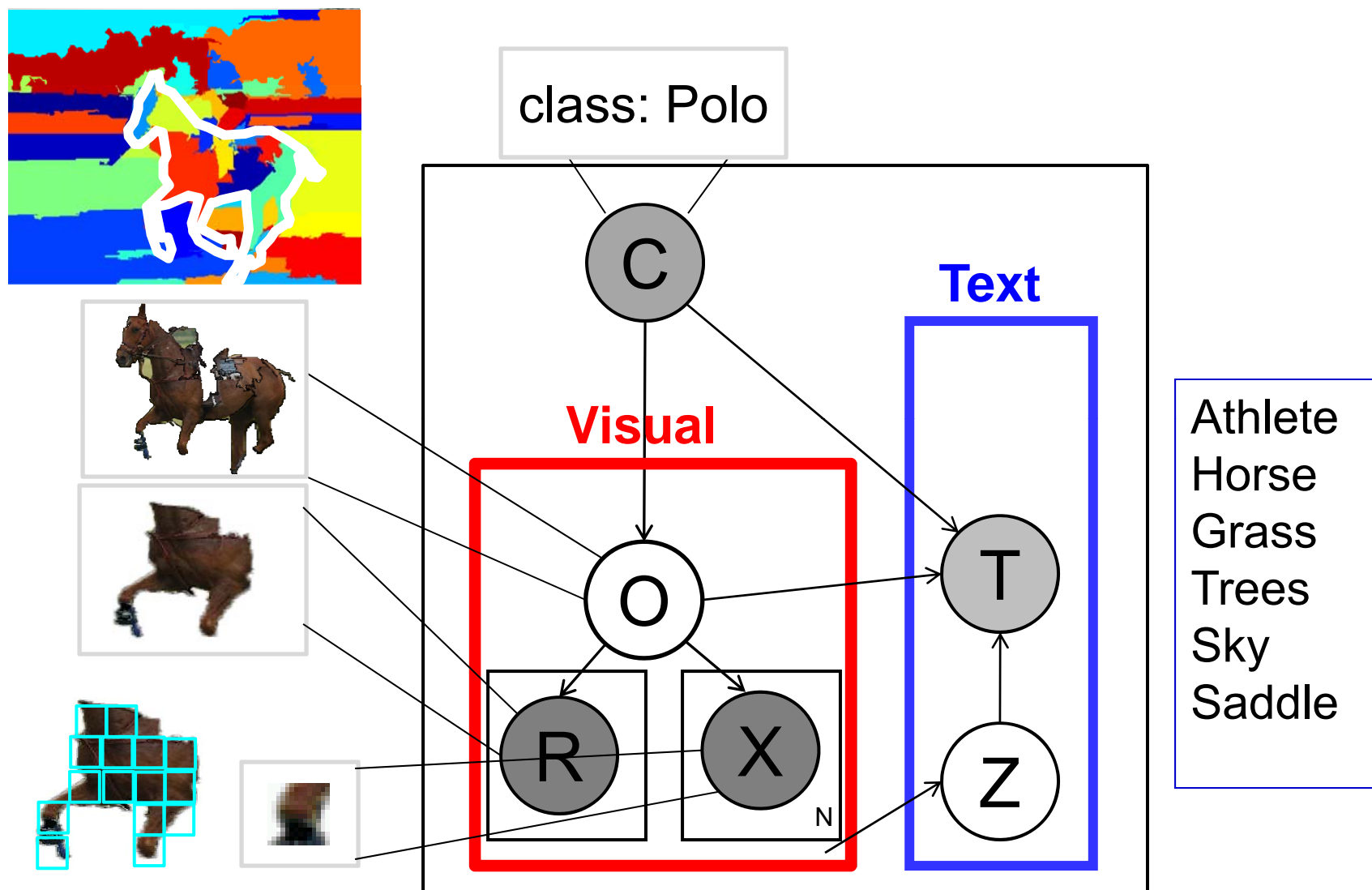
Text Component

A joint model for image classification, annotation & segmentation



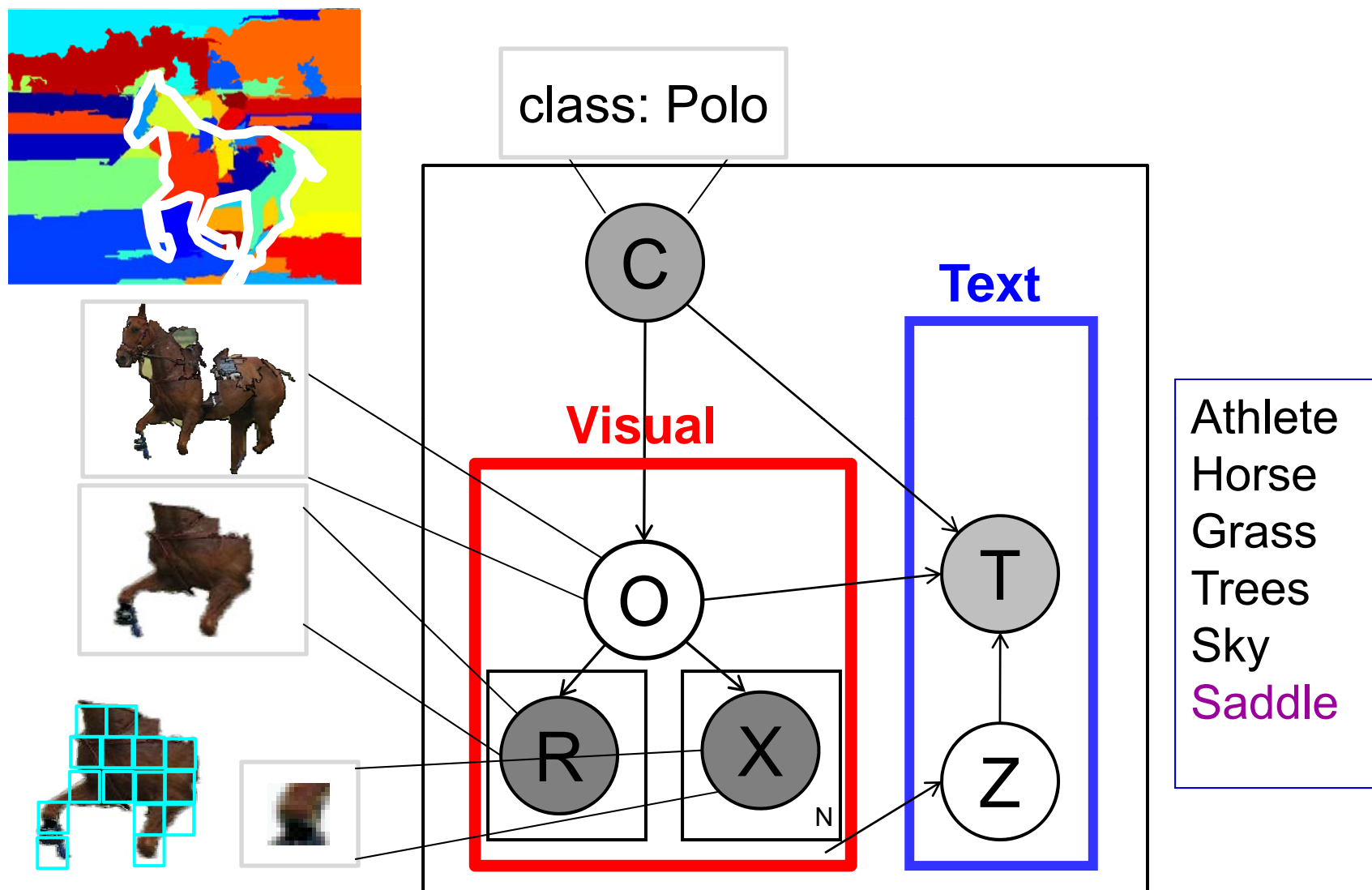
$$p(C, O, R, X, S, T, Z | \eta, \alpha, \beta, \gamma, \theta, \varphi) = p(C) \cdot \left(\prod_{n=1}^{N_r} p(O_n | \eta, C) \right) \prod_{r=1}^{N_r} \left(\left(\prod_{i=1}^{N_F} p(R_{ni} | O_n, \alpha_i) \right) \cdot \prod_{r=1}^{A_r} p(X_{nr} | O_n, \beta) \right) \cdot p(T_m | O_{Z_m}, S_m, \theta,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

A joint model for image classification, annotation & segmentation



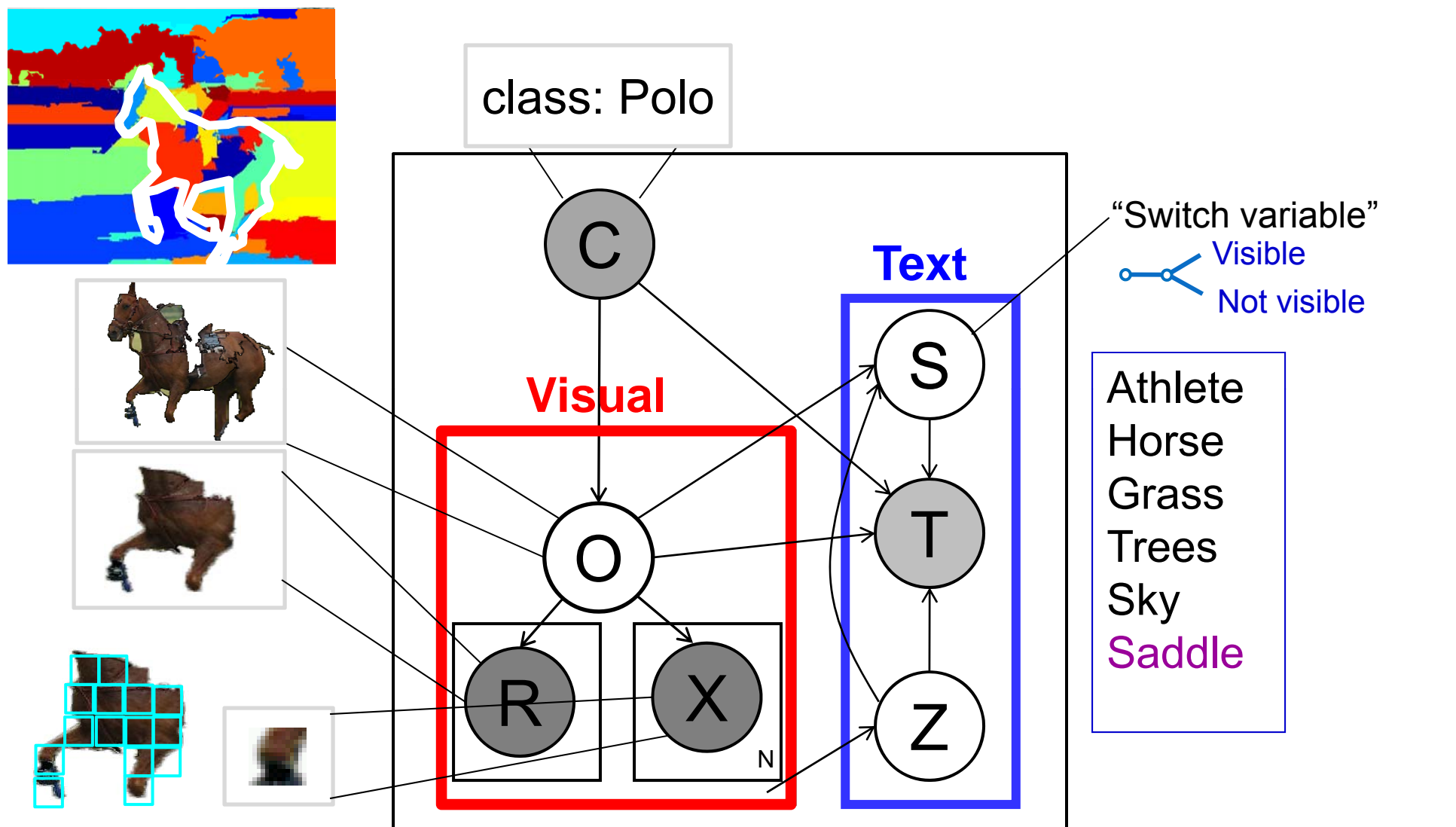
$$p(C, O, R, X, S, T, Z | \eta, \alpha, \beta, \gamma, \theta, \varphi) = p(C) \cdot \left(\prod_{n=1}^{N_r} p(O_n | \eta, C) \right) \prod_{p=1}^{N_r} \left(\left(\prod_{i=1}^{N_F} p(R_{ni} | O_n, \alpha_i) \right) \cdot \prod_{r=1}^{A_r} p(X_{nr} | O_n, \beta) \right) \cdot \prod_{m=1}^{N_t} p(Z_m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

A joint model for image classification, annotation & segmentation



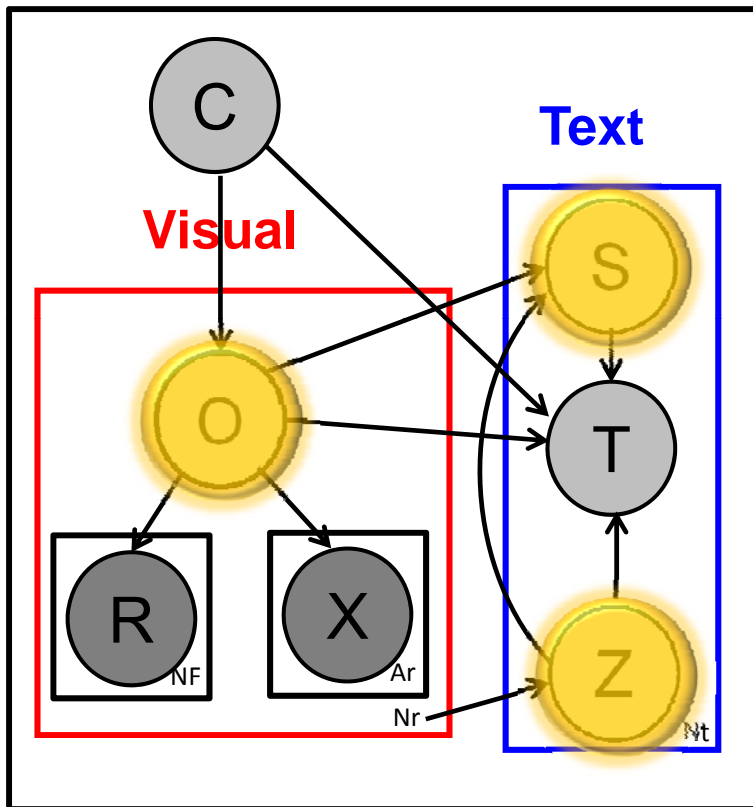
$$p(C, O, R, X, S, T, Z | \eta, \alpha, \beta, \gamma, \theta, \varphi) = p(C) \cdot \left(\prod_{n=1}^{N_r} p(O_n | \eta, C) \right) \prod_{p=1}^{N_r} \left(\left(\prod_{i=1}^{N_F} p(R_{ni} | O_n, \alpha_i) \right) \cdot \prod_{r=1}^{A_r} p(X_{nr} | O_n, \beta) \right) \cdot \prod_{m=1}^{N_t} p(Z_m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

A joint model for image classification, annotation & segmentation



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

Learning



Exact Inference is Intractable!

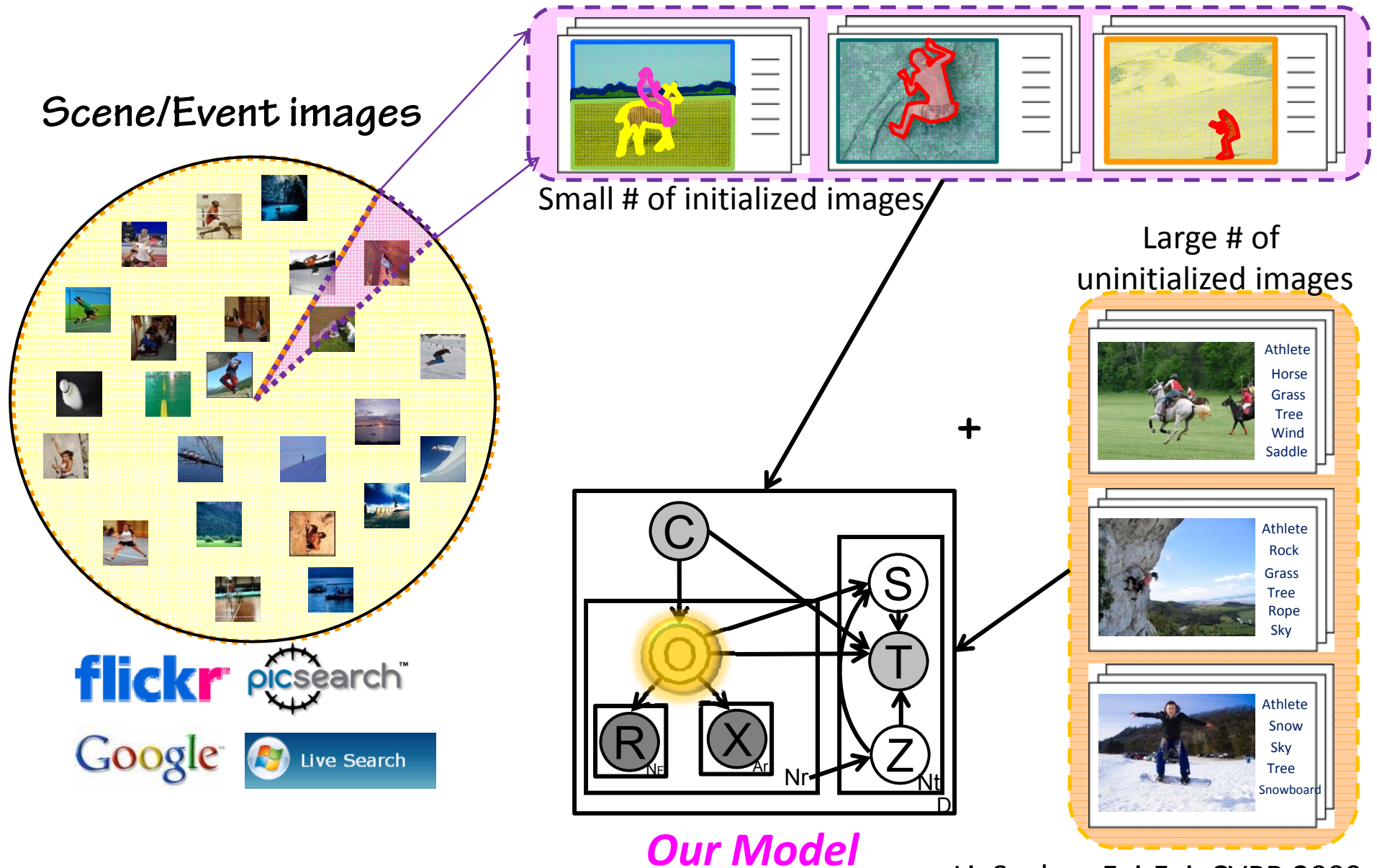
Relationship of the random variables

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

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

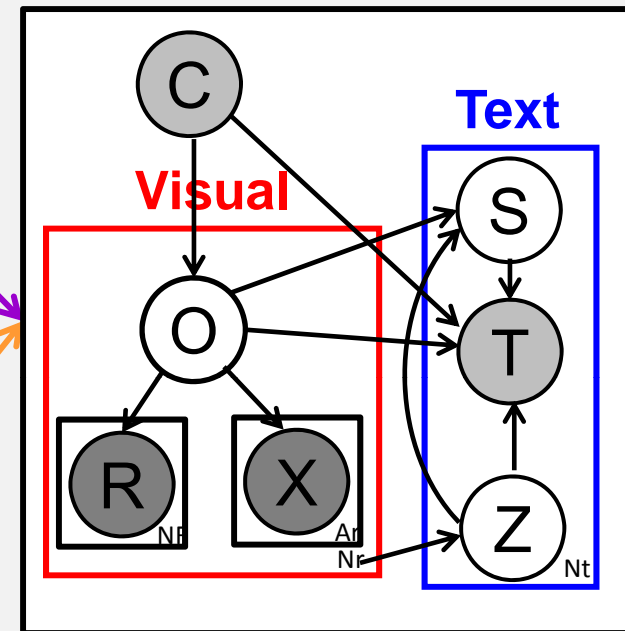
Learning

flickr Small # of automatically initialized images



Large # of uninitialized images

Model



Recognition & Experiment

- Dataset
- Learned Model
- Results

flickr™ 8 Event/Scene Classes

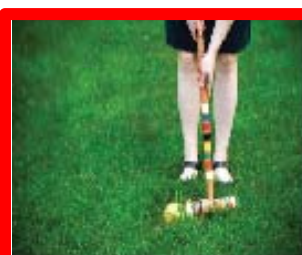
Badminton



Bocce



Croquet



Polo



Remark: Tags are not used during testing

flickr[™] 8 Event/Scene Classes

Rock
climbing



Rowing



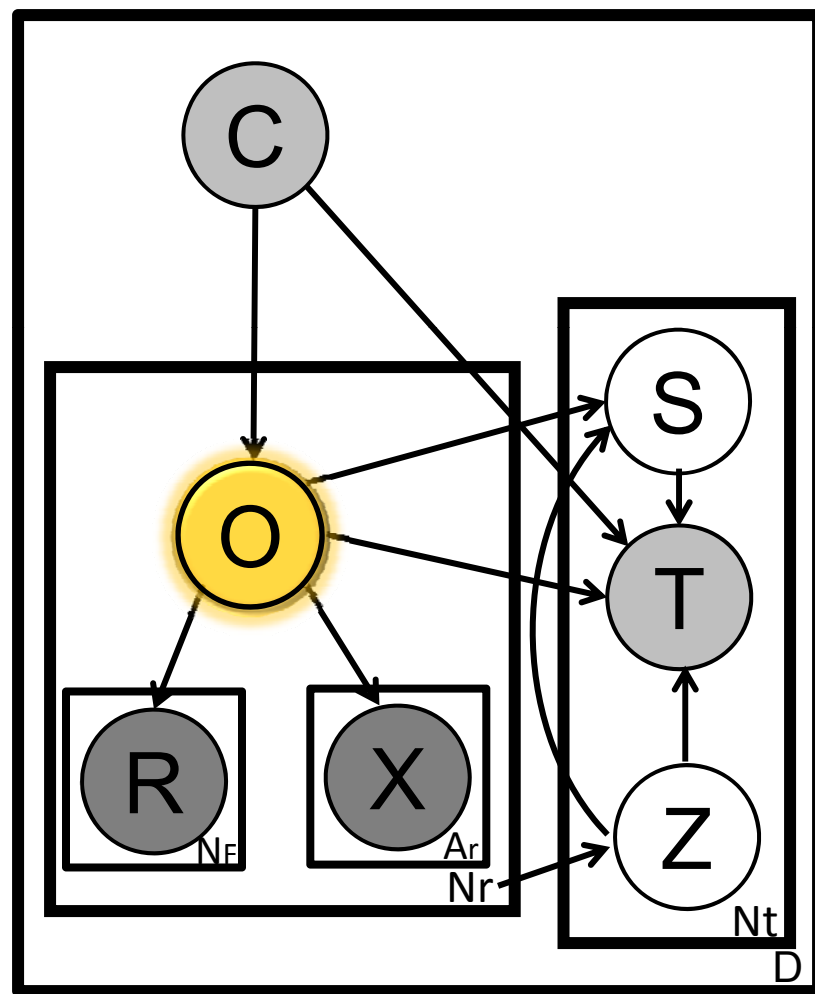
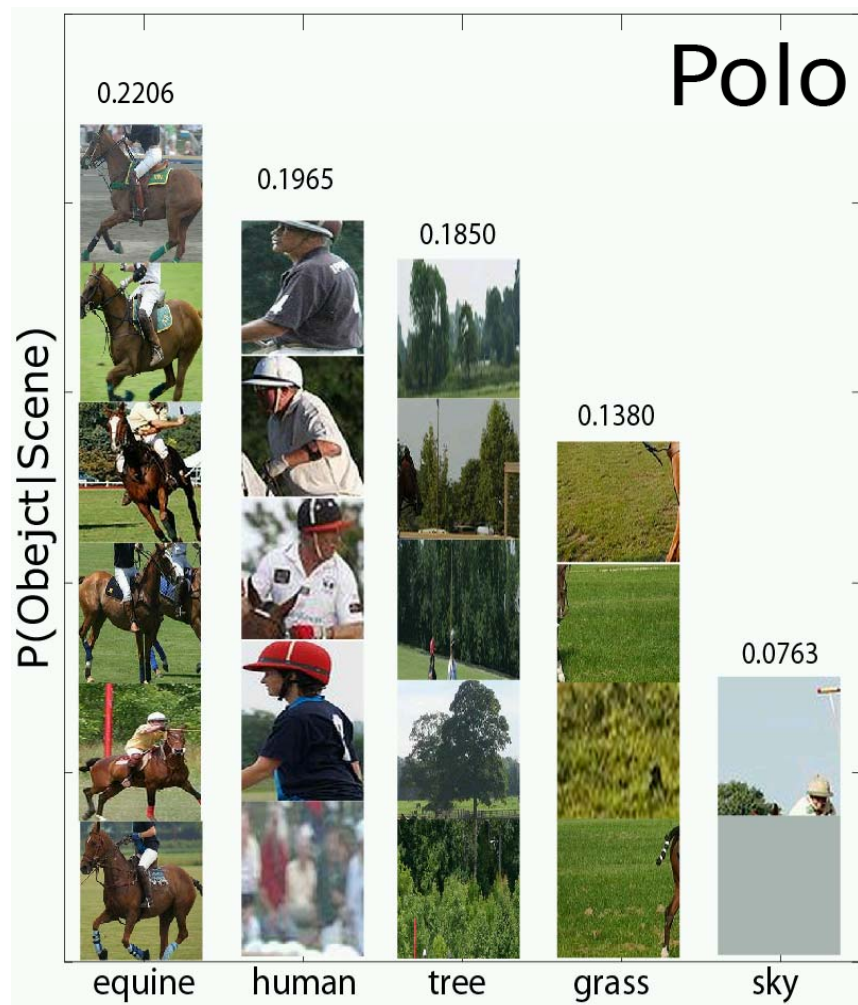
Sailing



Snow
boarding



Learned model: O

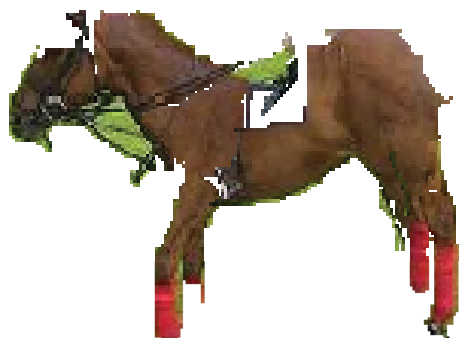




Athlete

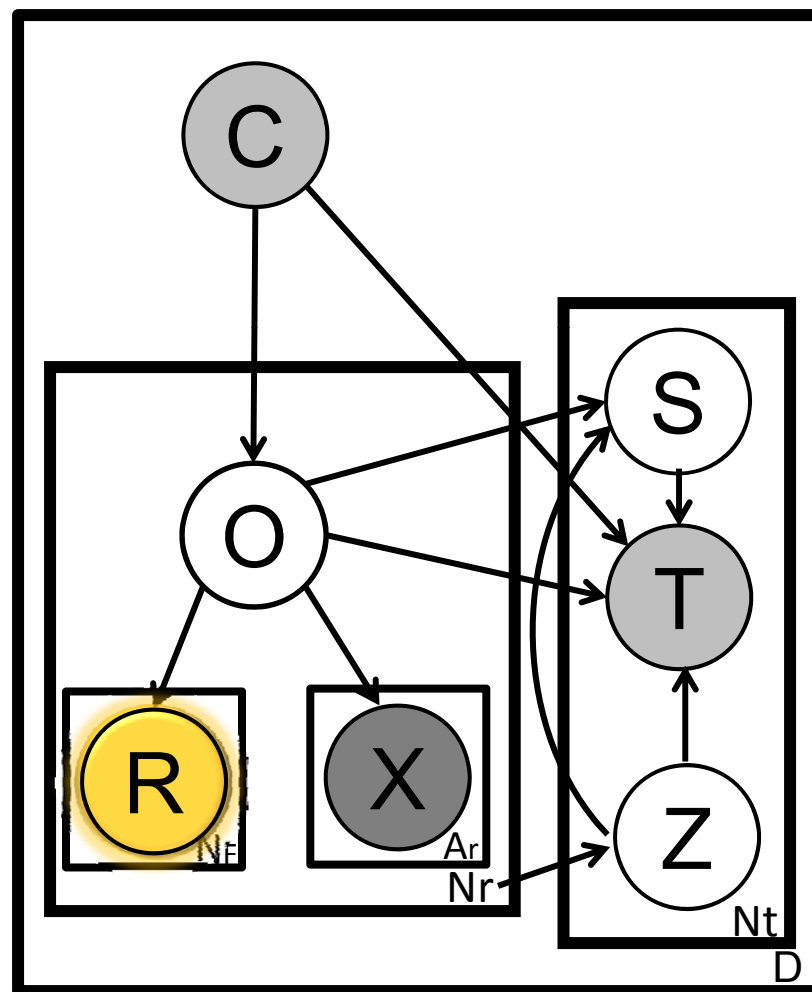


Grass

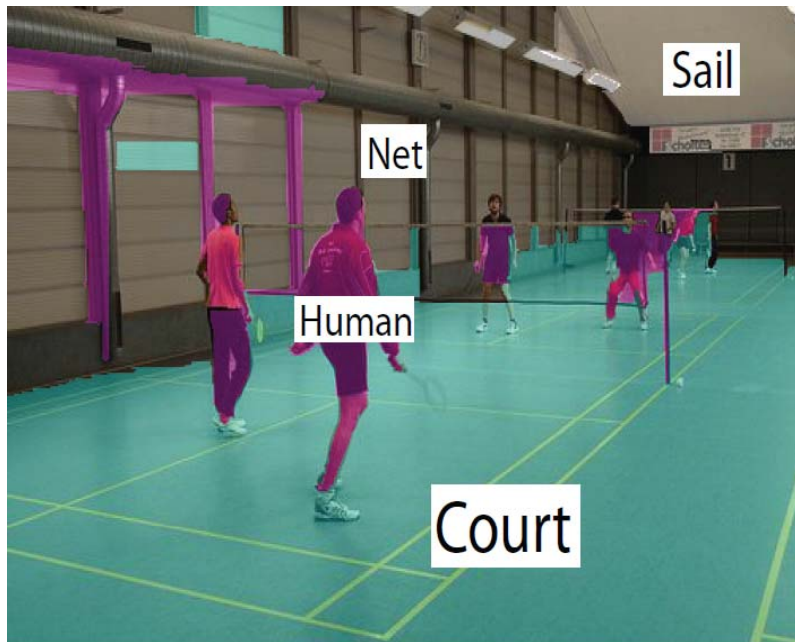


Horse

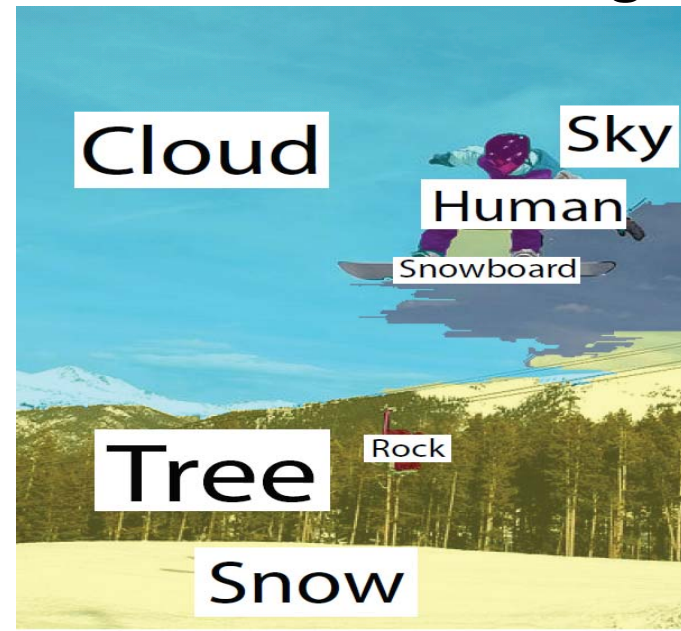
Learned model: \mathcal{R}



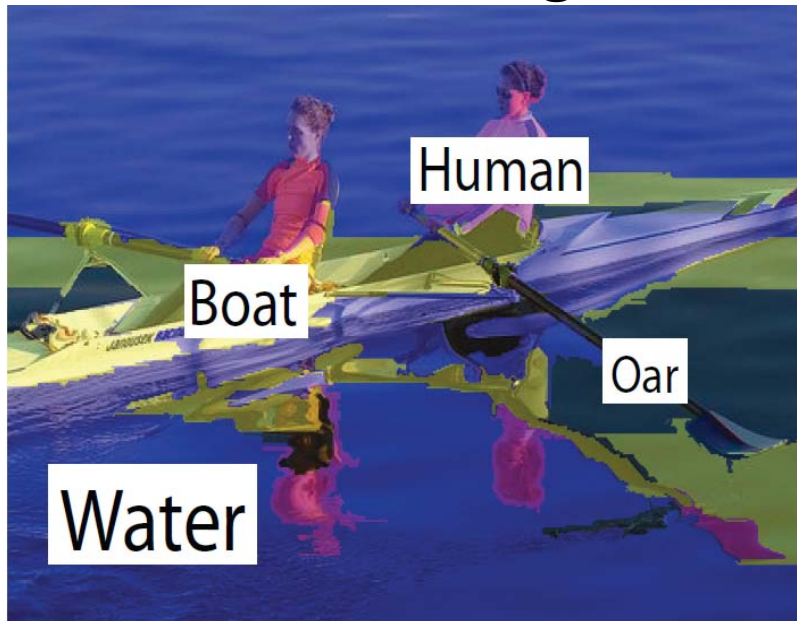
Class: Badminton



Class: Snowboarding



Class: Rowing



Class: Rock Climbing

